
cryptlib

Security Toolkit

Version 3.3.2

Copyright Peter Gutmann 1992-2007

July 2007

You may print a reasonable number of copies of this work for personal use in conjunction with cryptlib software development provided that no fee is charged.

INTRODUCTION	1
cryptlib Overview	1
cryptlib features	2
Architecture	2
S/MIME	3
PGP/OpenPGP	3
Secure Sessions	4
Plug-and-play PKI	4
Certificate Management	4
CA Operations	6
Crypto Devices and Smart Card Support	8
Certificate Store Interface	8
User Interface	9
Security Features	9
Embedded Systems	10
Performance	10
Cryptographic Random Number Management	11
Programming Interface	11
Documentation	11
Algorithm Support	12
Standards Compliance	12
Y2K Compliance	13
Configuration Options	13
cryptlib Applications	13
Encryption Code Example	14
Secure Session Code Example	14
Certificate Management Code Example	15
Document conventions	15
Recommended Reading	15
INSTALLATION	17
AMX	17
BeOS	17
ChorusOS	17
DOS	17
DOS32	17
eCOS	17
μ C/OS-II	18
Embedded Linux	18
μ ITRON	18
Macintosh OS X	18
MVS	18
OS2	18
PalmOS	19
QNX Neutrino	19
RTEMS	19
Tandem	19
uClinux	19
Unix	20
VM/CMS	21
VxWorks	21
Windows 3.x	21
Windows 95/98/ME and Windows NT/2000/XP/Vista	21
Windows CE / Pocket PC / SmartPhone	22
Xilinx XMK	23
Other Systems	23
Key Database Setup	23
Configuration Issues	24
Customised and Cut-down cryptlib Versions	25

Debug vs. Release Versions of cryptlib	25
cryptlib Version Information	25
Support for Vendor-specific Algorithms	26
CRYPTLIB BASICS	27
Programming Interfaces	28
High-level Interface	28
Mid-level Interface	28
Low-level Interface	28
Objects and Interfaces	29
Objects and Attributes	29
Interfacing with cryptlib	30
Initialisation	30
C / C++	31
C# / .NET	31
Delphi	32
Java	32
Python	33
Tcl	33
Visual Basic	33
Return Codes	33
Working with Object Attributes	34
Attribute Types	36
Attribute Lists and Attribute Groups	38
Attribute Cursor Management	39
Object Security	42
Role-based Access Control	44
Managing User Roles	44
Creating and Destroying Users and Roles	45
Miscellaneous Issues	46
Multi-threaded cryptlib Operation	46
Safeguarding Cryptographic Operations	47
Interaction with External Events	48
SECURITY USABILITY FUNDAMENTALS	49
Security (Un-)Usability	49
Theoretical vs. Effective Security	50
User Conditioning	52
Certificates and Conditioned Users	54
SSL Certificates: Indistinguishable from Placebo	55
The User is Trusting... What?	57
Password Mismanagement	60
Other Languages, Other Cultures	61
THE PSYCHOLOGY OF SECURITY USABILITY	66
How Users Make Decisions	66
Consequences of the Human Decision-making Process	68
Security and Rationality	72
Security at Layers 8 and 9	73
The ‘Simon Says’ Problem	77
User Education, and Why it Doesn’t Work	81
SECURITY USABILITY DESIGN	88
Ease of Use	88
Automation vs. Explicitness	90
Safe Defaults	92
Requirements and Anti-requirements	94

Interaction with other Systems	96
Matching Users' Mental Models	97
Activity-Based Planning	99
Design Example: Key Generation	102
Use of Familiar Metaphors	104
SECURITY USER INTERACTION	108
Speaking the User's Language	108
Effective Communication with Users	108
Design Example: Connecting to a Server whose Key has Changed	110
Design Example: Inability to Connect to a Required Server	114
Use of Visual Cues	116
Design Example: TLS Password-based Authentication	120
Design Example: Other Password Protection Mechanisms	121
Design Example: Strengthening Passwords against Dictionary Attacks	123
Legal Considerations	124
SECURITY USABILITY TESTING	128
Pre-implementation Testing	128
Stereotypical Users	128
Input from Users	130
Post-implementation Testing	131
User Testing	131
Usability Testing Examples	132
Encrypted Email	132
Browser Cookies	133
Key Storage	133
Banking Passwords	134
Password Managers	135
File Sharing	136
Site Images	137
Signed Email	139
Post-delivery Reviews	140
DATA ENVELOPING	143
Creating/Destroying Envelopes	143
The Data Enveloping Process	144
Data Size Considerations	146
Basic Data Enveloping	147
Compressed Data Enveloping	149
Password-based Encryption Enveloping	149
Conventional Encryption Enveloping	151
Authenticated Enveloping	152
De-enveloping Mixed Data	153
De-enveloping with a Large Envelope Buffer	154
Obtaining Envelope Security Parameters	155
Enveloping Large Data Quantities	155
Alternative Processing Techniques	157
Enveloping with Many Enveloping Attributes	158
ADVANCED ENVELOPING	160
Public-Key Encrypted Enveloping	160
Digitally Signed Enveloping	164
Enveloping with Multiple Attributes	166
Processing Multiple De-enveloping Attributes	167
Nested Envelopes	169
S/MIME	171

S/MIME Enveloping	171
Encrypted Enveloping	172
Digitally Signed Enveloping	174
Detached Signatures	175
Alternative Detached Signature Processing	176
Extra Signature Information	177
Timestamping	178
PGP	180
PGP Enveloping	180
Encrypted Enveloping	180
Digitally Signed Enveloping	182
Detached Signatures	183
FROM ENVELOPES TO EMAIL	185
S/MIME email	185
Data	185
Signed Data	185
Detached Signature	185
Encrypted Data	186
Nested Content	186
PGP email	186
Implementing S/MIME and PGP email using cryptlib	187
c-client/IMAP4	187
Eudora	188
MAPI	188
Windows 95/98/ME and NT/2000/XP/Vista Shell	188
SECURE SESSIONS	190
Creating/Destroying Session Objects	190
Client vs. Server Sessions	192
Server Names/URLs	192
Server Private Keys	193
Establishing a Session	194
Persistent Connections	194
SSH Sessions	195
SSH Client Sessions	195
SSH Server Sessions	196
SSH Channels	198
SSH Subsystems	199
SSH Port Forwarding	200
SSH Multiple Channels	201
SSL/TLS Sessions	202
SSL/TLS Client Sessions	203
SSL/TLS with Shared Keys	203
SSL/TLS with Client Certificates	204
SSL/TLS Server Sessions	205
SSL/TLS Servers with Shared Keys	205
SSL/TLS Servers with Client Certificates	206
Request/Response Protocol Sessions	207
RTCS Server Sessions	207
OCSP Server Sessions	207
TSP Server Sessions	208
Obtaining Session Status Information	209
Obtaining Session Security Parameters	209
Authenticating the Host with Key Fingerprints	209
Authenticating the Host or Client using Certificates	209
Authenticating the Client via Port and Address	210

Exchanging Data	210
Network Issues	212
Secure Sessions with Proxies	212
Network Timeouts	212
Managing your Own Network Connections and I/O	213
KEY GENERATION AND STORAGE	216
Key Generation	216
Generating a Key Pair into an Encryption Context	216
Asynchronous Key Generation	217
Keyset Types	218
Creating/Destroying Keyset Objects	219
File Keysets	220
HTTP Keysets	221
Database Keysets	222
LDAP Keysets	224
Reading a Key from a Keyset	226
Obtaining a Key for a User	226
General Keyset Queries	228
Handling Multiple Certificates with the Same Name	230
Key Group Management	230
Writing a Key to a Keyset	231
Changing a Private Key Password	232
Deleting a Key	233
CERTIFICATES AND CERTIFICATE MANAGEMENT	234
High-level vs. Low-level Certificate Operations	234
Plug-and-play PKI	234
Mid-level Certificate Management	234
Low-level Certificate Management	234
Certificates and Keys	235
Using Separate Signature and Encryption Certificates	235
Plug-and-play PKI	236
Simple Certificate Creation	237
The Certification Process	239
Obtaining Certificates using CMP	242
CMP Certificate Requests	243
CMP Operation Types	244
CMP Sessions	245
Obtaining Certificates using SCEP	247
SCEP Certificate Requests	248
SCEP Sessions	248
Certificate Status Checking using RTCS	249
Basic RTCS Queries	250
Creating an RTCS Request	250
Communicating with an RTCS Responder	251
Advanced RTCS Queries	252
Certificate Revocation Checking using OCSP	253
Creating an OCSP Request	253
Communicating with an OCSP Responder	254
Advanced OCSP Queries	255
MANAGING A CERTIFICATION AUTHORITY	256
Creating the Top-level (Root) CA Key	256
Initialising PKI User Information	258
Other PKI User Information	259

PKI User IDs	260
Managing a CA using CMP or SCEP	261
Making Certificates Available Online	262
Managing a CA Directly	264
Recording Incoming Requests	264
Retrieving Stored Requests	264
CA Management Operations	265
Issuing and revoking a Certificate	266
Issuing a CRL	266
Expiring Certificates	266
Recovering after a Restart	266
ENCRYPTION AND DECRYPTION	268
Creating/Destroying Encryption Contexts	268
Generating a Key into an Encryption Context	269
Deriving a Key into an Encryption Context	270
Loading a Key into an Encryption Context	271
Working with Initialisation Vectors	271
Loading Public/Private Keys	272
Loading Multibyte Integers	272
Querying Encryption Contexts	274
Using Encryption Contexts to Process Data	274
Conventional Encryption	275
Public-key Encryption	276
Hashing	276
EXCHANGING KEYS	278
Exporting a Key	278
Exporting using Conventional Encryption	279
Importing a Key	280
Importing using Conventional Encryption	280
Querying an Exported Key Object	281
Extended Key Export/Import	281
Key Agreement	282
SIGNING DATA	284
Querying a Signature Object	285
Extended Signature Creation/Checking	285
CERTIFICATES IN DETAIL	288
Overview of Certificates	288
Certificates and Standards Compliance	288
Certificate Compliance Level Checking	289
Creating/Destroying Certificate Objects	291
Obtaining a Certificate	291
Certificate Structures	292
Attribute Certificate Structure	292
Certificate Structure	294
Certification Request Structure	295
CRL Structure	296
Certificate Attributes	297
Basic Certificate Management	297
Certificate Identification Information	299
DN Structure for Business Use	300
DN Structure for Private Use	301

DN Structure for Use with a Web Server	301
Other DN Structures	301
Working with Distinguished Names	301
Creating Customised DNs	302
Extended Certificate Identification Information	304
Working with GeneralName Components	305
Certificate Fingerprints	306
Importing/Exporting Certificates	306
Signing/Verifying Certificates	308
Certificate Chains	310
Working with Certificate Chains	311
Signing Certificate Chains	311
Checking Certificate Chains	312
Exporting Certificate Chains	313
Certificate Revocation using CRLs	314
Working with CRLs	314
Creating CRLs	314
Advanced CRL Creation	315
Checking Certificates against CRLs	316
Automated CRL Checking	316
Certificate Trust Management	317
Controlling Certificate Usage	317
Implicitly Trusted Certificates	317
Working with Trust Settings	318
CERTIFICATE EXTENSIONS	320
Extension Structure	320
Working with Extension Attributes	320
Composite Extension Attributes	321
X.509 Extensions	322
Alternative Names	322
Basic Constraints	322
Certificate Policies, Policy Mappings, Policy Constraints, and Policy Inhibiting	323
CRL Distribution Points/Freshest CRL and Subject/Authority Information Access	324
Directory Attributes	325
Key Usage, Extended Key Usage, and Netscape certificate type	325
Name Constraints	328
Private Key Usage Period	329
Subject and Authority Key Identifiers	329
CRL Extensions	329
CRL Reasons, CRL Numbers, Delta CRL Indicators	329
Hold Instruction Code	331
Invalidity Date	331
Issuing Distribution Point and Certificate Issuer	331
Digital Signature Legislation Extensions	332
Certificate Generation Date	332
Other Restrictions	332
Reliance Limit	332
Signature Delegation	333
Qualified Certificate Extensions	333
Biometric Info	333
QC Statements	333
SET Extensions	334
SET Card Required and Merchant Data	334
SET Certificate Type, Hashed Root Key, and Tunnelling	334
Application-specific Extensions	335

OCSP Extensions	335
Vendor-specific Extensions	335
Netscape Certificate Extensions	336
Thawte Certificate Extensions	336
Generic Extensions	336
OTHER CERTIFICATE OBJECT EXTENSIONS	338
CMS/SMIME Attributes	338
Content Type	338
Countersignature	339
Message Digest	339
Signing Description	339
Signing Time	339
Extended CMS/SMIME Attributes	339
AuthentiCode Attributes	340
Content Hints	341
DOMSEC Attributes	341
Mail List Expansion History	341
Nonce	342
Receipt Request	342
SCEP Attributes	342
Security Label, Equivalent Label	343
Signature Policy	344
S/MIME Capabilities	345
Signing Certificate	345
OCSP Attributes	346
CRYPTLIB USER INTERFACE COMPONENTS	347
Displaying Certificates	347
Key/Certificate Generation	347
ENCRYPTION DEVICES AND MODULES	350
Creating/Destroying Device Objects	350
Activating and Controlling Cryptographic Devices	351
Device Initialisation	351
User Authentication	352
Device Zeroisation	353
Working with Device Objects	353
Key Storage in Crypto Devices	354
Querying Device Information	354
Considerations when Working with Devices	355
Fortezza Cards	356
PKCS #11 Devices	356
Installing New PKCS #11 Modules	356
Accessing PKCS #11 Devices	357
CryptoAPI	357
MISCELLANEOUS TOPICS	359
Querying cryptlib's Capabilities	359
Working with Configuration Options	359
Querying/Setting Configuration Options	362
Saving Configuration Options	363
Obtaining Information About Cryptlib	363
Random Numbers	364
Gathering Random Information	364
Obtaining Random Numbers	365
Working with Newer Versions of cryptlib	365

ERROR HANDLING	367
Extended Error Reporting	369
EMBEDDED SYSTEMS	372
Embedded OS Types	372
AMX	372
ChorusOS	373
DOS	373
eCOS	373
μ C/OS-II	373
Embedded Linux	373
μ ITRON	373
PalmOS	374
QNX Neutrino	374
RTEMS	374
uClinux	374
Windows CE	374
VxWorks	374
Xilinx XMK	375
Embedded cryptlib Configuration Options	375
Debugging with Embedded cryptlib	377
Porting to Devices without a Filesystem	377
Porting to Devices without Dynamic Memory Allocation	377
Memory Allocation Strategy	378
cryptlib Memory Usage	378
Tracking Memory Usage	378
Porting to Devices without Randomness/Entropy Sources	379
DATABASE AND NETWORKING PLUGINS	380
The Database Plugin Interface	380
Database Plugin Functions	381
The Network Plugin Interface	384
Network Plugin Functions	384
The Crypto Plugin Interface	385
ALGORITHMS AND STANDARDS CONFORMANCE	387
AES	387
Blowfish	387
CAST-128	387
DES	388
Triple DES	388
Diffie-Hellman	389
DSA	389
Elgamal	389
HMAC-MD5	390
HMAC-SHA1	390
HMAC-RIPEMD-160	390
IDEA	390
MD2	391
MD4	391
MD5	392
RC2	392
RC4	392
RC5	393

RIPEMD-160	393
RSA	393
SHA	393
SHA2	394
Skipjack	394
DATA TYPES AND CONSTANTS	395
CRYPT_ALGO_TYPE	395
CRYPT_ATTRIBUTE_TYPE	396
CRYPT_CERTFORMAT_TYPE	396
CRYPT_CERTTYPE_TYPE	397
CRYPT_DEVICE_TYPE	397
CRYPT_FORMAT_TYPE	397
CRYPT_KEYID_TYPE	398
CRYPT_KEYOPT_TYPE	398
CRYPT_KEYSET_TYPE	398
CRYPT_MODE_TYPE	399
CRYPT_OBJECT_TYPE	399
CRYPT_SESSION_TYPE	399
Data Size Constants	400
Miscellaneous Constants	400
DATA STRUCTURES	402
CRYPT_OBJECT_INFO Structure	402
CRYPT_PKCINFO_xxx Structures	402
CRYPT_QUERY_INFO Structure	403
FUNCTION REFERENCE	404
cryptAddCertExtension	404
cryptAddPrivateKey	404
cryptAddPublicKey	404
cryptAddRandom	405
cryptAsyncCancel	405
cryptAsyncQuery	405
cryptCAAddItem	405
cryptCACertManagement	406
cryptCAGetItem	406
cryptCheckCert	407
cryptCheckSignature	407
cryptCheckSignatureEx	407
cryptCreateCert	408
cryptCreateContext	408
cryptCreateEnvelope	408
cryptCreateSession	409
cryptCreateSignature	409
cryptCreateSignatureEx	409
cryptDecrypt	410
cryptDeleteAttribute	410
cryptDeleteCertExtension	411
cryptDeleteKey	411

cryptDestroyCert	411
cryptDestroyContext	412
cryptDestroyEnvelope	412
cryptDestroyObject	412
cryptDestroySession	412
cryptDeviceClose	412
cryptDeviceCreateContext	413
cryptDeviceOpen	413
cryptDeviceQueryCapability	413
cryptEncrypt	414
cryptEnd	414
cryptExportCert	414
cryptExportKey	415
cryptExportKeyEx	415
cryptFlushData	416
cryptGenerateKey	416
cryptGenerateKeyAsync	417
cryptGetAttribute	417
cryptGetAttributeString	417
cryptGetCertExtension	418
cryptGetPrivateKey	418
cryptGetPublicKey	419
cryptImportCert	419
cryptImportKey	420
cryptInit	420
cryptKeysetClose	420
cryptKeysetOpen	420
cryptPopData	421
cryptPushData	421
cryptQueryCapability	422
cryptQueryObject	422
cryptSetAttribute	422
cryptSetAttributeString	423
cryptSignCert	423
cryptUIDisplayCert	423
cryptUIGenerateKey	424
ACKNOWLEDGEMENTS	425

Introduction

The information age has seen the development of electronic pathways that carry vast amounts of valuable commercial, scientific, and educational information between financial institutions, companies, individuals, and government organisations. Unfortunately the unprecedented levels of access provided by systems like the Internet also expose this data to breaches of confidentiality, disruption of service, and outright theft. As a result, there is an enormous (and still growing) demand for the means to secure these online transactions. One report by the Computer Systems Policy Project (a consortium of virtually every large US computer company, including Apple, AT&T, Compaq, Digital, IBM, Silicon Graphics, Sun, and Unisys) estimated that the potential revenue arising from these security requirements in the US alone could be as much as US\$30-60 billion in the next few years, and the potential exposure to global users from a lack of this security is projected to reach between US\$320 and 640 billion.

Unfortunately the security systems required to protect data are generally extremely difficult to design and implement, and even when available tend to require considerable understanding of the underlying principles in order to be used. This has led to a proliferation of “snake oil” products that offer only illusionary security, or to organisations holding back from deploying online information systems because the means to secure them aren’t readily available, or because they employed weak, easily broken security that was unacceptable to users.

The cryptlib security toolkit provides the answer to this problem. A complete description of the capabilities provided by cryptlib is given below.

cryptlib Overview

cryptlib is a powerful security toolkit that allows even inexperienced crypto programmers to easily add encryption and authentication services to their software. The high-level interface provides anyone with the ability to add strong security capabilities to an application in as little as half an hour, without needing to know any of the low-level details that make the encryption or authentication work. Because of this, cryptlib dramatically reduces the cost involved in adding security to new or existing applications.

At the highest level, cryptlib provides implementations of complete security services such as S/MIME and PGP/OpenPGP secure enveloping, SSL/TLS and SSH secure sessions, CA services such as CMP, SCEP, RTCS, and OCSP, and other security operations such as secure timestamping (TSP). Since cryptlib uses industry-standard X.509, S/MIME, PGP/OpenPGP, and SSH/SSL/TLS data formats, the resulting encrypted or signed data can be easily transported to other systems and processed there, and cryptlib itself runs on virtually any operating system — cryptlib doesn’t tie you to a single platform. This allows email, files, and EDI transactions to be authenticated with digital signatures and encrypted in an industry-standard format.

cryptlib provides an extensive range of other capabilities including full X.509/PKIX certificate handling (all X.509 versions from X.509v1 to X.509v4) with additional support for SET, Microsoft AuthenticCode, Identrus, SigG, S/MIME, SSL, and Qualified certificates, PKCS #7 certificate chains, handling of certification requests and CRLs including automated checking of certificates against CRLs and online checking using RTCS and OCSP, and issuing and revoking certificates using CMP and SCEP. In addition cryptlib implements a full range of certification authority (CA) functions, as well as providing complete CMP, SCEP, RTCS, and OCSP server implementations to handle online certificate enrolment/issue/revocation and certificate status checking. Alongside the certificate handling, cryptlib provides a sophisticated key storage interface that allows the use of a wide range of key database types ranging from PKCS #11 devices, PKCS #15 key files, and PGP/OpenPGP key rings through to commercial-grade RDBMS’ and LDAP directories with optional SSL protection.

In addition to its built-in capabilities, cryptlib can make use of the crypto capabilities of a variety of external crypto devices such as hardware crypto accelerators, Fortezza cards, PKCS #11 devices, hardware security modules (HSMs), and crypto smart cards. For particularly demanding applications cryptlib can be used with a variety of crypto devices that have received appropriate FIPS 140 or ITSEC/Common Criteria certification. The crypto device interface also provides a convenient general-purpose plug-in capability for adding new functionality that will be automatically used by cryptlib.

cryptlib is supplied as source code for AMX, BeOS, ChorusOS, DOS, DOS32, eCOS, µC/OS-II, embedded Linux, IBM MVS, µITRON, Macintosh/OS X, OS/2, PalmOS, RTEMS, Tandem, a variety of Unix versions (including AIX, Digital Unix, DGUX, FreeBSD/NetBSD/OpenBSD, HP-UX, IRIX, Linux, MP-RAS, OSF/1, QNX, SCO/UnixWare, Solaris, SunOS, Ultrix, and UTS4), uClinux, VM/CMS, VxWorks, Windows 3.x, Windows 95/98/ME, Windows CE/PocketPC/SmartPhone, Windows NT/2000/XP/Vista, and Xilinx XMK. cryptlib's highly portable nature means that it is also being used in a variety of custom embedded system environments. In addition, cryptlib is available as a standard Windows DLL and an ActiveX control.. cryptlib comes with language bindings for C / C++, C# / .NET, Delphi, Java, Python, and Visual Basic (VB).

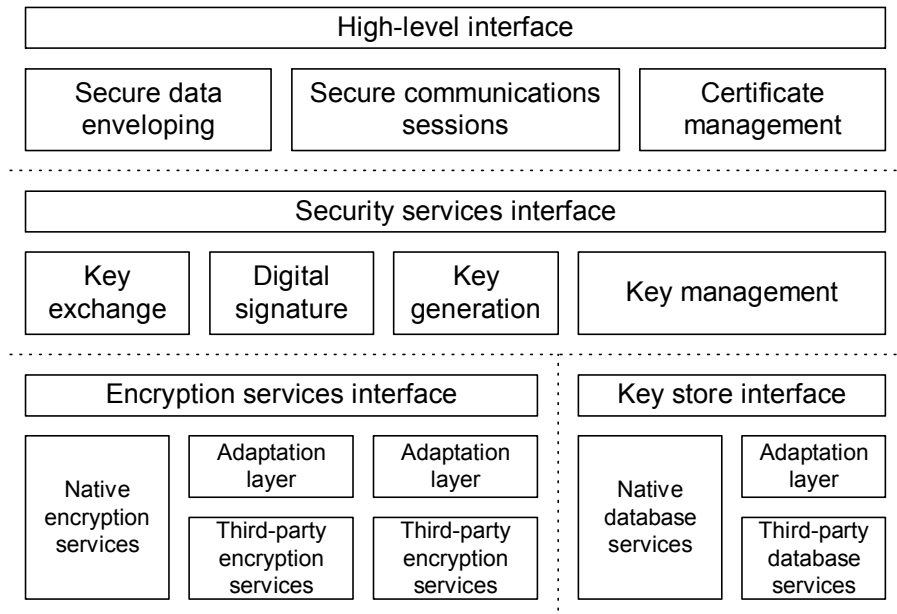
cryptlib features

cryptlib provides a standardised interface to a number of popular encryption algorithms, as well as providing a high-level interface that hides most of the implementation details and uses operating-system-independent encoding methods that make it easy to transfer secured data from one operating environment to another. Although use of the high-level interface is recommended, experienced programmers can directly access the lower-level encryption routines for implementing custom encryption protocols or methods not directly provided by cryptlib.

Architecture

cryptlib consists of a set of layered security services and associated programming interfaces that provide an integrated set of information and communications security capabilities. Much like the network reference model, cryptlib contains a series of layers that provide each level of abstraction, with higher layers building on the capabilities provided by the lower layers.

At the lowest level are basic components such as core encryption and authentication routines, which are usually implemented in software but may also be implemented in hardware (due to the speed of the software components used in cryptlib, the software is usually faster than dedicated hardware). At the next level are components that wrap up the specialised and often quite complex core components in a layer that provides abstract functionality and ensures complete cross-platform portability of data. These functions typically cover areas such as “create a digital signature” or “exchange an encryption key”. At the highest level are extremely powerful and easy-to-use functions such as “encrypt a message”, “sign a message”, “open a secure link”, and “create a digital certificate” that require no knowledge of encryption techniques, and that take care of complex issues such as key management, data encoding, en/decryption, and digital signature processing.



cryptlib's powerful object management interface provides the ability to add encryption and authentication capabilities to an application without needing to know all the low-level details that make the encryption or authentication work. The automatic object-management routines take care of encoding issues and cross-platform portability problems, so that a handful of function calls is all that's needed to wrap up data in signed or encrypted form with all of the associated information and parameters needed to recreate it on the other side of a communications channel. This provides a considerable advantage over other encryption toolkits that often require hundreds of lines of code and the manipulation of complex encryption data structures to perform the same task.

S/MIME

cryptlib employs the IETF-standardised Cryptographic Message Syntax (CMS, formerly called PKCS #7) format as its native data format. CMS is the underlying format used in the S/MIME secure mail standard, as well as a number of other standards covering secure EDI and related systems like HL7 medical messaging and the Session Initiation Protocol (SIP) for services such as Internet telephony and instant messaging. As an example of its use in secure EDI, cryptlib provides security services for the Symphonia EDI messaging toolkit which is used to communicate medical lab reports, patient data, drug prescription information, and similar information requiring a high level of security.

The S/MIME implementation uses cryptlib's enveloping interface which allows simple, rapid integration of strong encryption and authentication capabilities into existing email agents and messaging software. The resulting signed enveloped data format provides message integrity and origin authentication services, the encrypted enveloped data format provides confidentiality. In addition cryptlib's S/MIME implementation allows external services such as trusted timestamping authorities (TSAs) to be used when a signed message is created, providing externally-certified proof of the time of message creation. The complexity of the S/MIME format means that the few other toolkits that are available require a high level of programmer knowledge of S/MIME processing issues. In contrast cryptlib's enveloping interface makes the process as simple as pushing raw data into an envelope and popping the processed data back out, a total of three function calls, plus one more call to add the appropriate encryption or signature key.

PGP/OpenPGP

Alongside the PKCS #7/CMS/SMIME formats, cryptlib supports the PGP/OpenPGP message format, allowing it to be used to send and receive PGP-encrypted email and

data. As with the S/MIME implementation, the PGP implementation uses cryptlib's enveloping interface to allow simple, rapid integration of strong encryption and authentication capabilities into existing email agents and messaging software. Since the enveloping interface is universal, the process involved in creating PGP and S/MIME messages is identical except for the envelope format specifier, allowing a one-off development effort to handle any secure message format.

Secure Sessions

cryptlib secure sessions can include SSH, SSL, and TLS sessions, and general communications sessions can include protocols such as the certificate management protocol (CMP), simple certificate enrolment protocol (SCEP), real-time certificate status protocol (RTCS), online certificate status protocol (OCSP), and timestamping (TSP). As with envelopes, cryptlib takes care of the session details for you so that all you need to do is provide basic communications information such as the name of the server or host to connect to and any other information required for the session such as a password or certificate. cryptlib takes care of establishing the session and managing the details of the communications channel and its security parameters.

cryptlib provides both client and server implementations of all session types. By tying a key or certificate store to the session, you can let cryptlib take care of any key management issues for you. For example, with an SSH, SSL or TLS server session cryptlib will use the key/certificate store to authenticate incoming connections, and with a CMP or SCEP server session cryptlib will use the certificate store to handle the certificate management process. In this way a complete CMP-based CA that handles enrolment, certificate update and renewal, and certificate revocation, can be implemented with only a handful of function calls.

Plug-and-play PKI

Working with certificates can be complex and painful, requiring the use of a number of arcane and difficult-to-use mechanisms to perform even the simplest operations. To eliminate this problem, cryptlib provides a plug-and-play PKI interface that manages all certificate processing and management operations for you, requiring no special knowledge of certificate formats, protocols, or operations. Using the plug-and-play PKI interface with an appropriately-configured CA means that cryptlib will automatically and transparently handle key generation, certificate enrolment, securely obtaining trusted CA certificates, and certifying the newly-generated keys for the user, all in a single operation. Similarly, certificate validity checking can be performed using an online real-time status check that avoids the complexity and delayed status information provided by mechanisms such as CRLs. The plug-and-play PKI interface removes most of the complexity and difficulty involved in working with certificates, making it easier to use certificates than with any of the conventional certificate management mechanisms.

Certificate Management

cryptlib implements full X.509 certificate support, including all X.509 version 3, version 4, and version 5 extensions as well as extensions defined in the IETF PKIX certificate profile. cryptlib also supports additional certificate types and extensions including SET certificates, Microsoft AuthentiCode and Netscape and Microsoft server-gated crypto certificates, Identrus certificates, qualified certificates, S/MIME and SSL client and server certificates, SigG extensions, and various vendor-specific extensions such as Netscape certificate types and the Thawte secure extranet.

In addition to certificate handling, cryptlib allows the generation of certification requests suitable for submission to certification authorities (CAs) in order to obtain a certificate. Since cryptlib is itself capable of processing certification requests into certificates, it is also possible to use cryptlib to provide full CA services. cryptlib also supports the creating and handling of the certificate chains required for S/MIME, SSL, and other applications, and the creation of certificate revocation lists (CRLs) with the capability to check certificates against existing or new CRLs either automatically or under programmer control. In addition to CRL-based revocation

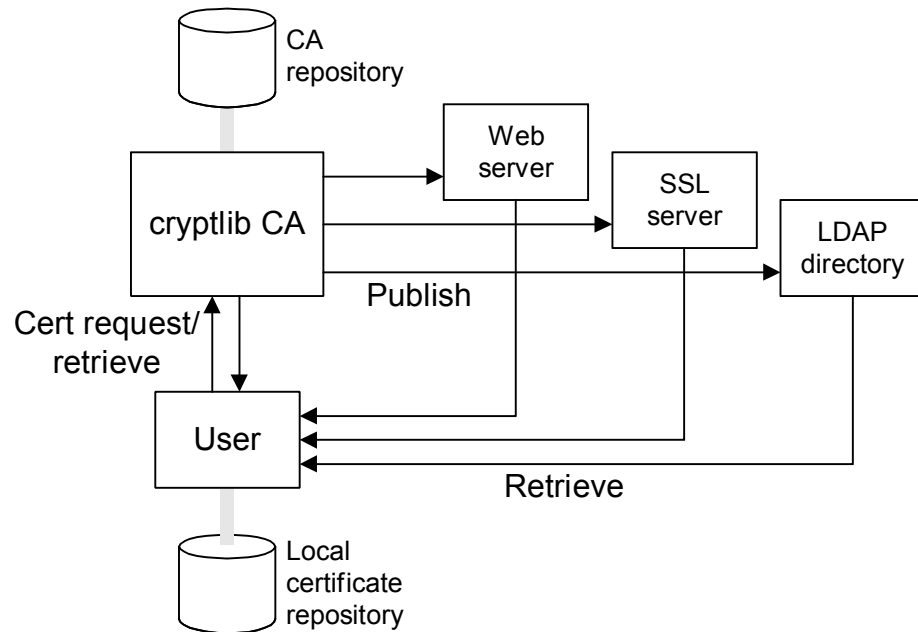
checking, cryptlib also supports online status protocols such as RTCS and OCSP. cryptlib also implements the CMP protocol which fully automates the management of certificates, allowing online certificate enrolment, issue, update/replacement, and revocation of certificates, and the SCEP protocol, which automates the certificate issue process. Using CMP removes from the user any need for technical knowledge of certificate management, since all details are managed by the CA.

cryptlib can import and export certification requests, certificates, certificate chains, and CRLs, covering the majority of certificate transport formats used by a wide variety of software such as web browsers and servers. The certificate types that are supported include:

- Basic X.509 version 1 and 2 certificates
- Extended X.509 version 3, 4, and 5 certificates
- Certificates conformant to the IETF PKIX profile
- SSL/TLS server and client certificates
- S/MIME email certificates
- SET certificates
- SigG certificate extensions
- AuthentiCode code signing certificates
- Identrus certificates
- Qualified certificates
- IPsec server, client, end-user, and tunnelling certificates
- Server-gated crypto certificates
- Timestamping certificates

In addition cryptlib supports X.509v3, X.509v4, X.509v5, IETF, S/MIME, SET, and SigG certificate extensions and many vendor-specific extensions including ones covering public and private key usage, certificate policies, path and name constraints, policy constraints and mappings, and alternative names and other identifiers. This comprehensive coverage makes cryptlib a single solution for almost all certificate processing requirements.

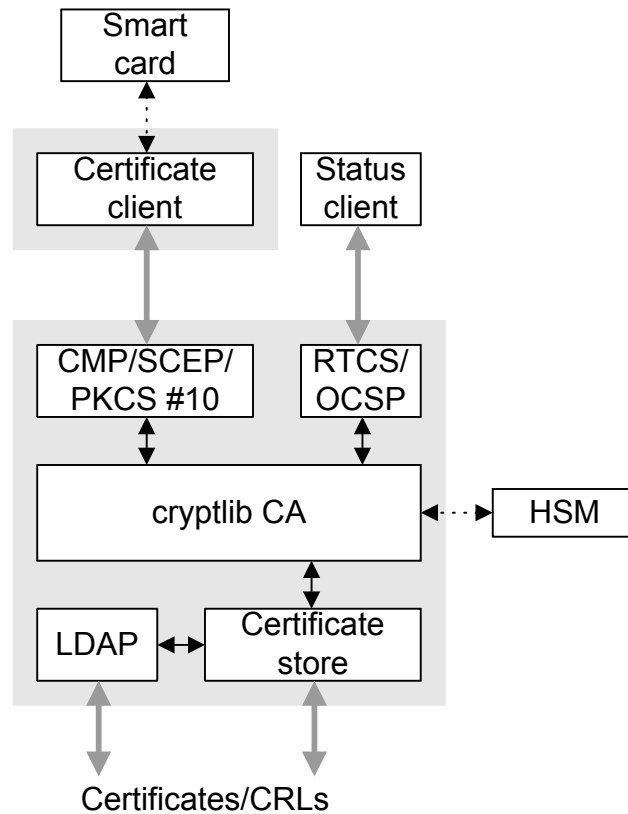
The diagram below shows a typical cryptlib application, in which it provides the full functionality of both a CA (processing certification requests, storing the issued certificates locally in a certificate database, and optionally publishing the certificates on the web or in an LDAP directory) and an end entity (generating certification requests, submitting them to a CA, and retrieving the result from the web or a directory service).



To handle certificate trust and revocation issues, cryptlib includes a certificate trust manager that can be used to automatically manage CA trust settings. For example a CA can be designated as a trusted issuer that will allow cryptlib to automatically evaluate trust along certificate chains. Similarly, cryptlib can automatically check certificates against RTCS and OCSP responders and CRLs published by CAs, removing from the user the need to perform complex manual checking.

CA Operations

cryptlib includes a scalable, flexible Certificate Authority (CA) engine built on the transaction-processing capabilities of a number of proven, industrial-strength relational databases running on a variety of hardware platforms. The CA facility provides an automated means of handling certificate issuance without dealing directly with the details of processing request, signing certificates, saving the resulting certificates in keys stores, and assembling CRLs. This constitutes a complete CA system for issuance and management of certificates and CRLs. A typical cryptlib CA configuration is shown below.



Available CA operations include:

- Certificate enrolment/initialisation operations
- Certificate issue
- Certificate update/key update
- Certificate expiry management
- Revocation request processing
- CRL issue

All CA operations are recorded to an event log using cryptlib's built-in CA logging/auditing facility, which provides a full account of certificate requests, certificates issued or renewed, revocations requested and issued, certificates expired, and general CA management operations. The logs may be queried for information on all events or a specified subset of events, for example all certificates that were issued on a certain day.

cryptlib contains a full implementation of a CMP server (to handle online certificate management), and SCEP server (to handle online certificate issue), a RTCS server (to handle real-time certificate status checking), and an OCSP server (to handle revocation checking). All of these servers are fully automated, requiring little user intervention beyond the initial enrolment process in which user eligibility for a certificate is established. These services make it easier than ever to manage your own CA. Certificate expiration and revocation are handled automatically by the CA engine. Expired certificates are removed from the certificate store, and CRLs are assembled from previously processed certificate revocation requests. These operations are handled with a single function call.

The CA keys can optionally be generated and held in tamper-resistant hardware security modules, with certificate signing being performed by the hardware module. Issued certificates can be stored on smart cards or similar crypto devices in addition to being managed using software-only implementations. The CA facility supports the simultaneous operation of multiple CAs, for example to manage users served through

divisional CAs certified by a root CA. Each CA can issue multiple certificates to users, allowing the use of separate keys bound to signature and encryption certificates.

Crypto Devices and Smart Card Support

In addition to its built-in capabilities, cryptlib can make use of the crypto capabilities of a variety of external crypto devices such as:

- Hardware crypto accelerators
- Fortezza cards
- PKCS #11 devices
- Crypto smart cards
- Hardware security modules (HSMs)
- PCI crypto cards
- Dallas iButtons
- Datakeys/iKeys
- PCMCIA crypto tokens
- USB tokens

These devices will be used by cryptlib to handle functions such as key generation and storage, certificate creation, digital signatures, and message en- and decryption.

Typical applications include:

- Running a certification authority inside tamper-resistant hardware
- Smart-card based digital signatures
- Message encryption/decryption in secure hardware

cryptlib manages any device-specific interfacing requirements so that the programming interface for any crypto device is identical to cryptlib's native interface, allowing existing applications that use cryptlib to be easily and transparently migrated to using crypto devices. The ability to mix and match crypto devices and the software-only implementation allows appropriate tradeoffs to be chosen between flexibility, cost, and security.

Certificate Store Interface

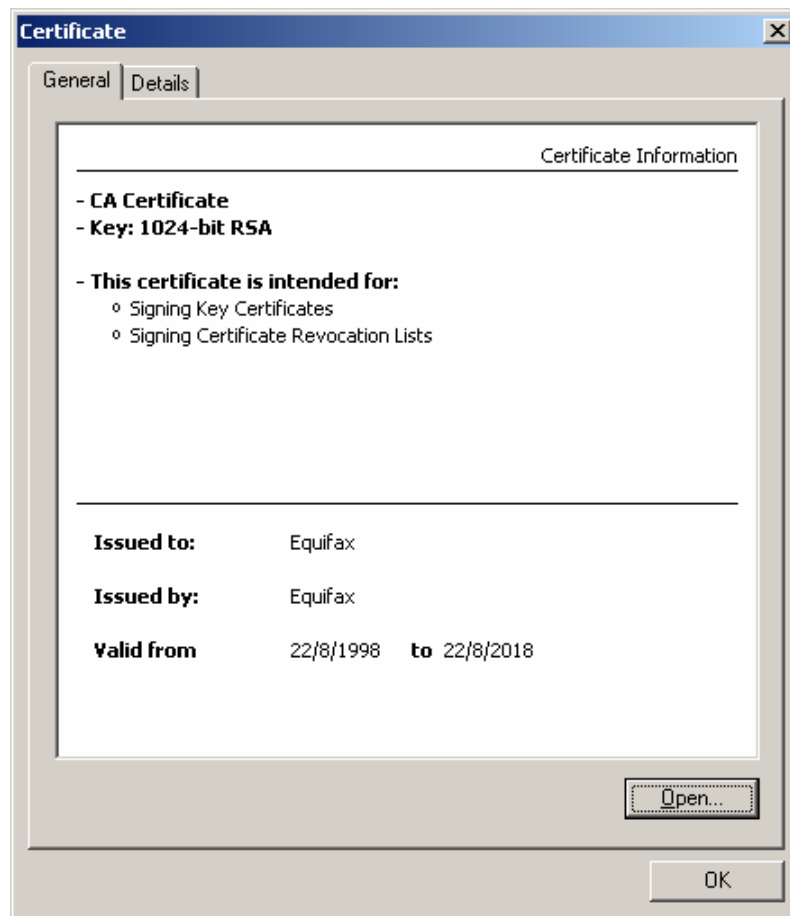
cryptlib utilizes commercial-strength RDBMS' to store keys in the internationally standardised X.509 format. The certificate store integrates seamlessly into existing databases and can be managed using existing tools. For example a key database stored on an MS SQL Server might be managed using Visual Basic or MS Access; a key database stored on an Oracle server might be managed through SQL*Plus.

In addition to standard certificate stores, cryptlib supports the storage and retrieval of certificates in LDAP directories, HTTP access for keys accessible via the web, and external flat-file key collections such as PKCS #15 soft-tokens and PGP/OpenPGP key rings. The key collections may be freely mixed (so for example a private key could be stored in a PKCS #15 soft-token, a PGP/OpenPGP key ring or on a smart card with the corresponding X.509 certificate being stored in a certificate store, an LDAP directory, or on the web).

Private keys may be stored on disk encrypted with an algorithm such as triple DES or AES (selectable by the user), with the password processed using several thousand iterations of a hashing algorithm such as SHA-1 (also selectable by the user). Where the operating system supports it, cryptlib will apply system security features such as ACLs under Windows NT/2000/XP/Vista and file permissions under Unix to the private key file to further restrict access.

User Interface

In addition to its general security functionality, cryptlib includes a number of user interface components that simplify the task of working with keys and certificates. Components such as the certificate viewer shown below allow users to browse the contents of certificates, certificate chains, requests, and other certificate objects. The key generation wizard simplifies the task of key and certificate generation by handling most of the details of the process automatically, producing a complete public/private key pair and certificate request suitable for submission to a CA, or a self-signed certificate for immediate use. These user interface components remove much of the complexity of the key and certificate management process, allowing developers to concentrate on applying the completed keys and certificates towards securing data, email, or communications sessions rather than on the process needed to create them.



Security Features

cryptlib is built around a security kernel with Orange Book B3-level security features to implement its security mechanisms. This kernel provides the interface between the outside world and the architecture's objects (intra-object security) and between the objects themselves (inter-object security). The security kernel is the basis of the entire cryptlib architecture — all objects are accessed and controlled through it, and all object attributes are manipulated through it. The kernel is implemented as an interface layer that sits on top of the objects, monitoring all accesses and handling all protection functions.

Each cryptlib object is contained entirely within the security perimeter, so that data and control information can only flow in and out in a very tightly-controlled manner, and objects are isolated from each other within the perimeter by the security kernel. For example once keying information has been sent to an object, it can't be retrieved

by the user except under tightly-controlled conditions. In general keying information isn't even visible to the user, since it's generated inside the object itself and never leaves the security perimeter. This design is ideally matched to hardware implementations that perform strict red/black separation, since sensitive information can never leave the hardware.

Associated with each object is a set of mandatory ACLs that determine who can access a particular object and under which conditions the access is allowed. If the operating system supports it, all sensitive information used will be page-locked to ensure that it's never swapped to disk from where it could be recovered using a disk editor. All memory corresponding to security-related data is managed by cryptlib and will be automatically sanitised and freed when cryptlib shuts down even if the calling program forgets to release the memory itself.

Where the operating system supports it, cryptlib will apply operating system security features to any objects that it creates or manages. For example under Windows NT/2000/XP/Vista cryptlib private key files will be created with an access control list (ACL) that allows only the key owner access to the file; under Unix the file permissions will be set to achieve the same result.

Embedded Systems

cryptlib's high level of portability and configurability makes it ideal for use in embedded systems with limited resources or specialised requirements, including ones based on ARM7, ARM9, ARM TDMI, Fujitsu FR-V, Hitachi SuperH, MIPS IV, MIPS V, Motorola ColdFire, NEC V8xx series, NEC VRxxxx series, Panasonic/Matsushita AM33/AM34, PowerPC, Samsung CalmRISC, SH3, SH4, SPARC, SPARClite, StrongArm, TI OMAP, and Intel XScale processors. cryptlib doesn't perform any floating-point operations and runs directly on processors without an FPU.

The code is fully independent of any underlying storage or I/O mechanisms, and works just as easily with abstractions such as named memory segments in flash memory as it does with standard key files on disk. It has been deployed on embedded systems without any conventional I/O capabilities (stdio) or dynamic memory allocation facilities, with proprietary operating system architectures and services including ATMs, printers, web-enabled devices, POS systems, embedded device controllers, and similar environments, and even in devices with no operating system at all (cryptlib runs on the bare metal). It can also run independent of any form of operating system, and has been run on the bare metal in environments with minimal available resources, in effect functioning as a complete crypto operating system for the underlying hardware.

Because cryptlib functions identically across all supported environments, it's possible to perform application development in a full-featured development environment such as Windows or Unix and only when the application is complete and tested move it to the embedded system. This flexibility saves countless hours of development time, greatly reducing the amount of time that needs to be spent with embedded systems debuggers or in-circuit emulators since most of the development and code testing can be done on the host system of choice.

If required the cryptlib developers can provide assistance in moving the code to any new or unusual environments.

Performance

cryptlib is re-entrant and completely thread-safe, allowing it to be used with multithreaded applications under BeOS, OS/2, Windows 95/98/ME, Windows NT/2000/XP/Vista, Windows CE, and Unix systems that support threads. Because it is thread-safe, lengthy cryptlib operations can be run in the background if required while other processing is performed in the foreground. In addition cryptlib itself is multithreaded so that computationally intensive internal operations take place in the background without impacting the performance of the calling application.

Most of the core algorithms used in cryptlib have been implemented in assembly language in order to provide the maximum possible performance, and will take advantage of crypto hardware acceleration facilities present in some CPUs such as the Via C3 family. These routines provide an unprecedented level of performance, in most cases running faster than expensive, specialised encryption hardware designed to perform the same task. This means that cryptlib can be used for high-bandwidth applications such as video/audio encryption and online network and disk encryption without the need to resort to expensive, specialised encryption hardware.

Cryptographic Random Number Management

cryptlib contains an internal secure random data management system that provides the cryptographically strong random data used to generate session keys and public/private keys, in public-key encryption operations, and in various other areas that require secure random data. The random data pool is updated with unpredictable process-specific information as well as system-wide data such as current disk I/O and paging statistics, network, assorted client/server network protocol traffic, packet filter statistics, multiprocessor statistics, process information, users, VM statistics, process statistics, battery/power usage statistics, system thermal management data, open files, inodes, terminals, vector processors, streams, and loaded code, objects in the global heap, loaded modules, running threads, process, and tasks, and an equally large number of system performance-related statistics covering virtually every aspect of the operation of the system.

The exact data collected depends on the hardware and operating system, but generally includes extremely detailed and constantly changing operating statistics and information. In addition if a `/dev/random`, EGD, or PRNGD-style randomness driver (which continually accumulates random data from the system) is available, cryptlib will use this as a source of randomness. Finally, cryptlib supports a number of cryptographically strong hardware random number generators, either built into the CPU or system chipset or available as external crypto devices, that can be used to supplement the internal generator. As a post-processing stage, cryptlib employs an ANSI X9.17/X9.31 generator for additional security and for FIPS 140 compliance. This level of secure random number management ensures that security problems such as those present in Netscape's web browser (which allowed encryption keys to be predicted without breaking the encryption because the "random" data wasn't at all random) can't occur with cryptlib.

Programming Interface

The application programming interface (API) serves as an interface to a range of plug-in encryption modules that allow encryption algorithms to be added in a fairly transparent manner, so that adding a new algorithm or replacing an existing software implementation with custom encryption hardware can be done without any trouble. The standardised API allows any of the algorithms and modes supported by cryptlib to be used with a minimum of coding effort. In addition the easy-to-use high-level routines allow for the exchange of encrypted or signed messages or the establishment of secure communications channels with a minimum of programming overhead. Language bindings are available for C / C++, C# / .NET, Delphi, Java, Python, Tcl, and Visual Basic (VB).

cryptlib has been written to be as foolproof as possible. On initialisation it performs extensive self-testing against test data from encryption standards documents, and the APIs check each parameter and function call for errors before any actions are performed, with error reporting down to the level of individual parameters. In addition logical errors such as, for example, a key exchange function being called in the wrong sequence, are checked for and identified.

Documentation

cryptlib comes with extensive documentation in the form of a 310-page user manual and a 320-page technical reference manual. The user manual is intended for everyday cryptlib use and contains detailed documentation on every aspect of

cryptlib's functionality. In most cases the code needed to secure an application can be cut and pasted directly from the appropriate section of the manual, avoiding the need to learn yet another programming API. The user manual concludes with a reference section covering the various cryptlib API functions, constants, and data types.

The technical reference manual covers the design and internals of cryptlib itself, including the cryptlib security model and security mechanisms that protect every part of cryptlib's operation. In addition the technical manual provides a wealth of background information to help users understand the security foundations on which cryptlib is built.

Algorithm Support

Included as core cryptlib components are implementations of the most popular encryption and authentication algorithms, AES, Blowfish, CAST, DES, triple DES, IDEA, RC2, RC4, RC5, and Skipjack, conventional encryption, MD2, MD4, MD5, RIPEMD-160, SHA-1, and SHA-2 hash algorithms, HMAC-MD5, HMAC-SHA, and HMAC-RIPEMD-160 algorithms, and Diffie-Hellman, DSA, Elgamal, and RSA public-key encryption, with elliptic-curve encryption under development. The algorithm parameters are summarised below:

Algorithm	Key size	Block size
AES	128/192/256	128
Blowfish	448	64
CAST-128	128	64
DES	56	64
Triple DES	112 / 168	64
IDEA	128	64
RC2	1024	64
RC4	2048	8
RC5	832	64
Skipjack	80	64
MD2	—	128
MD4	—	128
MD5	—	128
RIPEMD-160	—	160
SHA-1	—	160
SHA-2 / SHA-256	—	256
HMAC-MD5	128	128
HMAC-SHA	160	160
HMAC-RIPEMD-160	160	160
Diffie-Hellman	4096	—
DSA	4096 ¹	—
Elgamal	4096	—
RSA	4096	—

Standards Compliance

All algorithms, security methods, and data encoding systems in cryptlib either comply with one or more national and international banking and security standards, or are implemented and tested to conform to a reference implementation of a particular algorithm or security system. Compliance with national and international security standards is automatically provided when cryptlib is integrated into an application. These standards include ANSI X3.92, ANSI X3.106, ANSI X9.9, ANSI X9.17, ANSI X9.30-1, ANSI X9.30-2, ANSI X9.31-1, ANSI X9.42, ANSI X9.52, ANSI X9.55, ANSI X9.57, ANSI X9.73, ETSI TS 101 733, ETSI TS 101 861, ETSI TS 101 862, ETSI TS 102, FIPS PUB 46-2, FIPS PUB 46-3, FIPS PUB 74, FIPS PUB 81, FIPS PUB 113, FIPS PUB 180, FIPS PUB 180-1, FIPS PUB 186, FIPS PUB 198, ISO/IEC

¹ The DSA standard only defines key sizes from 512 to 1024 bits, cryptlib supports longer keys but there is no extra security to be gained from using these keys.

8372, ISO/IEC 8731 ISO/IEC 8732, ISO/IEC 8824/ITU-T X.680, ISO/IEC 8825/ITU-T X.690, ISO/IEC 9797, ISO/IEC 10116, ISO/IEC 10118, ISO/IEC 15782, ITU-T X.842, ITU-T X.843, PKCS #1, PKCS #3, PKCS #5, PKCS #7, PKCS #9, PKCS #10, PKCS #11, PKCS #15, RFC 1319, RFC 1320, RFC 1321, RFC 1750, RFC 1991, RFC 2040, RFC 2104, RFC 2144, RFC 2202, RFC 2246, RFC 2268, RFC 2311 (cryptography-related portions), RFC 2312, RFC 2313, RFC 2314, RFC 2315, RFC 2437, RFC 2440, RFC 2459, RFC 2510, RFC 2511, RFC 2528, RFC 2560, RFC 2585, RFC 2630, RFC 2631, RFC 2632, RFC 2633 (cryptography-related portions), RFC 2634, RFC 2785, RFC 2876, RFC 2898, RFC 2984, RFC 2985, RFC 2986, RFC 3039, RFC 3058, RFC 3114, RFC 3126, RFC 3161, RFC 3174, RFC 3183, RFC 3211, RFC 3218, RFC 3261 (cryptography-related portions), RFC 3268, RFC 3274, RFC 3279, RFC 3280, RFC 3281, RFC 3369, RFC 3370, RFC 3447, RFC 3546, RFC 3565, RFC 3739, RFC 3770, RFC 3851, RFC 3852, RFC 4055, RFC 4086, RFC 4108, RFC 4134, RFC 4210, RFC 4211, RFC 4231, RFC 4250, RFC 4251, RFC 4252, RFC 4253, RFC 4254, RFC 4256, RFC 4262, RFC 4279, RFC 4325, RFC 4334, RFC 4346, RFC 4366, RFC 4387, RFC 4419, RFC 4476, RFC 4648, RFC 4680, RFC 4681, and the Payment Card Industry (PCI) Data Security Standard (cryptography-related portions). Because of the use of internationally recognised and standardised security algorithms, cryptlib users will avoid the problems caused by home-grown, proprietary algorithms and security techniques that often fail to provide any protection against attackers, resulting in embarrassing bad publicity and expensive product recalls.

Y2K Compliance

cryptlib handles all date information using the ANSI/ISO C time format, which does not suffer from Y2K problems. Although earlier versions of the X.509 certificate format do have Y2K problems, cryptlib transparently converts the dates encoded in certificates to and from the ANSI/ISO format, so cryptlib users will never see this. cryptlib's own time/date format is not affected by any Y2K problems, and cryptlib itself conforms to the requirements in the British Standards Institution's DISC PD2000-1:1998 Y2K compliance standard.

Configuration Options

cryptlib works with a configuration database that can be used to tune its operation for different environments. This allows a system administrator to set a consistent security policy which is then automatically applied by cryptlib to operations such as key generation and data encryption and signing, although they can be overridden on a per-application or per-user basis if required.

cryptlib Applications

The security services provided by cryptlib can be used in virtually any situation that requires the protection or authentication of sensitive data. Some areas in which cryptlib is currently used include:

- Protection of medical records transmitted over electronic links.
- Protection of financial information transmitted between branches of banks.
- Transparent disk encryption.
- Strong security services added to web browsers with weak, exportable security.
- Running a CA.
- Encrypted electronic mail.
- File encryption.
- Protecting content on Internet servers.
- Digitally signed electronic forms.
- S/MIME mail gateway.

- Secure database access.
- Protection of credit card information.

Encryption Code Example

The best way to illustrate what cryptlib can do is with an example. The following code encrypts a message using public-key encryption.

```
/* Create an envelope for the message */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_SMIME );

/* Push in the message recipient's name */
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_RECIPIENT,
    recipientName, recipientNameLength );

/* Push in the message data and pop out the signed and encrypted
   result */
cryptPushData( cryptEnvelope, message, messageSize, &bytesIn );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, encryptedMessage, encryptedSize,
    &bytesOut );

/* Clean up */
cryptDestroyEnvelope( cryptEnvelope );
```

This performs the same task as a program like PGP using just 6 function calls (to create a PGP/OpenPGP message, just change the `CRYPT_FORMAT_SMIME` to `CRYPT_FORMAT_PGP`). All data management is handled automatically by cryptlib, so there's no need to worry about encryption modes and algorithms and key lengths and key types and initialisation vectors and other details (although cryptlib provides the ability to specify all this if you feel the need).

The code shown above results in cryptlib performing the following actions:

- Generate a random session key for the default encryption algorithm (usually triple DES or AES).
- Look up the recipient's public key in a key database.
- Encrypt the session key using the recipient's public key.
- Encrypt the signed data with the session key.
- Pass the result back to the user.

However unless you want to call cryptlib using the low-level interface, you never need to know about any of this. cryptlib will automatically know what to do with the data based on the resources you add to the envelope — if you add a signature key it will sign the data, if you add an encryption key it will encrypt the data, and so on.

Secure Session Code Example

Establishing a secure session using SSL/TLS is similarly easy:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSL );

/* Add the server name and activate the session */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    serverName, serverNameLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

If you prefer SSH to SSL, just change the `CRYPT_SESSION_SSL` to `CRYPT_SESSION_SSH` and add a user name and password to log on. As with the encryption code example above, cryptlib provides a single unified interface to its secure session mechanisms, so you don't have to invest a lot of effort in adding special-case handling for different security protocols and mechanisms.

The corresponding SSL/TLS (or SSH if you prefer) server is:

```

CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSL_SERVER
);

/* Add the server key/certificate and activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY, privateKey
);
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );

```

As with the secure enveloping example, cryptlib is performing a large amount of work in the background, but again there's no need to know about this since it's all taken care of automatically.

Certificate Management Code Example

The following code illustrates cryptlib's plug-and-play PKI interface:

```

CRYPT_SESSION cryptSession;

/* Create the CMP session and add the server name/address */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_CMP );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER, server,
serverLength );

/* Add the username, password, and smart card */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
userName, userNameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
password, passwordLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_CMP_PRIVKEYSET,
cryptDevice );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, TRUE );

```

This code takes a smart card and generates separate encryption and signing keys in it, requests a signature certificate from the CA for the signing key, uses that to obtain a certificate for the encryption key, obtains any further certificates that may be needed from the CA (for example for S/MIME signing or SSL server operation), and stores everything in the smart card. Compare this to the hundreds or even thousands of lines of code required to do the same thing using other toolkits.

Oh yes, and cryptlib provides the CA-side functionality as well — there's no need to pay an expensive commercial CA for your certificates, since cryptlib can perform the same function.

Document conventions

This manual uses the following document conventions:

Example	Description
<code>cryptlib.h</code>	This font is used for filenames.
cryptCreateContext	Bold type indicates cryptlib function names.
<i>Value</i>	Words or portions of words in italics indicate placeholders for information that you need to supply.
<code>if(i == 0)</code>	This font is used for sample code and operating system commands.

Recommended Reading

One of the best books to help you understand how to use cryptlib is *Network Security* by Charlie Kaufman, Radia Perlman, and Mike Speciner, which covers general security principles, encryption techniques, and a number of potential cryptlib applications such as X.400/X.500 security, PEM/S/MIME/PGP, Kerberos, and various other security, authentication, and encryption techniques. The book also

contains a wealth of practical advice for anyone considering implementing a cryptographic security system. *Security Engineering: A Guide to Building Dependable Distributed Systems* by Ross Anderson also contains a large amount of useful information and advice on engineering secure systems. *Building Secure Software* by John Viega and Gary McGraw and *Writing Secure Software* by Michael Howard and David LeBlanc contain a wealth of information on safe programming techniques and related security issues.

Cryptographic Security Architecture Design and Verification by Peter Gutmann is the technical documentation for cryptlib and complements the cryptlib user manual. It contains full details of the architectural and security features of cryptlib, as well as a wealth of background material to help you understand the security foundations on which cryptlib is built.

A tutorial in 8 parts totalling over 700 slides and covering all aspects of encryption and general network security, including encryption and security basics, algorithms, key management and certificates, CAs, certificate profiles and policies, PEM, PGP, S/MIME, SSL, SSH, SET, smart cards, and a wide variety of related topics, is available from <http://www.cs.auckland.ac.nz/~pgut001/tutorial/>. If you want to do anything with certificates, you should definitely read *Everything you Never Wanted to Know about PKI but were Forced to Find Out*, available from <http://www.cs.auckland.ac.nz/~pgut001/pubs/-pkitutorial.pdf>, to find out what you're in for if you have to work with certificates.

In addition to this, there are a number of excellent books available that will help you in understanding the cryptography used in cryptlib. The foremost of these are *Applied Cryptography* by Bruce Schneier and the *Handbook of Applied Cryptography* by Alfred Menezes, Paul van Oorschot, and Scott Vanstone. *Applied Cryptography* provides an easy-to-read overview while the *Handbook of Applied Cryptography* provides extremely comprehensive, in-depth coverage of the field.

For general coverage of computer security issues, *Security in Computing* by Charles Pfleeger provides a good overview of security, access control, and secure operating systems and databases, and also goes into a number of other areas such as ethical issues that aren't covered by most books on computer security. *Computer Security: Art and Science* by Matt Bishop provides in-depth coverage of all aspects of computer security modelling and design, with a particular emphasis on access control and security models and high-assurance systems.

Installation

This chapter describes how to install cryptlib for a variety of operating systems.

AMX

The AMX Multitasking Executive is a real-time OS (RTOS) with development hosted under Unix or Windows. You can build cryptlib for AMX using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for AMX is `target-amx`, so you'd build cryptlib with `make target-amx`. Details on building and using cryptlib for AMX, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

BeOS

The BeOS version of cryptlib can be built using a procedure which is identical to that given for Unix on page 20. Any current version of BeOS can build the code directly from the Unix makefile. Old versions of BeOS using the Be development environment will require that you edit the Unix makefile slightly by un-commenting the marked lines at the start of the file.

ChorusOS

ChorusOS is an embedded OS with development hosted under Unix. You can build cryptlib for ChorusOS using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for ChorusOS is `target-chorus`, so you'd build cryptlib with `make target-chorus`. Details on building and using cryptlib for ChorusOS, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

DOS

The 16-bit DOS version of cryptlib can be built from the same files as the 16-bit Windows version, so no separate makefile is provided. Because DOS is so limited in its capabilities, it is in effect an embedded systems OS. Details on building and using cryptlib for DOS, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

DOS32

The 32-bit DOS version of cryptlib can be built using the supplied makefile, which requires the `djgpp` compiler. The DOS32 version of cryptlib uses the same 32-bit assembly language code used by the Win32 and 80x86 Unix versions, so it runs significantly faster than the 16-bit DOS version. Like the 16-bit DOS version, any attempt to use the high-level key export routines will fail with a `CRYPT_ERROR_RANDOM` error code unless a `/dev/random`-style driver is available because there isn't any way to reliably obtain random data under DOS. You can however treat DOS as an embedded systems environment and use the random seeding capability described in "Porting to Devices without Randomness/Entropy Sources" on page 379.

eCOS

eCOS is an embedded/real-time OS (RTOS) with development hosted under Unix or Windows. You can build cryptlib for eCOS using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for eCOS is `target-ecos`, so you'd build cryptlib with `make target-ecos`. Details on building and using cryptlib for eCOS, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

μC/OS-II

μC/OS-II is an embedded/real-time OS (RTOS) with development usually hosted under Windows. You can build cryptlib for μC/OS-II using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for μC/OS-II is `target-ucos`, so you'd build cryptlib with **make target-ucos**. Details on building and using cryptlib for μC/OS-II, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

Embedded Linux

The embedded Linux version of cryptlib can be built using the standard Linux development tools. Since this environment is identical to the generic Unix one, the installation instructions for Unix on page 20 apply here.

μITRON

μITRON is an embedded/real-time OS (RTOS) with development usually hosted under Unix or a Unix-like OS. You can build cryptlib for μITRON using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for μITRON is `target-itron`, so you'd build cryptlib with **make target-itron**. Details on building and using cryptlib for μITRON, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

Macintosh OS X

The standard Macintosh build environment uses Apple's Mac OS X Developer Tools, driven by the standard makefile, for which the instructions in the section on building cryptlib for Unix on page 20 apply. Alternatively, you can build cryptlib using Metroworks' Codewarrior with the `Mac.mcp` project file. This can build cryptlib either as a static or shared library for both 68K and PowerPC Macs, although since this isn't the primary build environment the project file may apply to a slightly older cryptlib release and require a little updating to match the current configuration (the standard makefile will always be current). In addition it's possible to build it using Apple's free MrC compiler, with the same caveat about updating of configuration files.

MVS

The MVS version of cryptlib can be built using the standard IBM C/C++ compiler and accompanying tools. Since this environment is very similar to the Unix one, the installation instructions for Unix on page 20 apply here also. Note that PTF UQ50384 (which fixes a bug in the macro version of the `strcat` function as described in APAR PQ43130) is required if you're using the V2R10 C/C++ compiler.

You can control the use of ddnames with the `DDNAME_IO` define. If `DDNAME_IO` is defined when building the code, cryptlib will use ddnames for all I/O, and user options will be saved in dynamically allocated datasets `userid.CRYPTLIB.filename`. If `DDNAME_IO` is not defined when building the code, cryptlib will use HFS for all I/O, and user options will be saved in `$HOME/.cryptlib`.

After you've built cryptlib, you should run the self-test program to make sure that everything is working OK. You can use the `ussalloc` USS shell script to allocate MVS data sets for testlib, and the `usscopy` shell script to copy the files in the test directory to the MVS data sets allocated with `ussalloc`. `testlib.jcl` is the JCL needed to execute testlib.

OS2

The OS/2 version of cryptlib can be built using the command-line version of the IBM compiler. The supplied makefile will build the DLL version of cryptlib, and can also build the cryptlib self-test program, which is a console application. You should run

the self-test program after you've built cryptlib to make sure that everything is working OK.

If you're using the IBM OS/2 compiler you should set enumerated types to always be 32-bit values because the compiler by default uses variable-length types depending on the enum range (so one enum could be an 8-bit type and another 32). cryptlib is immune to this "feature", and function calls from your code to cryptlib should also be unaffected because of type promotion to 32-bit integers, but the variable-range enums may cause problems in your code if you try to work with them under the assumption that they have a fixed type.

PalmOS

PalmOS is the operating system for the Palm series of PDAs, with development hosted under Unix or Windows. You can build cryptlib for PalmOS using the PalmOS 6 SDK and the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for the PalmOS SDK is `target-palmos` and for the alternative PRC development tools is `target-palmos-prc`, so you'd build cryptlib with **make target-palmos** or **make target-palmos-prc**. Details on building and using cryptlib for PalmOS, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

QNX Neutrino

The QNX Neutrino version of cryptlib can be built using the standard QNX development tools. Since this environment is identical to the generic Unix one, the installation instructions for Unix on page 20 apply here.

RTEMS

The Real-Time Operating System for Multiprocessor Systems (RTEMS) is a real-time OS (RTOS) with development hosted under Unix or Windows. You can build cryptlib for RTEMS using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for RTEMS is `target-rtems`, so you'd build cryptlib with **make target-rtems**. Details on building and using cryptlib for RTEMS, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

Tandem

The Tandem version of cryptlib can be built using the standard c89 compiler and accompanying tools under the OSS environment. Since this environment is very similar to the Unix one, the installation instructions for Unix on page 20 apply here also. The default target is Tandem OSS, you can re-target the built for NSK using the `-Wstype=guardian` directive in the makefile.

The Guardian sockets implementation changed in newer releases of the OS. Older releases required the use of non-standard `nowait` sockets handled via `AWAITIOX()` instead of the standard BSD sockets interface. If you're running an older version of the OS and need to use any of the secure networking protocols such as SSL/TLS, SSH, CMP, SCEP, RTCS, or OCSP, you'll need to use cryptlib's alternative network data-handling strategy described in "Network Issues" on page 212.

uClinux

uClinux is a real-mode/embedded version of Linux with development hosted under Unix. You can build cryptlib for uClinux using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for uClinux is `target-uclinux`, so you'd build cryptlib with **make target-uclinux**. Details on building and using cryptlib for uClinux, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

Unix

To unzip the code under Unix use the `-a` option to ensure that the text files are converted to the Unix format. The makefile by default will build the statically-linked library when you invoke it with `make`. To build the shared library, use `make shared`. Once cryptlib has been built, use `make testlib` to build the cryptlib self-test program `testlib`, or `make stestlib` to build the shared-library self-test program `stestlib`. This will run fairly extensive self-tests of cryptlib that you can run after you've built it to make sure that everything is working OK. `testlib` needs to be run from the cryptlib root directory (the one that the main data files are in) since it uses a large number of pre-generated data files that are located in a subdirectory below this one. Depending on your system setup and privileges you may need to either copy the shared library to `/usr/lib` or set the `LD_LIBRARY_PATH` environment variable (or an OS-specific equivalent) to make sure that the shared library is used.

If you're using the statically-linked form of cryptlib in your application rather than the shared library, you'll probably need to link in additional (system-specific) static libraries to handle threads, network access, and system-specific odds and ends. The makefile contains a list of the needed additional libraries, ordered by system type and version. The shared-library version of cryptlib doesn't require these additional libraries to be linked in, since the references are automatically resolved by the OS.

If your system doesn't come pre-configured with a `/dev/random`, EGD, or PRNGD-style randomness driver (which continually accumulates random data from the system), you may want to download one and install it, since cryptlib will make use of it for gathering entropy. cryptlib has a built-in randomness polling subsystem so it will function without an external randomness driver, but it never hurts to have one present to supplement the internal entropy polling.

If you're using a key database or certificate store, you need to enable the use of the appropriate interface module for the database backend. Details are given in "Key Database Setup" on page 23. For the cryptlib self-test code you can define the database libraries using the `TESTLIBS` setting at the start of the makefile. If you don't enable the use of a database interface, the self-test code will issue a warning that no key database is present and continue without testing the database interface.

If you're using an LDAP directory, you need to install the required LDAP client library on your system, enable the use of LDAP using the `USE_LDAP` define before you build cryptlib, and link the LDAP client library into your executable (on most systems the cryptlib build scripts will take care of this automatically). If you don't enable the use of an LDAP directory interface, the self-test code will issue a warning that no LDAP directory interface is present and continue without testing the LDAP interface.

If you're using special encryption hardware or an external encryption device such as a PCMCIA card or smart card, you need to install the required device drivers on your system and enable their use when you build cryptlib by linking in the required interface libraries. If you don't enable the use of a crypto device, the self-test code will issue a warning that no devices are present and continue without testing the crypto device interface.

If your application forks, you shouldn't need to take any special actions for cryptlib beyond the usual precautions with forking a process. In particular forking a process that contains multiple threads has system-specific semantics, with the behaviour depending on whether the system implements *fork1* or *forkall* behaviour. With *fork1* behaviour (the Posix default), only the thread that calls `fork()` is copied to the child. With *forkall*, all threads in the process are copied. The *fork1* behaviour can lead to deadlock if a thread other than the one that called `fork()` holds a lock, since the fact that it's not copied to the child means that it'll never be released. You can work around this with `pthread_atfork()` to handle lock management, but a better approach is to simply not mix threads and forking unless you follow the `fork()` with

an `exec()`. Note that this isn't a cryptlib issue, it's specific to the interaction of `fork()` and threads.

For any common Unix system, cryptlib will build without any problems, but in some rare cases you may need to edit `random/unix.c` and possibly `io/file.h` and `io/tcp.h` if you're running an unusual Unix variant that puts include files in strange places or has broken Posix or sockets support.

VM/CMS

The VM/CMS version of cryptlib can be built using the standard C/370 compiler and accompanying tools. The supplied EXEC2 file `VMBUILD EXEC` will build cryptlib as a `TXTLIB` and then build the self-test program as an executable `MODULE` file. Since VM sites typically have different system configurations, this file and possibly portions of the source code may require tuning in order to adjust it to suit the build process normally used at your site.

VxWorks

VxWorks is an embedded/real-time OS (RTOS) with development hosted under Unix or Windows. You can build cryptlib for VxWorks using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for VxWorks is `target-vxworks`, so you'd build cryptlib with `make target-vxworks`. Details on building and using cryptlib for VxWorks, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

Windows 3.x

The 16-bit cryptlib DLL can be built using the `crypt16.mak` makefile, which is for version 1.5x of the Visual C++ compiler. The mixed C/assembly language encryption and hashing code will give a number of warnings, the remaining code should compile without warnings. Once the DLL has been built, `test.mak` will build the cryptlib self-test program, which is a console application. You can run this after you've built cryptlib to make sure that everything is working OK.

If you're using a key database or certificate store, you need to set up an ODBC data source for this. Details are given in "Key Database Setup" on page 23.

Windows 95/98/ME and Windows NT/2000/XP/Vista

The 32-bit cryptlib DLL can be built using the `crypt32` project file, which is for Visual C++ 6 and Visual C++ .NET. Once the DLL has been built, the `test32` project file will build the cryptlib self-test program `test32`, which is a console application. You can run this after you've built cryptlib to make sure that everything is working OK. `test32` needs to be run from the cryptlib root directory (the one that the main data files are in) since it uses a large number of pre-generated data files that are located in a subdirectory below this one. If you'll be using the cryptlib user interface components you need to install the cryptlib user interface library `cl32ui.dll` alongside cryptlib itself.

If you're using an older version of Visual C++ .NET, a bug in its version 6 project file import process results in files having the `$(NoInherit)` property set, so that a define made at the project level won't be passed down to other files. If you want to enable options based on global defines, you need to disable this property before the defines will propagate down to other files.

If you're using a key database or certificate store, you need to set up an ODBC data source for this. Details are given in "Key Database Setup" on page 23.

If you're using special encryption hardware or an external encryption device such as a PCMCIA card or smart card, you need to install the required device drivers on your system, and if you're using a generic PKCS #11 device you need to configure the appropriate driver for it as described in "Encryption Devices and Modules" on page 350. cryptlib will automatically detect and use any devices that it recognises and that

have drivers present. If you don't enable the use of a crypto device, the self-test code will issue a warning that no devices are present and continue without testing the crypto device interface.

Personal firewall products from some vendors can interfere with network operations for devices other than standard web browsers and mail clients. If you're experiencing odd behaviour when using cryptlib for network operations (for example you can connect but can't exchange data, or you get strange error messages when you connect), you can try temporarily disabling the personal firewall to see if this fixes the problem. If it does, you should contact the personal firewall vendor to fix their product, or switch to a different product.

If you're using Borland C++ rather than Visual C++, you'll need to set up the `.def` and `.lib` files for use with the Borland compiler. To do this, run the following commands in the cryptlib directory:

```
impdef cl32 cl32
implib cl32 cl32.def
```

The first one will produce a Borland-specific `.def` file from the DLL, the second one will produce a Borland-specific `.lib` file from the DLL and `.def` file.

To install the ActiveX control, put the cryptlib DLL and the ActiveX wrapper `clcom.dll` into the Windows system directory and register the ActiveX wrapper with:

```
regsvr32 clcom.dll
```

To use the ActiveX control with Visual Basic, use Project | Reference to add `clcom.dll`, after which VB will recognise the presence of the ActiveX wrapper.

Windows CE / Pocket PC / SmartPhone

The 32-bit cryptlib DLL for Windows CE/PocketPC/SmartPhone can be built using the `crypt32ce` project file, which is for version 3 or 4 of the eMbedded Visual C++ compiler. Once the DLL has been built, the `test32ce` project file will build the cryptlib self-test program `test32ce`, which is a (pseudo-)console application that produces its output on the debug console. You can run this after you've built cryptlib to make sure that everything is working OK. `test32ce` needs to be run from the cryptlib root directory (the one that the main data files are in) since it uses a large number of pre-generated data files that are located in a subdirectory below this one.

The cryptlib Windows CE self-test uses the 'Storage Card' pseudo-folder to access the files needed for the self-test. Depending on the system setup, you need to either copy the files to the storage card or (the easier alternative) use folder sharing to access the directory containing the test files. From the Windows CE menu, select Folder Sharing and share the `testdata` subdirectory, which will appear as `\\Storage Card\` on the Windows CE device.

Windows CE is a Unicode environment, which means that all text strings are passed to and from cryptlib as Unicode strings. For simplicity the examples in this manual are presented using the standard `char` data type used on most systems, however under Windows CE all character types and strings are Unicode in line with standard Windows CE practice. When you're using the examples, you should treat any occurrence of characters and strings as standard Unicode data types.

A few older versions of eVC++ for some platforms don't include the ANSI/ISO C standard `time.h` header, which is a required file for a conforming ANSI/ISO C compiler. If you have a version of eVC++ that doesn't include this standard header, you need to add it from another source, for example an eVC++ distribution that does include it or the standard (non-embedded) VC++ distribution.

When compiling cryptlib under eVC++ 4.0 for the Arm architecture with optimisation enabled, a compiler bug may prevent three files from compiling. If you get an internal compiler error trying to compile `context/kg_rsa.c`, `crypt/rc2skey.c`, or `misc/base64.c`, you can work around the problem by disabling optimisation using `#pragma optimize("g", off)` / `#pragma optimize("g", on)`

around the functions `initCheckRSAkey()` in `kg_rsa.c`, `RC2_set_key()` in `rc2skey.c`, and `adjustPKIUserValue()` in `base64.c`.

Xilinx XMK

The Xilinx Microkernel (XMK) is a real-time OS (RTOS) with development hosted under Unix or Windows. You can build cryptlib for XMK using the cross-compilation capabilities of the standard makefile, see the entry for Unix on page 20 for more details on working with the makefile. The make target for XMK is `target-xmk-mb` for the MicroBlaze core and `target-xmk-ppc` for the PowerPC core, so you'd build cryptlib with `make target-xmk-mb` or `make target-xmk-ppc`. Details on building and using cryptlib for XMK, and on embedded cryptlib in general, are given in "Embedded Systems" on page 372.

Other Systems

cryptlib should be fairly portable to other systems, the only part that needs special attention is the randomness-gathering in `random/os_name.c` (cryptlib won't work without this, the code will produce a link error). The idea behind the randomness-gathering code is to perform a comprehensive poll of every possible entropy source in the system in a separate thread or background task ("slowPoll"), as well as providing a less useful but much faster poll of quick-response sources ("fastPoll"). In addition the filesystem I/O code in `io/file.c` may need system-specific code and definitions added to it if the system you're running on doesn't use a standard form of file I/O, for example a system that has its own file I/O layer that isn't compatible with standard models or one that doesn't have file I/O at all such as an embedded device that uses flash memory for storage.

To find out what to compile, look at the Unix makefile, which contains all of the necessary source files (the `group_name_OBJS` dependencies) and compiler options. Link all of these into a library (as the makefile does) and then compile and link the modules in the `test` subdirectory with the library to create the self-test program. There is additional assembly-language code included that will lead to noticeable speedups on some systems, you should modify your build options as appropriate to use these if possible.

Depending on your compiler you may get a few warnings about some of the encryption and hashing code (one or two) and the bignum code (one or two). This code mostly relates to the use of C as a high-level assembler and changing things around to remove the warnings on one system could cause the code to break on another system.

Key Database Setup

If you want to work with a key database or certificate store, you need to configure a database for cryptlib to use. Under Windows, go to the Control Panel and click on the ODBC/ODBC32 item. Click on "Add" and select the ODBC data source (that is, the database type) that you want to use. If it's on the local machine, this will probably be an Access database, if it's a centralised database on a network this will probably be SQL Server. Once you've selected the data source type, you need to give it a name for cryptlib to use. "Public Keys" is a good choice (the self-test code uses two sources called `testkeys` and `testcertstore` during the self-test procedure, and will create these itself if possible). In addition you may need to set up other parameters like the server that the database is located on and other access information. Once the data source is set up, you can access it as a `CRYPT_-KEYSET_ODBC` keyset using the name that you've assigned to it.

Under Unix or similar systems the best way to work with a key database or certificate store is to use the ODBC interface, either via a layered driver such as `unixODBC` or `iODBC`, or directly via interfaces such as `MyODBC`. Alternatively, you can use the cryptlib generic database interface to compile database-specific support code directly into cryptlib, or the database network plugin capability to make a network connection to a database server such as IBM DB2, Informix, Ingres, Oracle, Postgres, or Sybase.

The easiest interface to use is the ODBC one, which hides all of the low-level database interface details. The ODBC configuration process follows the same pattern as the one given above for ODBC under Windows, with OS-specific variations depending on the platform that you're running it under. You can enable the use of the ODBC interface using the `USE_ODBC` define before you build cryptlib, and if you're not using Windows (which uses dynamic binding to the ODBC interface) you need to link the ODBC client library into your executable (on most systems the cryptlib build scripts will take care of this automatically).

For Unix and Unix-like systems the two most common ODBC implementations are unixODBC and iODBC, although a variety of other products are also available, and some databases have native ODBC support, examples being MySQL (via MyODBC) and IBM DB2. These interfaces support a wide range of commercial database including AdabasD, IBM DB2, Informix, Ingres, Interbase, MySQL, Oracle, Postgres, and Sybase. unixODBC uses the `ODBCConfig` GUI application to configure data sources and drivers in a manner identical to the standard Windows interface, and also provides the `odbcinst` CLI utility to configure data sources and drivers. `odbcinst` can be used to automatically install and configure database drivers for ODBC using template files that contain information about the driver such as the location of the driver binaries, usually somewhere under `/usr/local`. For example to configure the Oracle drivers for ODBC using a prepared template file you'd use:

```
odbcinst -i -d -f oracle.tmpl
```

iODBC provides drivers as platform-specific binaries that are installed using the iODBC installation shell scripts. See the documentation for the particular ODBC interface that you're using for more information on installation and configuration issues.

If you don't want to use the ODBC interface, you can either compile database-specific interface code directly into cryptlib or use the database network plugin capability to make a network connection to a database server. To use cryptlib's generic database interface you need to define `USE_DATABASE` when you build cryptlib and add the appropriate interface code to communicate with the database back-end of your choice, as described in "Database and Networking Plugins" on page 380. In addition you need to link the database library or libraries (for example `libmysql.a`) into your executable.

To use the database plugin capability to make a network connection to a database server such as Informix, Ingres, Oracle, Postgres, or Sybase, you need to create the appropriate plugin for your server as described in "Database and Networking Plugins" on page 380.

If you need to use a database keyset on an embedded system, you can use a system like the SQLite embedded database engine, <http://sqlite.org/>. SQLite is a self-contained, embeddable, zero-configuration SQL database engine that provides all of the capabilities needed by cryptlib database keysets.

Configuration Issues

For compatibility with existing deployed code, cryptlib supports a wide variety of encryption, signature, and hash algorithms, key types, and security mechanisms. Some of these backwards-compatible items are obsolete, unsound, or even entirely broken. For this reason the encryption algorithms RC2, RC4, and Skipjack, the hash algorithms MD2 and MD4, and the SSHv1 protocol, are disabled by default. If you want to enable these obsolete and insecure items, you can do so via the cryptlib configuration file `misc/config.h`. Note that by enabling these unsafe items, you are voiding cryptlib's security guarantees and agree to indemnify the cryptlib authors against any claims or losses from any problems that may arise. In other words you really, really shouldn't do this.

cryptlib also contains two algorithms, IDEA and RC5, that may be covered by patents in some countries. If you're unsure over whether you can use the algorithms, you should disable them as described below. Note that disabling IDEA will remove the

ability to read PGP 2 keys and messages, since this version requires the use of the IDEA algorithm for en/decryption of data.

Customised and Cut-down cryptlib Versions

In some cases you may want to customise the cryptlib build or create a cut-down version that omits certain capabilities in order to reduce code size for constrained environments. You can do this by editing the configuration build file `misc/config.h`, which allows almost every part of cryptlib's functionality to be selectively enabled or disabled (some functionality is used by all of cryptlib and can't be disabled). Each portion of functionality is controlled by a `USE_name` define, by undefining the value before you build cryptlib the named functionality will be removed. For example, undefining `USE_SSH1` would disable the use of SSHv1 (this is disabled by default, since it's been superseded by SSHv2); undefining `USE_SKIPJACK` would disable the use of the Skipjack algorithm (this is also disabled by default, since it's obsolete and no longer considered secure). In addition you can use the build file to disable the use of the two patented algorithms IDEA and RC5 (see "Algorithms" on page 387 for more information on whether these two patents affect your use of cryptlib) by undefining `USE_PATENTED_ALGORITHMS`. More details on tuning cryptlib's size and capabilities (particularly for use in embedded systems) is given in "Embedded Systems" on page 372.

Debug vs. Release Versions of cryptlib

cryptlib can be built in one of two forms, a debug version and a release version. The main difference between the two is that the release version is built with the `NDEBUG` value defined, which disables the large number of internal consistency checks that are present in the debug build of cryptlib. These consistency checks are used to catch conditions such as inappropriate error codes being returned from internal functions, invalid data values being passed to functions inside cryptlib, configuration errors, and general sanity checks that ensure that everything is operating as it should. If one of these internal checks is triggered, cryptlib will throw an exception and display an error message indicating that an assertion in the code has failed. These assertions are useful for tracking down areas of code that may need revision in later releases.

If you don't want to see these diagnostic messages, you should build cryptlib with the `NDEBUG` value defined (this is the default under Unix and is done automatically under Windows when you build a release version of the code with Visual C++). Building a version in this manner will disable the extra consistency checks that are present in the debug version so that, for example, error conditions will be indicated by cryptlib returning an error code for a function call rather than throwing an exception. This will have the slight downside that it'll make tracking the exact location of a problem a bit more complex, since the error code which is returned probably won't be checked until the flow of execution has progressed a long way from where the problem was detected. On the other hand the release version of the code is significantly smaller than the debug version.

As always, if you're working with a debug build of the code and perform an operation that triggers an internal consistency check you should report the details and the code necessary to recreate it to the cryptlib developers in order to allow the exception condition to be analysed and corrected.

cryptlib Version Information

cryptlib uses 3-digit version numbers, available at runtime through the configuration options `CRYPT_OPTION_INFO_MAJORVERSION`, `CRYPT_OPTION_INFO_MINORVERSION`, and `CRYPT_OPTION_INFO_STEPPING`, and at compile time through the define `CRYPTLIB_VERSION`. `CRYPTLIB_VERSION` contains the current version as a 3-digit decimal value with the first digit being the major version number (currently 3), the second digit being the minor version number, and the third digit being the update or stepping number. For example, cryptlib version 3.2.1 would have a `CRYPTLIB_VERSION` value of 321.

All cryptlib releases with the same stepping version number are binary-compatible. This means that if you move from (for example) cryptlib version 3.2.1 to 3.2.2, all you need to do is replace the cryptlib DLL or shared library to take advantage of new cryptlib features and updates. All cryptlib releases with the same minor version number are source-compatible, so that if you move from (for example) 3.2.1 to 3.3.5, you need to recompile your application to match new features in cryptlib.

Support for Vendor-specific Algorithms

cryptlib supports the use of vendor-specific algorithm types with the predefined values `CRYPT_ALGO_VENDOR1`, `CRYPT_ALGO_VENDOR2`, and `CRYPT_ALGO_VENDOR3`. For each of the algorithms you use, you need to add a call to initialise the algorithm capability information to `device/system.c` alongside the existing algorithm initialisation, and then provide your implementation of the algorithm to compile and link into cryptlib. When you rebuild cryptlib with the preprocessor define `USE_VENDOR_ALGOS` set, the new algorithm types will be included in cryptlib's capabilities.

For example if you wanted to add support for the Foo256 cipher to cryptlib you would create the file `vendalgo.c` containing the capability definitions and then rebuild cryptlib with `USE_VENDOR_ALGOS` defined. The Foo256 algorithm would then become available as algorithm type `CRYPT_ALGO_VENDOR1`.

cryptlib Basics

cryptlib works with two classes of objects, container objects and action objects. A container object is an object that contains one or more items such as data, keys or certificates. An action object is an object which is used to perform an action such as encrypting or signing data. The container types used in cryptlib are envelopes (for data), sessions (for communications sessions), keysets (for keys), and certificates (for attributes such as key usage restrictions and signature information). Container objects can have items such as data or public/private keys placed in them and retrieved from them. In addition to containing data or keys, container objects can also contain other objects that affect the behaviour of the container object. For example pushing an encryption object into an envelope container object will result in all data which is pushed into the envelope being encrypted or decrypted using the encryption object.

Encryption contexts are the action objects used by cryptlib. Action objects are used to act on data, for example to encrypt or decrypt a piece of data or to digitally sign or check the signature on a piece of data.

The usual mechanism for processing data is to use an envelope or session container object. The process of pushing data into an envelope and popping the processed data back out is known as enveloping the data. The reverse process is known as de-enveloping the data. Session objects work in a similar manner, but are used to encapsulate a secure session with a remote client or server rather than a local data transformation. The first section of this manual covers the basics of enveloping data, which introduces the enveloping mechanism and covers various aspects of the enveloping process such as processing data streams of unknown length and handling errors. Once you have the code to perform basic enveloping in place, you can add extra functionality such as password-based data encryption to the processing. After the basic concepts behind enveloping have been explained, more advanced techniques such as public-key based enveloping and digital signature enveloping for S/MIME and PGP are covered.

Session objects are very similar to envelope objects except that they represent a communications session with a remote client or server. The next section covers the use of session objects for protocols such as SSL, TLS, and SSH to secure communications or work with protocols such as CMS, SCEP, RTCS, OCSP, and TSP that handle functions such as certificate status information and timestamping.

The use of public keys for enveloping requires the use of key management functions, and the next section covers key generation and storing and retrieving keys from keyset objects and crypto devices. The public portions of public/private key pairs are typically managed using X.509 certificates and certificate revocation lists. The next sections cover the management of certificates including certificate issue, certificate status checking, and certificate revocation list (CRL) creation and checking, as well as the full CA management process. This covers the full key life cycle from creation through certification to revocation and/or destruction.

So far all of the objects that have been covered are high-level container objects. The next section covers the creation of low-level action objects that you can either push into a container object or apply directly to data, including the various ways of loading or generating keys into them. The next sections explain how to apply the action objects to data and cover the process of encryption, key exchange, and signature generation and verification, working at a much lower level than the enveloping or session interface.

The next sections cover certificates and certificate-like objects in more detail than the earlier ones, covering such topics as DN structures, certificate chains, trust management, and certificate extensions. This deals with certificates at a very low level at which they're rather harder to manage than with the high-level certificate management functions.

The next section covers the use of encryption devices such as smart cards, crypto devices, HSMs, and Fortezza cards, and explains how to use them to perform many of the tasks covered in previous sections. Finally, the last sections cover miscellaneous topics such as random number management, the cryptlib configuration database, key database and network plugins, and use in embedded systems.

Programming Interfaces

cryptlib provides three levels of interface, of which the highest-level one is the easiest to use and therefore the recommended one. At this level cryptlib works with envelope and session container objects, an abstract object into which you can insert and remove data which is processed as required while it is in the object. Using envelopes and session objects requires no knowledge of encryption or digital signature techniques. At an intermediate level, cryptlib works with encryption action objects, and requires some knowledge of encryption techniques. In addition you will need to handle some of the management of the encryption objects yourself. At the very lowest level cryptlib works directly with the encryption action objects and requires you to know about algorithm details and key and data management methods.

Before you begin you should decide which interface you want to use, as each one has its own distinct advantages and disadvantages. The three interfaces are:

High-level Interface

This interface requires no knowledge of encryption and digital signature techniques, and is easiest for use with languages like Visual Basic and Java that don't interface to C data structures very well. The container object interface provides services to create and destroy envelopes and secure sessions, to add security attributes such as encryption information and signature keys to a container object, and to move data into and out of a container. Because of its simplicity and ease of use, it's strongly recommended that you use this interface if at all possible.

Mid-level Interface

This interface requires some knowledge of encryption and digital signature techniques. Because it handles encoding of things like session keys and digital signatures but not of the data itself, it's better suited for applications that require high-speed data encryption, or encryption of many small data packets (such as an encrypted terminal session). The mid-level interface provides services such as routines to export and import encrypted keys and to create and check digital signatures. The high-level interface is built on top of this interface.

Low-level Interface

This interface requires quite a bit of knowledge of encryption and digital signature techniques. It provides a direct interface to the raw encryption capabilities of cryptlib. The only reason for using these low-level routines is if you need them as building blocks for your own custom encryption protocol. Note though that cryptlib is designed to benefit the application of encryption in standard protocols and not the raw use of crypto in home-made protocols. Getting such security protocols right is very difficult, with many "obvious" and "simple" approaches being quite vulnerable to attack. This is why cryptlib encourages the use of vetted security protocols, and does not encourage roll-your-own security mechanisms. In particular if you don't know what PKCS #1 padding is, what CBC does, or why you need an IV, you shouldn't be using this interface.

The low-level interface serves as an interface to a range of plug-in encryption modules that allow encryption algorithms to be added in a fairly transparent manner, with a standardised interface allowing any of the algorithms and modes supported by cryptlib to be used with a minimum of coding effort. As such the main function of the action object interface is to provide a standard, portable interface between the underlying encryption routines and the user software. The mid-level interface is built on top of this interface.

Objects and Interfaces

The cryptlib object types are as follows:

Type	Description
CRYPT_CERTIFICATE	A key certificate objects that usually contain a key certificate for an individual or organisation but can also contain other information such as certificate chains or digital signature attributes.
CRYPT_CONTEXT	A encryption context objects that contain encryption, digital signature, hash, or MAC information.
CRYPT_DEVICE	A device object that provide a mechanism for working with crypto devices such as crypto hardware accelerators and PCMCIA and smart cards.
CRYPT_ENVELOPE	An envelope container object that provide an abstract container for performing encryption, signing, and other security-related operations on an item of data.
CRYPT_KEYSET	A key collection container object that contain collections of public or private keys.
CRYPT_SESSION	A secure session object that manage a secure session with a server or client.

These objects are referred to via arbitrary integer values, or handles, which have no meaning outside of cryptlib. All data pertaining to an object is managed internally by cryptlib, with no outside access to security-related information being possible. There is also a generic object handle of type CRYPT_HANDLE which is used in cases where the exact type of an object is not important. For example most cryptlib functions that require keys can work with either encryption contexts or key certificate objects, so the objects they use have a generic CRYPT_HANDLE which is equivalent to either a CRYPT_CONTEXT or a CRYPT_CERTIFICATE.

Objects and Attributes

Each cryptlib object has a number of attributes of type CRYPT_ATTRIBUTE_TYPE that you can get, set, and in some cases delete. For example an encryption context would have a key attribute, a certificate would have issuer name and validity attributes, and an envelope would have attributes such as passwords or signature information, depending on the type of the envelope. Most cryptlib objects are controlled by manipulating these attributes.

The attribute classes are as follows:

Type	Description
CRYPT_ATTRIBUTE_name	Generic attributes that apply to all objects.
CRYPT_CERTINFO_name	Certificate object attributes.
CRYPT_CTXINFO_name	Encryption context attributes.
CRYPT_DEVINFO_name	Crypto device attributes.
CRYPT_ENVINFO_name	Envelope attributes.
CRYPT_KEYINFO_name	Keyset attributes.
CRYPT_OPTION_name	cryptlib-wide configuration options.
CRYPT_PROPERTY_name	Object properties.

Type	Description
CRYPT_SESSIONINFO_name	Session attributes.

Some of the attributes apply only to a particular object type but others may apply across multiple objects. For example a certificate contains a public key, so the key size attribute, which is normally associated with a context, would also apply to a certificate. To determine the key size for the key in a certificate, you would read its key size attribute as if it were an encryption context.

Attribute data is either a single numeric value or variable-length data consisting of a (data, length) pair. Numeric attribute values are used for objects, boolean values and integers. Variable-length data attribute values are used for text strings, binary data blobs, and representations of time using the ANSI/ISO standard seconds-since-1970 format.

Interfacing with cryptlib

All necessary constants, types, structures, and function prototypes are defined in a language-specific header file as described below. You need to use these files for each module that makes use of cryptlib. Although many of the examples given in this manual are for C/C++ (the more widely-used ones are given for other languages as well), they apply equally for the other languages.

All language bindings for cryptlib are provided in the **bindings** subdirectory. Before you can use a specific language interface, you may need to copy the file(s) for the language that you're using into the cryptlib main directory or the directory containing the application that you're building. Alternatively, you can refer to the file(s) in the **bindings** directory by the absolute pathname.

Initialisation

Before you can use any of the cryptlib functions, you need to call the **cryptInit** function to initialise cryptlib. You also need to call its companion function **cryptEnd** at the end of your program after you've finished using cryptlib. **cryptInit** initialises cryptlib for use, and **cryptEnd** performs various cleanup functions including automatic garbage collection of any objects you may have forgotten to destroy. You don't have to worry about inadvertently calling **cryptInit** multiple times (for example if you're calling it from multiple threads), it will handle the initialisation correctly. However you should only call **cryptEnd** once when you've finished using cryptlib.

If you call **cryptEnd** and there are still objects in existence, it will return CRYPT_ERROR_INCOMPLETE to inform you that there were leftover objects present. cryptlib can tell this because it keeps track of each object so that it can erase any sensitive data that may be present in the object (**cryptEnd** will return a CRYPT_ERROR_INCOMPLETE error to warn you, but will nevertheless clean up and free each object for you).

To make the use of **cryptEnd** in a C or C++ program easier, you may want to use the `C atexit()` function or add a call to **cryptEnd** to a C++ destructor in order to have **cryptEnd** called automatically when your program exits.

If you're going to be doing something that needs encryption keys (which is pretty much everything), you should also perform a randomness poll fairly early on to give cryptlib enough random data to create keys:

```
cryptAddRandom( NULL, CRYPT_RANDOM_SLOWPOLL );
```

Randomness polls are described in more detail in "Random Numbers" on page 364. The randomness poll executes asynchronously, so it won't stall the rest of your code while it's running. The one possible exception to this polling on start-up is when you're using cryptlib as part of a larger application where you're not certain that cryptlib will actually be used. For example a PHP script that's run repeatedly from the command line may only use the encryption functionality on rare occasions (or not at all), so that it's better to perform the slow poll only when it's actually needed rather than unconditionally every time the script is invoked. This is a somewhat special

case though, and normally it's better practice to always perform the slow poll on start-up.

As the text above mentioned, you should initialise cryptlib when your program first starts and shut it down when your program is about to exit, rather than repeatedly starting cryptlib up and shutting it down again each time you use it. Since cryptlib consists of a complete crypto operating system with extensive initialisation, internal security self-tests, and full resource management, repeatedly starting and stopping it will unnecessarily consume resources such as processor time during each initialisation and shutdown. It can also tie up host operating system resources if the host contains subsystems that leak memory or handles (under Windows, ODBC and LDAP are particularly bad, with ODBC leaking memory and LDAP leaking handles. DNS is also rather leaky — this is one of the reasons why programs like web browsers and FTP clients consume memory and handles without bounds). To avoid this problem, you should avoid repeatedly starting up and shutting down cryptlib:

Right

```
cryptInit();
serverLoop:
    process data;
cryptEnd();
```

Wrong

```
serverLoop:
    cryptInit();
    process data;
    cryptEnd();
```

C / C++

To use cryptlib from C or C++ you would use:

```
#include "cryptlib.h"

cryptInit();

/* Calls to cryptlib routines */

cryptEnd();
```

C# / .NET

To use cryptlib from C# / .NET, add `cryptlib.cs` to your .NET project and the cryptlib DLL to your path, and then use:

```
using cryptlib;

crypt.Init();

// Calls to cryptlib routines

crypt.End();
```

If you're using a .NET language other than C# (for example VB.NET), you'll need to build `cryptlib.cs` as a class library first. From Visual Studio, create a new C# project of type Class Library, add `cryptlib.cs` to it, and compile it to create a DLL. Now go to your VB project and add the DLL as a Reference. The cryptlib classes and methods will be available natively using VB (or whatever .NET language you're using).

All cryptlib functions are placed in the `crypt` class, so that standard cryptlib functions like:

```
cryptSetAttribute( cryptContext, CRYPT_CTXINFO_KEYSIZE, 1024 / 8 );
```

become:

```
crypt.SetAttribute( cryptContext, crypt.CTXINFO_KEYSIZE, 1024 / 8 );
```

In general when calling cryptlib functions you can use Strings wherever the cryptlib interface requires a null-terminated C string, and byte arrays wherever the cryptlib interface requires binary data.

Instead of returning a status value like the native C interface, the .NET version throws `CryptException` for error status returns, and returns integer or string data return values as the return value:

```
value = crypt.GetAttribute( cryptContext, crypt.CTXINFO_ALGO );
stringValue = crypt.GetAttributeString( cryptContext,
    crypt.CTXINFO_ALGO_NAME );
```

Delphi

To use cryptlib from Delphi, add the cryptlib DLL to your path and then use:

```
implementation
uses cryptlib;

cryptInit;

{ Calls to cryptlib routines }

cryptEnd;
end;
```

The Delphi interface to cryptlib is otherwise mostly identical to the standard C/C++ one.

Java

To use cryptlib with Java, put **cryptlib.jar** on your classpath and use `System.loadLibrary()` to load the cryptlib shared library. You can then use:

```
import cryptlib.*;

class Cryptlib
{
    public static void main( String[] args )
    {
        System.loadLibrary( "cl" );    // cryptlib library name

        try
        {
            crypt.Init();

            //Calls to cryptlib routines

            crypt.End();
        }
        catch( CryptException e )
        {
            // cryptlib returned an error
            e.printStackTrace();
        }
    }
};
```

All cryptlib functions are placed in the `crypt` class, so that standard cryptlib functions like:

```
cryptSetAttribute( cryptContext, CRYPT_CTXINFO_KEYSIZE, 1024 / 8 );
```

become:

```
crypt.SetAttribute( cryptContext, crypt.CTXINFO_KEYSIZE, 1024 / 8 );
```

In general when calling cryptlib functions you can use Java strings wherever the cryptlib interface requires a null-terminated C string, and Java byte arrays wherever the cryptlib interface requires binary data. In addition as of JDK 1.4 there is a `nio.ByteBuffer` class that can be “direct”, which provides a more efficient alternative to standard byte arrays since there’s no need to perform any copying.

Instead of returning a status value like the native C interface, the JNI version throws `CryptException` for error status returns, and returns integer or string data return values as the return value:

```
value = crypt.GetAttribute( cryptContext, crypt.CTXINFO_ALGO );
stringValue = crypt.GetAttributeString( cryptContext,
    crypt.CTXINFO_ALGO_NAME );
```

Python

To build the Python interface to cryptlib, run `python setup.py install` to build and install the `python.c` extension module. On a Unix platform you may need to create a symlink from `cl` to the actual shared library before you do this. Once you've done this you can use:

```
from cryptlib_py import *

cryptInit()

# Calls to cryptlib routines

cryptEnd()
```

Tcl

To use cryptlib from Tcl, you use the Cryptkit extension. Cryptkit is a stubs-enabled extension that can be used with any modern Tcl interpreter (at least, Tcl 8.4 or later). To build Cryptkit you'll need a copy of Tcl that can interpret Starkits, either Tclkit, the single file Tcl/Tk executable available from

<http://www.equi4.com/pub/tk>, or ActiveTcl from

<http://www.activestate.com>. You'll also need to download the Critcl Starkit from <http://mini.net/sdarchive/critcl.kit>, and make sure that the current directory contains `cryptlib.h` and a copy of the cryptlib static library, named `libcl_${platform}.a`, where `$platform` is the current platform name as provided by the Critcl platform command. For example under x86 Linux the library would be called `libcl_Linux-x86.a`. Then run the following Critcl command:

```
critcl -pkg cryptkit
```

This will leave you with a `lib` directory containing the information ready for use in any Tcl application. Once you've done this you can use:

```
package require Cryptkit

cryptInit

# Calls to cryptlib routines

cryptEnd
```

Since Tcl objects already contain length information, there's no need to pass length parameters to cryptlib function calls. This applies for the `AddCertExtension`, `CheckSignature`, `CheckSignatureEx`, `CreateSignature`, `CreateSignatureEx`, `Decrypt`, `Encrypt`, `ExportCert`, `ExportKey`, `ExportKeyEx`, `GetCertExtension`, `ImportKey`, `PushData`, and `SetAttributeString` functions.

Visual Basic

To use cryptlib from Visual Basic you would use:

```
' Add cryptlib.bas to your project

cryptInit

' Calls to cryptlib routines

cryptEnd
```

The Visual Basic interface to cryptlib is otherwise mostly identical to the standard C/C++ one.

Return Codes

Every cryptlib function returns a status code to tell you whether it succeeded or failed. If a function executes successfully, it returns `CRYPT_OK`. If it fails, it returns one of the status values detailed in "Error Handling" on page 367. The sample code used in this manual omits the checking of status values for clarity, but when using cryptlib you should check return values, particularly for critical functions

such as any that perform active crypto operations like processing data in envelopes, activating and using secure sessions, signing and checking certificates, and encryption and signing in general.

The previous initialisation code, rewritten to include checking for returned status values, is:

```
int status;

status = cryptInit();
if( status != CRYPT_OK )
    /* cryptlib initialisation failed */;

/* Calls to cryptlib routines */

status = cryptEnd();
if( status != CRYPT_OK )
    /* cryptlib shutdown failed */;
```

The C/C++ versions of cryptlib provide the `cryptStatusOK()` and `cryptStatusError()` macros to make checking of these status values easier. The C#, Java, and Python versions throw exceptions of type `CryptException` instead of returning error codes. These objects contain both the status code and an English error message. In C# the `CryptException` class has `Status` and `Message` properties:

```
try
{
    crypt.Init();

    crypt.End();
}
catch( CryptException e )
{
    int status = e.Status;
    String message = e.Message;
}
```

In Java the `CryptException` class has `getStatus()` and `getMessage()` accessors:

```
try
{
    crypt.Init();

    crypt.End();
}
catch( CryptException e )
{
    int status = e.getStatus();
    String message = e.getMessage();
}
```

In Python the exception value is a tuple containing the status code, then the message:

```
try:
    cryptInit()

    cryptEnd()
except CryptException, e:
    status, message = e
```

Working with Object Attributes

All object attributes are read, written, and deleted using a common set of functions: **`cryptGetAttribute/cryptGetAttributeString`** to get the value of an attribute, **`cryptSetAttribute/cryptSetAttributeString`** to set the value of an attribute, and **`cryptDeleteAttribute`** to delete an attribute. Attribute deletion is only valid for a small subset of attributes for which it makes sense, for example you can delete the validity date attribute from a certificate before the certificate is signed but not after it's signed, and you can never delete the algorithm-type attribute from an encryption context.

cryptGetAttribute and **cryptSetAttribute** take as argument an integer value or a pointer to a location to receive an integer value:

```
int keySize;

cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_PUBLICKEY, cryptKey );
cryptGetAttribute( cryptContext, CRYPT_CTXINFO_KEYSIZE, &keySize );
```

cryptGetAttributeString and **cryptSetAttributeString** take as argument a pointer to the data value to get or set and a length value or pointer to a location to receive the length value:

```
char emailAddress[ 128 ]
int emailAddressLength;

cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    "1234", 4 );
cryptGetAttributeString( cryptCertificate, CRYPT_CERTINFO_RFC822NAME,
    emailAddress, &emailAddressLength );
```

This leads to a small problem: How do you know how big to make the buffer? The answer is to use **cryptGetAttributeString** to tell you. If you pass in a null pointer for the data value, the function will set the length value to the size of the data, but not do anything else. You can then use code like:

```
char *emailAddress;
int emailAddressLength;

cryptGetAttributeString( cryptCertificate, CRYPT_CERTINFO_RFC822NAME,
    NULL, &emailAddressLength );
emailAddress = malloc( emailAddressLength );
cryptGetAttributeString( cryptCertificate, CRYPT_CERTINFO_RFC822NAME,
    emailAddress, &emailAddressLength );
```

to obtain the data value. In most cases this two-step process isn't necessary, the standards that cryptlib conforms to generally place limits on the size of most attributes so that cryptlib will never return more data than the fixed limit. For example most strings in certificates are limited to a maximum length set by the **CRYPT_MAX_TEXTSIZE** constant. More information on these sizes is given with the descriptions of the different attributes.

The Visual Basic version is:

```
Dim emailAddress as String
Dim emailAddressLength as Integer

cryptGetAttributeString cryptCertificate, CRYPT_CERTINFO_RFC822NAME, _
    0, emailAddressLength
emailAddress = String( emailAddressLength, vbNullChar )
cryptGetAttributeString cryptCertificate, CRYPT_CERTINFO_RFC822NAME, _
    emailAddress, emailAddressLength
```

In Python you can use **cryptGetAttributeString** and **cryptSetAttributeString** as usual, or use a shortcut syntax to make accessing attributes less verbose. The normal syntax follows the C form but migrates the integer output values (the length from **cryptGetAttributeString** or the output value from **cryptGetAttribute**) to return values, and doesn't require a length for **cryptSetAttributeString**:

```
from array import *

emailAddress = array( 'c', 'x' * 128 )

cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    "1234" )
emailAddressLength = cryptGetAttributeString( cryptCertificate,
    CRYPT_CERTINFO_RFC822NAME, emailAddress )
```

The shortcut syntax allows you to get/set attributes as if they were integer and string members of the object (without the **CRYPT_** prefix):

```
cryptEnvelope.ENVINFO_PASSWORD = "1234"
emailAddress = cryptCertificate.CERTINFO_RFC822NAME
```

Just as with Python, C# and Java also migrate returned data to return values. In the C# and Java cases the string functions take byte arrays or Strings. When passing a byte array, you can optionally specify an offset following it for **cryptGetAttributeString** and an offset and length following it for **cryptSetAttributeString**. There is also a special version of **cryptGetAttributeString** that returns Strings for convenience:

```
crypt.SetAttributeString( cryptEnvelope, crypt.ENVINFO_PASSWORD,
    "1234" );
String emailAddress = crypt.GetAttributeString( cryptCertificate,
    crypt.CERTINFO_RFC822NAME );
```

Finally, **cryptDeleteAttribute** lets you delete an attribute in the cases where that's possible:

```
cryptDeleteAttribute( cryptCertificate, CRYPT_CERTINFO_VALIDFROM );
```

All access to objects and object attributes is enforced by cryptlib's security kernel. If you try to access or manipulate an attribute in a manner that isn't allowed (for example by trying to read a write-only attribute, trying to assign a string value to a numeric attribute, trying to delete an attribute that can't be deleted, trying to set a certificate-specific attribute for an envelope, or some similar action) cryptlib will return an error code to tell you that this type of access is invalid. If there's a problem with the object that you're trying to manipulate, cryptlib will return **CRYPT_ERROR_PARAM1** to tell you that the object handle parameter passed to the function is invalid. If there's a problem with the attribute type (typically because it's invalid for this object type) cryptlib will return **CRYPT_ERROR_PARAM2**. If there's a problem with the attribute value, cryptlib will return **CRYPT_ERROR_PARAM3**, and if there's a problem with the length (for the functions that take a length parameter) cryptlib will return **CRYPT_ERROR_PARAM4**. If you try to perform an attribute access which is disallowed (reading an attribute that can't be read, writing to or deleting a read-only attribute, or something similar) cryptlib will return **CRYPT_ERROR_PERMISSION**.

Finally, if you try to access an attribute that hasn't been initialised or isn't present, cryptlib will return **CRYPT_ERROR_NOTINITED** or **CRYPT_ERROR_NOTFOUND**, the only real distinction between the two is that the former is typically returned for fixed attributes that haven't had a value assigned to them yet while the latter is returned for optional attributes that aren't present in the object.

Attribute Types

Attribute values can be boolean or numeric values, cryptlib objects, time values, text strings, or binary data:

Type	Description
Binary	A binary data string that can contain almost anything.
Boolean	Flags that can be set to 'true' (any nonzero value) or 'false' (a zero value), and control whether a certain option or operation is enabled or not. For example the CRYPT_CERTINFO_CA attribute in a certificate controls whether a certificate is marked as being a CA certificate or not. Note that cryptlib uses the value 1 to represent 'true', some languages may represent this by the value -1.
Numeric	A numeric constant such as an integer value or a bitflag. For example the CRYPT_CTXINFO_KEYSIZE attribute specifies the size of a key (in bytes) and the CRYPT_CERTINFO_CRLREASON attribute specifies a bitflag that indicates why a CRL was issued.
Object	A handle to a cryptlib object. For example the CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO attribute specifies the public key to be added to a certificate.

Type	Description
String	<p>A text string that contains information such as a name, message, email address, or URL. Strings are encoded using the standard system local character set, usually ASCII or latin-1 or UTF-8 (depending on the system), however under Windows CE, which is a Unicode environment, these are Unicode strings. In (very rare) cases where the standard system character set doesn't support the characters used in the string (for example when encoding Asian characters), the characters used will be Unicode or widechars. For all intents and purposes you can assume that all strings are encoded in the standard character set that you'd normally use, cryptlib will perform all conversions for you.</p> <p>An example string attribute is CRYPT_CTXINFO_LABEL, which contains a human-readable label used to identify private keys.</p> <p>The most frequently used text string components are those that make up a certificate's distinguished name, which identifies the certificate owner. Most of these components are limited to a maximum of 64 characters by the X.500 standard that covers certificates and their components, and cryptlib provides the CRYPT_MAX_TEXTSIZE constant for use with these components (this value is also used for most other strings such as key labels). Since this value is specified in characters rather than bytes, Unicode strings can be several times as long as this value when their length is expressed in bytes, depending on which data type the system uses to represent Unicode characters.</p>
Time	<p>The ANSI/ISO C standard time value containing the local time expressed as seconds since 1970. This is a binary (rather than numeric) field, with the data being the time value (in C and C++ this is a time_t, usually a signed long integer).</p> <p>Due to the vagaries of international time zones and daylight savings time adjustments, it isn't possible to accurately compare two local times from different time zones, or made across a DST switch (consider for example a country switching to DST, which has two 2am times while another country only has one). Because of this ambiguity, times read from objects such as certificates may appear to be out by an hour or two.</p>

Since most text strings have a fixed maximum length, you can use code like:

```
char commonName[ CRYPT_MAX_TEXTSIZE + 1 ];
int commonNameLength;

/* Retrieve the component and null-terminate it */
cryptGetAttributeString( cryptCertificate, CRYPT_CERTINFO_COMMONNAME,
    commonName, &commonNameLength );
commonName[ commonNameLength ] = '\0';
```

to read the value, in this case the common name of a certificate owner.

Note the explicit addition of the terminating null character, since the text strings returned aren't null-terminated.

In Visual Basic this is:

```
Dim commonName As String
Dim commonNameLength As Long
```

```

commonName = String( CRYPT_MAX_TEXTSIZE + 1 , vbNullChar )
cryptGetAttributeString cryptCertificate, CRYPT_CERTINFO_COMMONNAME, _
    commonName, commonNameLength
commonName = Left( commonName, InStr( commonName, vbNullChar ) - 1 )

```

The description above assumes that the common name is expressed in a single-byte character set. Since the values passed to **cryptGetAttributeString** and **cryptSetAttributeString** are untyped, their length is given in bytes and not in characters (which may not be byte-sized). For Unicode strings, you need to multiply the size of the buffer by the size of a Unicode character on your system to get the actual size to pass to the function, or divide by the size of a Unicode character to get the number of characters returned. For example to perform the same operation as above in a Unicode environment you'd use:

```

wchar_t commonName[ CRYPT_MAX_TEXTSIZE + 1 ];
int commonNameLength;

/* Retrieve the component and null-terminate it */
cryptGetAttributeString( cryptCertificate, CRYPT_CERTINFO_COMMONNAME,
    commonName, &commonNameLength );
commonName[ commonNameLength / sizeof( wchar_t ) ] = L'\0';

```

Attribute Lists and Attribute Groups

Several of the container object types (certificates, envelopes, and sessions) contain large collections of attributes that you can process as a list rather than having to access each attribute individually by type. The list of attributes is managed through the use of an attribute cursor that cryptlib maintains for each container object. You can set or move the cursor either to an absolute position in the list of attributes or relative to the current position.

Object attributes are usually grouped into collections of related attributes. For example an envelope object may contain a group of attributes consisting of a signature, the key that generated the signature, associated signing attributes such as the time and data type being signed, and even a timestamp on the signature itself. Similarly, a session object may have a group of attributes consisting of a server name, server port, and server key. So instead of a straight linear list of attributes:

Object — Attr — Attr — Attr — Attr

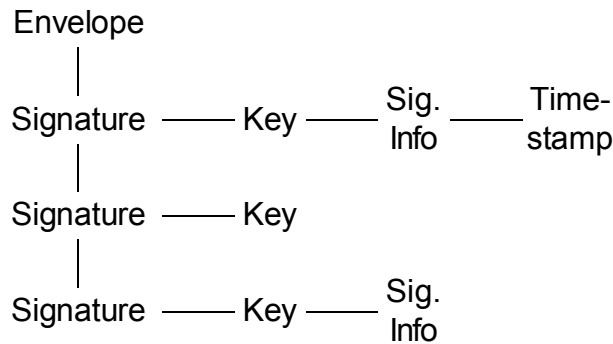
the attributes are arranged by group:

```

Object
|
Group — Attr — Attr — Attr
|
Group — Attr
|
Group — Attr — Attr

```

Some objects may contain multiple instances of attribute groups, each of which contains its own set of attributes. For example an envelope could contain several signature attribute groups, and each attribute group will contain its own signing keys, certificates, signature information such as the signing time, and so on. One particular instance of the abstract group/attribute view shown above would be:



In order to navigate across attribute groups, and across attributes within a group, cryptlib provides the attribute cursor functionality described in the section that follows. As well as moving the cursor back and forth across attribute groups and attributes within the group, you can also position it directly on a group or attribute. In the common case where only a single attribute group is present, for example an envelope object that contains a single signature or a session object that contains user information for a single user:



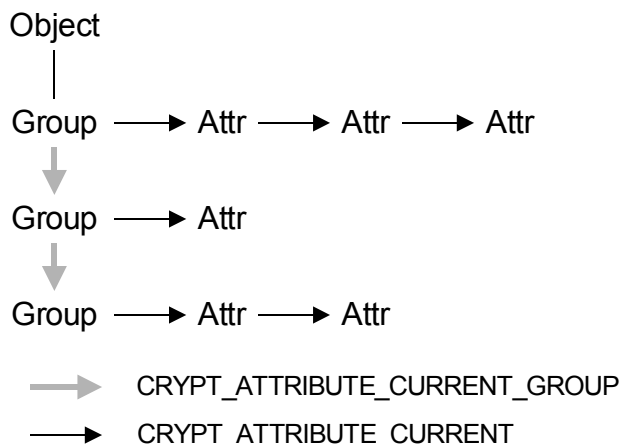
you can treat the attributes as a single flat list of attributes and not worry about the hierarchical arrangement into groups.

Attribute Cursor Management

You can move the attribute cursor by setting an attribute that tells cryptlib where to move it to. This attribute, either `CRYPT_ATTRIBUTE_CURRENT_GROUP` when moving by attribute group or `CRYPT_ATTRIBUTE_CURRENT` when moving by attribute within the current group, takes as value a cursor movement code that moves the cursor either to an absolute position (the first or last group or attribute in the list) or relative to its current position. The movement codes are:

Code	Description
<code>CRYPT_CURSOR_FIRST</code>	Move the cursor to the first group or attribute.
<code>CRYPT_CURSOR_LAST</code>	Move the cursor to the last group or attribute.
<code>CRYPT_CURSOR_NEXT</code>	Move the cursor to the next group or attribute.
<code>CRYPT_CURSOR_PREV</code>	Move the cursor to the previous group or attribute.

Moving by attribute group or attribute then works as follows:



Note that `CRYPT_ATTRIBUTE_CURRENT` only moves the cursor within the current group. Once you get to the start or end of the group, you need to use `CRYPT_ATTRIBUTE_CURRENT_GROUP` to move on to the next one. Moving the cursor from one group to another will reset the cursor position to the first attribute within the group if it's been previously moved to some other attribute within the group. For example to move the cursor to the start of the first attribute group in a certificate, you would use:

```
cryptSetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_CURRENT_GROUP,
                  CRYPT_CURSOR_FIRST );
```

To advance the cursor to the start of the next group, you would use:

```
cryptSetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_CURRENT_GROUP,
                  CRYPT_CURSOR_NEXT );
```

To advance the cursor to the next attribute in the current group, you would use:

```
cryptSetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_CURRENT,
                  CRYPT_CURSOR_NEXT );
```

In some cases multiple instances of the same attribute can be present, in which case you can use a third level of cursor movement, handled via the `CRYPT_ATTRIBUTE_CURRENT_INSTANCE` attribute, and relative cursor movement to step through the different instances of the attribute. Since the use of multi-valued attributes is rare, it's safe to assume one value per attribute in most cases, so that stepping through multiple attribute instances is unnecessary.

Once you've set the cursor position, you can work with the attribute group or attribute at that position in the usual manner. To obtain the group or attribute type at the current position, you would use:

```
CRYPT_ATTRIBUTE_TYPE groupID;

cryptGetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_CURRENT_GROUP,
                  &groupID );
```

This example obtains the attribute group type, to obtain the attribute type you would substitute `CRYPT_ATTRIBUTE_CURRENT` in place of `CRYPT_ATTRIBUTE_CURRENT_GROUP`. Attribute accesses are relative to the currently selected group, so for example if you move the cursor in an envelope to a signature attribute group and then read the signature key/certificate or signing time, it'll be the one for the currently-selected signature attribute group. Since there can be multiple signatures present in an envelope, you can use this mechanism to read the signing information for each of the ones that are present.

To delete the attribute group at the current cursor position you would use:

```
cryptDeleteAttribute( cryptCertificate,
                    CRYPT_ATTRIBUTE_CURRENT_GROUP );
```

Deleting the attribute group at the cursor position will move the cursor to the start of the group that follows the deleted one, or to the start of the previous group if the one being deleted was the last one present. This means that you can delete every attribute group simply by repeatedly deleting the one under the cursor.

The attribute cursor provides a convenient mechanism for stepping through every attribute group and attribute which is present in an object. For example to iterate through every attribute group you would use:

```
if( cryptSetAttribute( cryptCertificate,
                    CRYPT_ATTRIBUTE_CURRENT_GROUP, CRYPT_CURSOR_FIRST ) == CRYPT_OK )
do
{
    CRYPT_ATTRIBUTE_TYPE groupID;

    /* Get the ID of the attribute group under the cursor */
    cryptGetAttribute( cryptCertificate,
                    CRYPT_ATTRIBUTE_CURRENT_GROUP, &groupID );
```

```

/* Handle the attribute if required */
/* ... */
}
while( cryptSetAttribute( cryptCertificate,
CRYPT_ATTRIBUTE_CURRENT_GROUP, CRYPT_CURSOR_NEXT ) ==
CRYPT_OK );

```

The Visual Basic equivalent is:

```

Dim groupID As CRYPT_ATTRIBUTE_TYPE

If cryptSetAttribute( cryptCertificate, _
CRYPT_ATTRIBUTE_CURRENT_GROUP, CRYPT_CURSOR_FIRST ) == CRYPT_OK
Then
Do
' Get the type of the attribute group under the cursor
cryptGetAttribute cryptCertificate, CRYPT_ATTRIBUTE_CURRENT, _
groupID

' Handle the attribute if required
' ...
Loop While cryptSetAttribute( cryptCertificate, _
CRYPT_ATTRIBUTE_CURRENT_GROUP, CRYPT_CURSOR_NEXT ) == CRYPT_OK
End If

```

To extend this a stage further and iterate through every attribute in every group in the object, you would use:

```

if( cryptSetAttribute( cryptCertificate,
CRYPT_ATTRIBUTE_CURRENT_GROUP, CRYPT_CURSOR_FIRST ) == CRYPT_OK )
do
{
do
{
CRYPT_ATTRIBUTE_TYPE attributeID;

/* Get the ID of the attribute under the cursor */
cryptGetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_CURRENT,
&attributeID );

/* Handle the attribute if required */
/* ... */
}
while( cryptSetAttribute( cryptCertificate,
CRYPT_ATTRIBUTE_CURRENT, CRYPT_CURSOR_NEXT ) == CRYPT_OK );
}
while( cryptSetAttribute( cryptCertificate,
CRYPT_ATTRIBUTE_CURRENT_GROUP, CRYPT_CURSOR_NEXT ) ==
CRYPT_OK );

```

Note that iterating attribute by attribute works within the current attribute group, but as mentioned earlier won't jump from one group to the next — to do that, you need to iterate by group.

You can also position the attribute cursor directly by telling cryptlib which attribute you want to move the cursor to. For example to move the cursor in a certificate object to the extended key usage attribute group you would use:

```

cryptSetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_CURRENT_GROUP,
CRYPT_CERTINFO_EXTKEYUSAGE );

```

Usually the absolute cursor-positioning capability is only useful for certificate objects where you know that certain attributes will be present and that only one instance of the attribute will be present. For envelope and session objects you generally can't tell in advance which attributes will be present and it's quite possible that multiple attribute instances (such as multiple signatures on an envelope) will be present. In this case selecting an attribute will only select the first one that's present, so it's better to use the attribute cursor to walk the list to see what's there.

Using this absolute cursor positioning in a variation of the attribute enumeration operation given earlier, you can enumerate only the attributes of a single attribute group (rather than all groups) by first selecting the group and then stepping through the attributes in it. For example to read all of a certificate's extended key usage types you would use:

```

if( cryptSetAttribute( cryptCertificate,
CRYPT_ATTRIBUTE_CURRENT_GROUP, CRYPT_CERTINFO_EXTKEYUSAGE ) ==
CRYPT_OK )
do
{
CRYPT_ATTRIBUTE_TYPE attributeID;

/* Get the ID of the attribute under the cursor */
cryptGetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_CURRENT,
&attributeID );
}
while( cryptSetAttribute( cryptCertificate,
CRYPT_ATTRIBUTE_CURRENT, CRYPT_CURSOR_NEXT ) == CRYPT_OK );

```

Object Security

Each cryptlib object has its own security settings that affect the way that you can use the object. You can set these attributes, identified by `CRYPT_PROPERTY_name`, after you create an object to provide enhanced control over how it's used. For example on a system that supports threads you can bind an object to an individual thread within a process so that only the thread that owns the object can see it. For any other thread in the process, the object handle is invalid.

You can get and set an object's properties using **cryptGetAttribute** and **cryptSetAttribute**, passing as arguments the object whose property attribute you want to change, the type of property attribute to change, and the attribute value or a pointer to a location to receive the attribute's value. The object property attributes that you can get or set are:

Property/Description	Type
<code>CRYPT_PROPERTY_FORWARDCOUNT</code>	Numeric

The number of times an object can be forwarded (that is, the number of times the ownership of the object can be changed). Each time the object's ownership is changed, the forwarding count decreases by one; once it reaches zero, the object can't be forwarded any further. For example if you set this attribute's value to 1 then you can forward the object to another thread, but that thread can't forward it further.

After you set this attribute (and any other security-related attributes), you should set the `CRYPT_PROPERTY_LOCKED` attribute to ensure that it can't be changed later.

<code>CRYPT_PROPERTY_HIGHSECURITY</code>	Boolean
--	---------

This is a composite value that sets all general security-related attributes to their highest security setting. Setting this value will make an object owned, non-exportable (if appropriate), non-forwardable, and locked. Since this is a composite value representing a number of separate attributes, its value can't be read or unset after being set.

<code>CRYPT_PROPERTY_LOCKED</code>	Boolean
------------------------------------	---------

Locks the security-related object attributes so that they can no longer be changed. You should set this attribute once you've set other security-related attributes such as `CRYPT_PROPERTY_FORWARDCOUNT`.

This attribute is a write-once attribute, once you've set it can't be reset.

<code>CRYPT_PROPERTY_NONEXPORTABLE</code>	Boolean
---	---------

Whether a key in an encryption action object can be exported from the object in encrypted form. Normally only session keys can be exported, and only in encrypted form, however in some cases private keys are also exported in encrypted form when they can be saved to a keyset. By setting this attribute you can make them non-exportable in any form (some keys, such as those held in crypto devices, are non-exportable by default).

This attribute is a write-once attribute, once you've set it can't be reset.

<code>CRYPT_PROPERTY_OWNER</code>	Numeric
-----------------------------------	---------

The identity of the thread that owns the object. The thread's identity is

Property/Description	Type
specified using a value that depends on the operating system, but is usually a thread handle or thread ID. For example under Windows 95/98/ME, NT/2000/XP/Vista, and Windows CE the thread ID is the value returned by the <code>GetCurrentThreadID</code> function, which returns a system-wide unique handle for the current thread.	

You can also pass in a value of `CRYPT_UNUSED`, which unbinds the object from the thread and makes it accessible to all threads in the process.

<code>CRYPT_PROPERTY_USAGECOUNT</code>	Numeric
The number of times an action object can be used before it deletes itself and becomes unusable. Every time an action object is used (for example when a signature encryption object is used to create a signature), its usage count is decremented; once the usage count reaches zero, the object can't be used to perform any further actions (although you can still perform non-action operations such as reading its attributes).	

This attribute is useful when you want to restrict the number of times an object can be used by other code. For example, before you change the ownership of a signature object to allow it to be used by another thread, you would set the usage count to 1 to ensure that it can't be used to sign arbitrary numbers of messages or transactions. This eliminates a troubling security problem with objects such as smart cards where, once a user has authenticated themselves to the card, the software can ask the card to sign arbitrary numbers of (unauthorised) transactions alongside the authorised ones.

This attribute is a write-once attribute, once you've set it can't be reset.

For example to create a triple DES encryption context in one thread and transfer ownership of the context to another thread you would use:

```
CRYPT_CONTEXT cryptContext;

/* Create a context and claim it for exclusive use */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );
cryptSetAttribute( cryptContext, CRYPT_PROPERTY_OWNER, threadID );

/* Generate a key into the context */
cryptGenerateKey( cryptContext );

/* Transfer ownership to another thread */
cryptSetAttribute( cryptContext, CRYPT_PROPERTY_OWNER,
    otherThreadID );
```

The other thread now has exclusive ownership of the context containing the loaded key. If you wanted to prevent the other thread from transferring the context further, you would also have to set the `CRYPT_PROPERTY_FORWARDCOUNT` property to 1 (to allow you to transfer it) and then set the `CRYPT_PROPERTY_LOCKED` attribute (to prevent the other thread from changing the attributes you've set).

Note that in the above code the object is claimed as soon as it's created (and before any sensitive data is loaded into it) to ensure that another thread isn't given a chance to use it when it contains sensitive data. The use of this type of object binding is recommended when working with sensitive information under Windows 95/98/ME, Windows NT/2000/XP/Vista, and Windows CE, since the Win32 API provides several security holes whereby any process in the system may interfere with resources owned by any other process in the system. The checking for object ownership which is performed typically adds a few microseconds to each call, so in extremely time-critical applications you may want to avoid binding an object to a thread. On the other hand for valuable resources such as private keys, you should always consider binding them to a thread, since the small overhead becomes insignificant compared to the cost of the public-key operation.

Although the example shown above is for encryption contexts, the same applies to other types of objects such as keysets and envelopes (although in that case the

information they contain isn't as sensitive as it is for encryption contexts). For container objects that can themselves contain objects (for example keysets), if the container is bound to a thread then any objects that are retrieved from it are also bound to the thread. For example if you're reading a private key from a keyset, you should bind the keyset to the current thread after you open it (but before you read any keys) so that any keys read from it will also automatically be bound to the current thread. In addition if a key which is used to generate another key (for example the key that imports a session key) is bound, then the resulting generated key will also be bound.

On non-multithreaded systems, `CRYPT_PROPERTY_OWNER` and `CRYPT_PROPERTY_FORWARDCOUNT` have no effect, so you can include them in your code for any type of system.

Role-based Access Control

cryptlib implements a form of access control called role-based access control or RBAC in which operations specific to a certain user role can't be performed by a user acting in a different role. For example in many organisations a cheque can only be issued by an accountant and can only be signed by a manager, which prevents a dishonest accountant or manager from both issuing a cheque to themselves and then signing it as well. This security measure is referred to as separation of duty, in which it takes at least two people to perform a critical operation. Similarly, cryptlib uses RBAC to enforce a strong separation of duty between various roles, providing the same effect as the corporate accounting controls that prevent an individual from writing themselves cheques.

cryptlib recognises a variety of user types or roles. The default user type has access to most of the standard functions in cryptlib but can't perform CA management operations or specialised administrative functions that are used to manage certain aspects of cryptlib's operation. When you use cryptlib in the role of a standard user, it functions as a normal crypto/security toolkit.

In addition to the standard user role, it's also possible to use cryptlib in the role of a security officer (SO), a special administrative user who can create new users and perform other administrative functions but can't perform general crypto operations like a normal user. This provides a clear separation of duty between administrative and end-user functionality.

Another role recognised by cryptlib is that of a certification authority that can make use of cryptlib's certificate management functionality but can't perform general administrative functions or non-CA-related crypto operations. Again, this provides a clear separation of duty between the role of the CA and the role of a general user or SO.

Managing User Roles

When a cryptlib object is created, it is associated with a user role which is specified at creation time and can't be accessed by any other user. For example if a private key is created by a CA for signing certificates, it can't be accessed by a normal user because it's only visible to the user acting in the role of the CA. Similarly, although a normal user may be able to see a certificate store, only a CA user can use it for the purpose of issuing certificates. The use of RBAC therefore protects objects from misuse by unauthorised users.

The identity of the user who owns the object is specified as a parameter for the object creation function. Throughout the rest of the cryptlib documentation this parameter is denoted through the use of the `cryptUser` variable. Usually this parameter is set to `CRYPT_UNUSED` to indicate that the user is acting in the role of a normal user and doesn't care about role-based controls. This is typically used in cases where there's only one cryptlib user, for example where cryptlib is running on an end-user PC (e.g. Windows, Macintosh) or a multi-user system that provides each user with the illusion of being on a single-user machine (e.g. Unix). In almost all cases therefore you'd pass in `CRYPT_UNUSED` as the user identity parameter.

In a few specialised cases where the user is acting in a role other than that of a normal user the default user role isn't enough. For example when you want to access a CA certificate store you can't use the role of a normal user to perform the access because only a CA can manipulate a certificate store. This prevents a normal user from issuing themselves certificates in the name of the CA and assorted other mischief such as revoking another user's certificate.

When acting in a different role than that of the default, normal user, you specify the identity of the user whose role you're acting in as a parameter of the object creation function as before, this time passing in the handle of the user identity instead of `CRYPT_UNUSED`. When the object is created, it is associated with the given user and role instead of the default user. The creation and use of user objects is covered in the next section.

Creating and Destroying Users and Roles

The following section is provided purely for forwards compatibility with functionality included in future versions of cryptlib. For the current version of cryptlib the user identity parameter should always be `CRYPT_UNUSED` since user object management isn't enabled in this version.

User objects can only be created and destroyed by an SO user, this being one of the special administrative functions mentioned earlier that can only be performed by an SO. You create a user object with **`cryptCreateUser`**, specifying the identity of the SO who is creating the user object, type of user role that the object is associated with, the name of the user, and a password that protects access to the user object:

```
CRYPT_USER cryptUser;

cryptCreateUser( &cryptUser, cryptSO, type, name, password );
```

The available user types or roles are:

Role	Description
<code>CRYPT_USER_CA</code>	A certification authority who can perform CA management functions but can't perform general-purpose crypto operations.
<code>CRYPT_USER_NORMAL</code>	A standard cryptlib user.
<code>CRYPT_USER_SO</code>	A security officer who can perform administrative functions such as creating or deleting users but who can't perform any other type of operation.

For example to create a normal user object for "John Doe" with the password "password" and a CA user object for "John's Certificate Authority" with the password "CA password" you would use:

```
CRYPT_USER cryptUser, cryptCAUser;

cryptCreateUser( &cryptUser, cryptSO, CRYPT_USER_NORMAL, "John Doe",
    "password" );
cryptCreateUser( &cryptUser, cryptSO, CRYPT_USER_CA, "John's
    Certification Authority", "CA password" );
```

Once a user object is created it can't be used immediately because it's still under the nominal control of the SO who created it rather than the user it's intended for. Before it can be used, control over the object needs to be handed over to the user that it's intended for. After the object is created by the SO, it is said to be in the SO initialised state. Any attempt to use an object when it's in the SO initialised state will result in cryptlib returning `CRYPT_ERROR_NOTINITED`.

To move the newly-created object into a usable state, it's necessary to change its password from the initial one set by the SO to one chosen by the user. Once this change occurs, the object is moved into the user initialised state and is ready for use. You can change the password from the initial one set by the SO to a user-chosen one with **`cryptChangePassword`**:

```
cryptChangePassword( cryptUser, oldPassword, newPassword );
```

When the password has been changed from the one set by the SO to the one chosen by the user, the user object is ready for use.

User objects can also be destroyed by the SO who created them:

```
cryptDeleteUser( cryptUser, "John Doe" );
```

Miscellaneous Issues

This section contains additional information that may be useful when working with cryptlib.

Multi-threaded cryptlib Operation

cryptlib is re-entrant and completely thread-safe (the threading model used is sometimes known as the free-threading model), allowing it to be used with multithreaded applications in systems that support threading. When you use cryptlib in a multithreaded application, you should take standard precautions to ensure that a single resource shared across multiple threads doesn't become a bottleneck, with all threads being forced to wait on a single shared object. For example if you're timestamping large numbers of messages then creating a single timestamping session object (see "Secure Sessions" on page 190) and using that for all timestamping services will result in all of the operations waiting on a single session object, which can often take several seconds to turn around a transaction with a remote server. A better option in this case would be to create a pool of timestamping session objects and use the next free one when required.

A similar situation occurs with other objects such as crypto devices and keysets that may be shared across multiple threads. For example cryptlib provides a facility for automatically fetching a decryption key from a keyset in order to decrypt data (see "Public-Key Encrypted Enveloping" on page 160). This is convenient when handling one or two messages since cryptlib will automatically take care of all of the processing for you, however if you're processing large numbers of messages then the need to read and decrypt the same private key for each message is very inefficient, not only in terms of CPU overhead but also because every message must wait for each of the previous messages to be processed before it gets its turn at the keyset.

A better alternative in this case is to read the private key from the keyset just once and then use it with each envelope, rather than having each envelope read and decrypt the key itself. Extending this even further, if you're using a very large private key, running on a slower processor, or processing large numbers of transactions, you may want to instantiate multiple copies of the private-key object to avoid the single private key again becoming a bottleneck.

In general most private-key operations, when performed on modern processors, are fairly quick, so there's no need to create large numbers of private-key objects for fear of them becoming a bottleneck. In this case the primary bottleneck is the need to read and decrypt the key for each message processed. However, when run on a multiple-CPU system, you should make some attempt to match objects to CPUs — creating a single private-key object on a four-CPU system guarantees that the overall performance will be no better than that of a single-CPU system, since the single object instance acts as a mutex that can only be acquired by one CPU at a time. Standard programming practice for efficient utilisation of resources on multiprocessor systems applies to cryptlib just as it does for other applications. Creating a pool of objects that can be picked up and used as required would be one standard approach to this problem. Some operating systems provide special support for this with functions for thread pooling management. For example, Windows 2000 and XP provide the `QueueUserWorkItem` function, which submits a work item to a thread pool for execution when the next thread becomes available. Windows Vista includes an enhanced version of the thread-pool API that replaces the basic `QueueUserWorkItem` with a more conventional `CreateThreadpoolWork/SubmitThreadpoolWork/CloseThreadpoolWork` combination that provides better control over thread pools, pool management, and pool cleanup.

In order to protect against potential deadlocks when multiple threads are waiting on a single object, cryptlib includes a watchdog timer that triggers after a certain amount of time has been spent waiting on the object. Once this occurs, cryptlib will return `CRYPT_ERROR_TIMEOUT` to indicate that an object is still in use after waiting for it to become available. If you experience timeouts of this kind, you should check your code to see if there are any bottlenecks due to a single object with a long response time being shared by several fast-response-time objects. Note that timeouts are also possible with normal cryptlib object use, for example when communicating data over a slow or stalled network link, so a `CRYPT_ERROR_TIMEOUT` status doesn't automatically mean that the watchdog timer signalled a problem.

To help diagnose situations of this kind, the debug build of cryptlib will display on the console output an indication that it waited on a particular object, along with the object type that it waited on. You can use this information to identify potential bottlenecks in your application.

Linux has a somewhat unusual threading implementation built around the `clone()` system call that can lead to unexpected behaviour with some kernel and/or glibc versions. Two common symptoms of glibc/kernel threading problems are phantom processes (which are actually glibc-internal threads created via `clone()`) being left behind when you application exits, and cryptlib's internal consistency-checking throwing an exception in the debug build when it detects a problem with threading. If you run into either of these situations, you can try different glibc and/or kernel versions to find a combination that works. Searching Internet newsgroups will provide a wealth of information and advice on problems with glibc and Linux threads.

Safeguarding Cryptographic Operations

Running cryptographic operations on general-purpose CPUs shared with other (often untrusted) programs can expose them to risk if the other programs can closely observe or even influence the behaviour of the crypto code. In addition it's possible for a remote system with the ability to precisely time network packet flows to deduce information about crypto operations like a SSL/TLS handshake that result in network traffic as the output of the crypto operation. The timing attacks only affect RSA private-key operations, and only those operations that are directly observable by another party, for example as a result of a network data exchange involving RSA decryption or signing.

There are several simple countermeasures that you can take to avoid this problem. The simplest approach is to use a different crypto mechanism that isn't vulnerable to this problem. By default cryptlib will try and use Diffie-Hellman key exchange in SSL/TLS, which isn't vulnerable to this type of attack because it uses a new random value each time, making it impossible to get repeatable timing measurements. In addition the cryptlib security kernel provides a good degree of protection since it isolates the RSA crypto operations from external observation, making it quite difficult to obtain timing information.

If you must use RSA in a manner in which its operation is visible to an external observer, you can enable randomisation of the RSA operations (known as "blinding") in order to provide the same protection that comes built into Diffie-Hellman. Enabling blinding adds a performance overhead of between two and five percent to each RSA operation. You can enable blinding for RSA operations (and a few other protection measures) by setting the `CRYPT_OPTION_MISC_SIDECHANNELPROTECTION` configuration option as described in "Working with Configuration Options" on page 359.

Many modern CPUs include sophisticated diagnostic and monitoring facilities that provide extensive insight into both the operation of the CPU and the data that it processes. If untrusted processes are running on a CPU alongside ones performing crypto operations, it may be possible for the untrusted processes to recover sensitive data or even crypto keys using built-in CPU monitoring facilities. This can occur even through indirect means such as observing memory access latencies for cached

vs. un-cached data, or branch times for cached vs. un-cached branch target information. Since the level of access provided by many of these diagnostic facilities is almost at the level of an in-circuit emulator (ICE), there are no truly effective defences against this level of threat.

If you're using a CPU that provides this detailed monitoring capability and you're also working with sensitive data or crypto keys, and in particular the private keys used in public-key encryption operations, you need to take precautions to ensure that other code can't misuse these monitoring capabilities to compromise your keys or sensitive data. The simplest and most effective defence is "don't do that, then": Don't allow untrusted code to run alongside your crypto code (or any other code processing sensitive information for that matter).

If you really need to run arbitrary untrusted code at the same time as code that's processing sensitive information, you'll need to use OS-level scheduling and CPU-control facilities to ensure that another process or thread can't run alongside your one and monitor its operation, and that out-of-band channels like CPU caches are flushed after your crypto operations have completed.

Interaction with External Events

Internally, cryptlib consists of a number of security-related objects, some of which can be controlled by the user through handles to the objects. These objects may also be acted on by external forces such as information coming from encryption and system hardware, which will result in a message related to the external action being sent to any affected cryptlib objects. An example of such an event is the withdrawal of a smart card from a card reader, which would result in a card removal message being sent to all cryptlib objects that were created using information stored on the card. This can affect quite a number of objects.

Typically, the affected cryptlib objects will destroy any sensitive information held in memory and disable themselves from further use. If you try to use any of the objects, cryptlib will return `CRYPT_ERROR_SIGNALLED` to indicate that an external event has caused a change in the state of the object.

After an object has entered the signalled state, the only remaining operation you can perform with the object is to destroy it using the appropriate function.

Security Usability Fundamentals

An important consideration when you're building an application is the usability of the security features that you'll be employing. Security experts frequently lament that security has been bolted onto applications as an afterthought, however the security community has committed the exact same sin in reverse, placing usability considerations in second place behind security, if they were considered at all. As a result, we spent the 1990s building and deploying security that wasn't really needed, and now that we're experiencing widespread phishing attacks with viruses and worms running rampant and the security *is* actually needed, we're finding that no-one can use it.

To understand the problem, it's necessary to go back to the basic definition of functionality and security. An application exhibits functionality if things that are supposed to happen, do happen. Similarly, an application exhibits security if things that aren't supposed to happen, don't happen. Security developers are interested in the latter, marketers and management tend to be more interested in the former.

Ensuring that things that aren't supposed to happen don't happen can be approached from both the application side and from the user side. From the application side, the application should behave in a safe manner, defaulting to behaviour that protects the user from harm. From the user side, the application should act in a manner in which the user's expectations of a safe user experience are met. The following sections look at some of the issues that face developers trying to create a user interface for a security application.

Security (Un-)Usability

Before you start thinking about potential features of your security user interface, you first need to consider the environment into which it'll be deployed. Now that we have 10-15 years of experience in (trying to) deploy Internet security, we can see, both from hindsight and because in the last few years people have actually started testing the usability of security applications, that a number of mechanisms that were expected to Solve The Problem don't really work in practice [1]. The idea behind security technology is to translate a hard problem (secure/safe communication and storage) into a simpler problem, not just to shift the complexity from one layer to another. This is an example of Fundamental Truth No.6 of the Twelve Networking Truths, "It is easier to move a problem around than it is to solve it" [2]. Security user interfaces are usually driven by the underlying technology, which means that they often just shift the problem from the technical level to the human level. Some of the most awkward technologies not only shift the complexity but add an extra level of complexity of their own (IPsec and PKI spring to mind).

Re:Good. (Score:2)

by jrockway (229604) *  <jon-nospam@jrock.us> on Wednesday July 12, @01:13AM (#15698213)

> The MiM is the hardest security problem by far there are no easy answers.

Umm, SSL was designed to solve this problem. When you visit your online bank, make sure the cert is valid and that the URL matches the one on your printed bankbook or credit card.

Pretty simple.

(People being too dumb/lazy to check, though, is the hard problem. Fortunately this is evolution at work.)

Figure 1: Blaming the user for security unusability

The major lesson that we've learned from the history of security (un-)usability is that technical solutions like PKI and access control don't align too well with usability conceptual models. As a result, calling in the usability people after the framework of the application's user interface measures have been set in concrete by purely

technology-driven considerations is doomed to failure, since the user interface will be forced to conform to the straightjacket constraints imposed by the security technology rather than being able to exploit the full benefits of years of usability research and experience. Blaming security problems on the user when they're actually caused by the user interface design (Figure 1) is equally ineffective.

This chapter covers some of the issues that affect security user interfaces, and looks at various problems that you'll have to deal with if you want to create an effective user interface for your security application.

Theoretical vs. Effective Security

There can be a significant difference between *theoretical* and *effective* security. In *theory*, we should all be using smart cards and PKI for authentication. However, these measures are so painful to deploy and use that they're almost never employed, making them far less *effectively* secure than basic usernames and passwords. Security experts tend to focus exclusively on the measures that provide the best (theoretical) security, however sometimes these measures provide very little effective security because they end up being misused, or turned off, or bypassed.

Worse yet, when they focus only on the theoretically perfect measures, they don't even try to get lesser security measures right. For example passwords are widely decried as being insecure, but this is mostly because security protocol designers have *chosen* to make them insecure. Both SSL and SSH, the two largest users of passwords for authentication, will connect to anything claiming to be a server and then hand over the password in plaintext after the handshake has completed. No attempt is made to provide even the most trivial protection through some form of challenge/response protocol, because everyone knows that passwords are insecure and so it isn't worth bothering to try and protect them.

This problem is exemplified by the IPsec protocol, which after years of discussion still doesn't have any standardised way to authenticate users based on simple mechanisms like one-time passwords or password-token cards. The IETF even chartered a special working group, IPSRA (IPsec Remote Access), for this purpose. The group's milestone list calls for an IPsec "user access control mechanism submitted for standards track" by March 2001, but six years later its sole output remains a requirements document [3] and an expired draft. As the author of one paper on effective engineering of authentication mechanisms points out, the design assumption behind IPsec was "all password-based authentication is insecure; IPsec is designed to be secure; therefore, you have to deploy a PKI for it"[4]. The result has been a system so unworkable that both developers and users have resorted to doing almost anything to bypass it, from using homebrew (and often insecure) "management tunnels" to communicate keys to hand-carrying static keying material to IPsec endpoints to avoiding IPsec altogether and using mechanisms like SSL-based VPNs, which were never designed to be used for tunnelling IP traffic but are being pressed into service because users have found that almost anything is preferable to having to use IPsec.

More than ten years after SSL was introduced, support for basic password challenge/response authentication is finally (reluctantly) being added, although even there it's only under the guise of enabling use with low-powered devices that can't handle the preferred PKI-based authentication, and still leads to prolonged arguments on the SSL developers list whenever the topic of allowing something other than certificates for user authentication comes up [5]. SSH, a protocol specifically created to protect passwords sent over the network, still operates in a manner in which the recipient ends up in possession of the plaintext password instead of having to perform a challenge-response authentication in its standard mode of authentication. This practice, under the technical label of a tunnelled authentication protocol, is known to be insecure [6][7][8] (what's required for proper security is a cryptographic binding between the outer tunnel and the inner authentication protocol, which SSL's recently-added challenge/response authentication finally performs), and yet both SSL and SSH persist in using it.

A lot of this problem arises from security's origin in the government crypto community. For cryptographers, the security must be perfect — anything less than perfect security would be inconceivable. In the past this has led to all-or-nothing attempts at implementing security such as the US DoD's "C2 in '92" initiative (a more modern form of this might be "PKI or Bust"), which resulted in nothing in '92 or at any other date — the whole multilevel-secure (MLS) operating system push could almost be regarded as a denial-of-service attack on security, since it largely drained security funding in the 1980s and was a significant R&D distraction. As security god Butler Lampson observed when he quoted Voltaire, "The best is the enemy of the good" ("Le mieux est l'ennemi du bien") — a product that offers generally effective (but less than perfect) security will be panned by security experts, who would prefer to see a theoretically perfect but practically unattainable or unusable product instead [9].

Psychologists refer to this phenomenon as zero-risk bias, the fact that people would rather reduce a risk (no matter how small) to zero than create a proportionally much larger decrease that doesn't reduce it to zero. Instead of reducing one risk from 90% to 10% they'll concentrate on reducing another risk from 1% to 0%, yielding a risk reduction of 1% instead of 80%. Zero-risk bias occurs because risk makes people worry, and reducing it to zero means that they don't have to worry about it any more. Obviously this only works if you're prepared to ignore other risks, which is why the phenomenon counts as a psychological bias. An example of such a zero-risk bias was the US' total ban on carcinogenic food additives in the 1950s, which increased the overall risk because (relatively) high-risk non-carcinogenic additives were substituted for (relatively) low-risk carcinogenic ones. The bias ignored the fact that many additives were potentially harmful and focused only on the single class of carcinogenic additives.

The striving for impossibly perfect security comes about because usability has never been a requirement put on those designing security protocols or setting security policies. For example one analysis of a military cryptosystem design reports that "the NSA designers focused almost exclusively on data confidentiality [...] if that meant that it was expensive, hard to use, and required extremely restrictive and awkward policy, or if it might lock out legitimate users from time to time, then so be it" [10]. This type of approach to usability issues was summed up by an early paper on security usability with the observation that "secure systems have a particularly rich tradition of indifference to the user, whether the user is a security administrator, a programmer, or an end user [...] Most research and development in secure systems has strong roots in the military. People in the military are selected and trained to follow rules and procedures precisely, no matter how onerous. This user training and selection decreased the pressure on early systems to be user friendly" [11].

Systems such as this, designed and implemented in a vacuum, can fail catastrophically when exposed to real-world considerations. As the report on the military system discussed above goes on to say, "once the nascent system left the NSA laboratories the emphasis on security above all changed dramatically. The people who approved the final design were not security experts at all. They were the Navy line officers who commanded the fleet. Their actions show that they were far more concerned with data availability rather than data confidentiality [...] any ship or station which became isolated by lack of key became an immediate, high-level issue and prompted numerous and vigorous complaints. A key compromise, by contrast, was a totally silent affair for the commander. Thus, commanders were prodded toward approving very insecure systems". A similar effect occurs with computer security software that pushes critical security decisions into the user interface, where users will find ways to work around the security because they don't understand it and it's preventing them from doing their job.

The best security measures are ones that you can easily explain to users so that they understand the risk and know how to respond appropriately. Don't be afraid to use simple but effective security measures, even if they're not the theoretical best that's available. You should however be careful not to use effective (as opposed to theoretically perfect) security as an excuse for weak security. Using weak or

homebrew encryption mechanisms when proven, industry-standard ones are available isn't effective security, it's weak security. Using appropriately secured passwords instead of PKI is justifiable, effective security (security researcher Simson Garfinkel has termed this "The principle of good security now" [12]).

An example of the conflict between theoretical and effective security is illustrated by what happens when we increase the usability of the security measures in an application. Computer users are supported by a vast and mostly informal network of friends, family, and neighbours (for home users) or office-mates and sysadmins (for work users) who are frequently given passwords and access codes in order to help the user with a problem. The theoretical security model says that once keys and similar secrets are in the hands of the user they'll take perfect care of them and protect them in an appropriate manner. However in practice the application interface to the keys is so hard to use that many users rely on help from others, who then need to be given access to the keys to perform their intended task. Increasing the usability of the security mechanisms helps close this gap between theory and practice by enabling users to manage their own security without having to outsource it to others.

In some cases usability is a fundamental component of a system's security. The Tor anonymity service was specifically designed to maximise usability (and therefore to maximise the number of users) because an unusable anonymity system that attracts few users can't provide much anonymity [13].

User Conditioning

It's often claimed that the way to address security issues is through better user education. As it turns out, we're been educating users for years about security, although unfortunately it's entirely the wrong kind of education. "Conditioning" might be a better term for what's been happening. Whenever users go online, they're subjected to a constant barrage of error messages, warnings, and popups: DNS errors, transient network outages, ASP errors, Javascript problems, missing plugins, temporary server outages, incorrect or expired certificates, problems connecting to the MySQL backend (common on any slashdotted web site), and a whole host of other issues. In one attack, covered in more detail in the section on usability testing below, researchers actually took advantage of this to replace security-related web site images with a message saying that they were being upgraded and would return at a later date.

To see just how tolerant browsers are of errors, enable script debugging (Internet Explorer), look at the error console (Firefox), or install Safari Enhancer and look at the error log (Safari). No matter which detection method you use, you can barely navigate to any Javascript-using page without getting errors, sometimes a whole cascade of them from a single web page. Javascript errors are so pervasive that browsers hide them by default because the web would be unusable if they even displayed them, let alone reacted to them. The result is a web ecosystem that bends over backwards to avoid exposing users to errors, and a user base that's become conditioned to ignoring anything that does leak through.

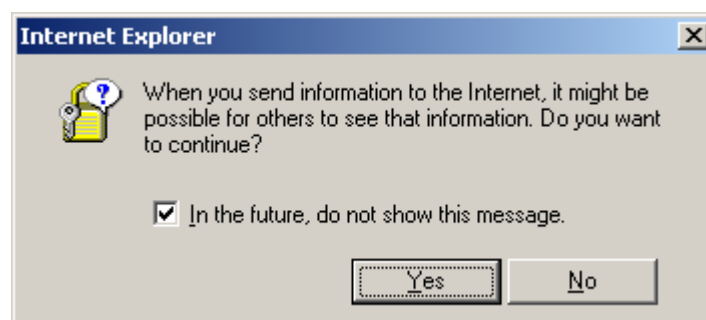


Figure 2: The user has clicked on a button, we'd better pop up a warning dialog

Sometimes the warnings don't even correspond to real errors but seem to exist only for their nuisance value. For example what is the warning in Figure 2 trying to

protect us from? Since we're using a web browser, it's quite obvious that we're about to send information over the Internet. Does a word-processor feel the need to warn users that it's about to perform a spell check, or a spreadsheet that it's about to recalculate a row? Since this warning is automatically displayed when anything at all is sent, we have no idea what the significance of the message is. Are we sending an online banking password, or just searching ebay for cheap dog food? (In this case the browser was trying to protect us from sending a query for dog food to ebay).

This warning would actually be useful in the situation where a user is entering their password on a US banks' insecure login page (discussed later on), but by then the dialog has long since been disabled due to all the false alarms.

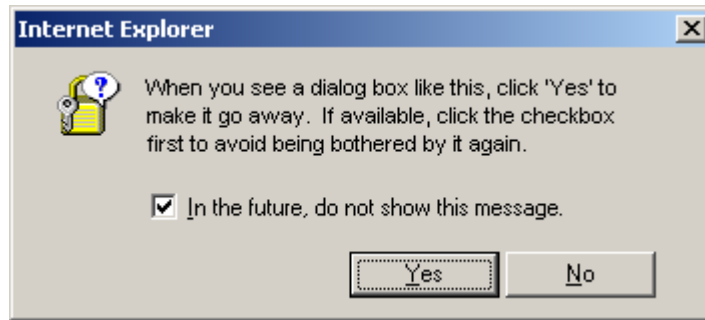


Figure 3: What the previous dialog is really saying

This (and many similarly pointless dialogs that web browsers and other applications pop up) are prime examples of conditioning users to ignore such messages — note the enabled-by-default “Do not show this message again” checkbox, in which the message’s creators admit that users will simply want it to go away and not come back again. The creation of such dialogs is very deeply ingrained in the programmer psyche. When Jeff Bezos came up with Amazon’s one-click shopping system, he had to go back and tell his developers that “one-click” really did mean that the customer only had to make one click, not one click plus a warning dialog plus another click (this works fine in the Amazon case since their order fulfilment system gives you several hours grace to change your mind).

Apple’s user interface design guidelines actually equate the appearance of frequent alerts with a design flaw in the underlying application. OpenBSD, a BSD distribution that concentrates specifically on security, has a policy of “no useless buttons” (unfortunately documented only in developer folklore), meaning that if a particular setting is secure and works for 95% of users then that’s what gets used. Microsoft has also finally acknowledged this problem in their Vista user interface guidelines with the design principle that Vista shouldn’t display error messages when users aren’t likely to change their behaviour as a result of the message, preferring that the message be suppressed if it’s not going to have any effect anyway (it remains to be seen how closely this guideline will be adhered to in practice).

Popups are a user interface instance of the Tragedy of the Commons. If they were less frequent they’d be more effective, but since they’re all over the place anyway there’s nothing to stop *my* application from popping up a few more than everyone else’s application in order to get the user’s attention. An economist would describe this situation by saying that popups have declining marginal utility.

Usability designer Alan Cooper describes these error boxes as “Kafkaesque interrogations with each successive choice leading to a yet blacker pit of retribution and regret [14]. They’re a bit like the land mines that sometimes feature in old war movies, you put your foot down and hear the click and know that although you’re safe now, as soon as you take the next step you’re in for a world of hurt. Unfortunately the war movie get-out-of-jail-free card of being the film’s leading character and therefore indispensable to the plot doesn’t work in the real world - you’re just another redshirt, and you’re not coming back from this mission.

The fix for all of these dialog-box problems is to click ‘Yes’, ‘OK’, or ‘Cancel’ as appropriate if these options are available, or to try again later if they aren’t. Any user

who's used the Internet for any amount of time has become deeply conditioned to applying this solution to all Internet/network problems. These warning dialogs don't warn, they just hassle. This warning message overload has actually been exploited by at least one piece of mobile malware, the Cabir virus, which reconnected to every device within range again and again and again until users eventually clicked 'OK' just to get rid of the message [15] (the situation wasn't helped by the fact that Symbian OS pops up a warning for every application, even a signed one, that originates from anywhere other than Symbian, training users to click 'OK' automatically).

Even when popups provide legitimate warnings of danger, user reactions to the warning may not be what the developers of the application were expecting. The developers of the TrustBar browser plugin, which warns users of phishing sites, found in one evaluation of the system that almost all users disabled the popups or even stopped using the plugin entirely because they found the popups disturbing and felt less safe due to the warnings [16]. Although the whole point of security warnings is to, well, warn of security issues, this makes users feel uneasy to the point where they'll disable the warnings in order to feel better². As security researcher Amir Herzberg puts it, "Defend, don't ask". Building something that relies on user education to be effective is a recipe for disaster. No-one has the time to learn how to use it, so they'll only be adopted by a small number of users, typically hard-core geeks and, in consumer electronics, gadget fanatics [17].

The best approach to the human-factors problem posed by warning dialogs is to redesign the way that the application works so that they're no longer needed. Since users will invariably click 'OK' (or whatever's needed to make the dialog disappear so that they get on with their job), the best way to protect the user is to actually do the right thing, rather than abrogating responsibility to the user. As Mr. Miyagi says in *Karate Kid II*, "Best block, not be there", or as rendered into a computing context by Gordon Bell, "The cheapest, fastest, and most reliable components of a computer system are those that aren't there". In a security user interface context, the best warning dialog is one that isn't there, with the application doing the right thing without having to bother the user.

Certificates and Conditioned Users

When certificates are used to secure network communications, a genuine attack displays symptoms that are identical to the dozens of other transient problems that users have been conditioned to ignore. In other words we're trying to detect attacks using certificates when an astronomical false positive rate (endless dialogs and warnings crying wolf) has conditioned users to ignore any warnings coming from the certificate layer. In order to be effective, the false positive rate must be close to zero to have any impact on the user.

An example of the effect of this user conditioning was revealed in a recent case where a large bank accidentally used an invalid certificate for its online banking services. An analysis of site access logs indicated that of the approximately 300 users who accessed the site, just one single user turned back when faced with the invalid certificate [18]. Although privacy concerns prevented a full-scale study of users' reactions from being carried out, an informal survey indicated that users were treating this as yet another transient problem to be sidestepped. Psychologists call this approach judgemental heuristics (non-psychologists call it "guessing"), a shortcut to having to think that works reasonably well most of the time at the cost of an occasional mistake, and the result of the use of these heuristics is termed an automatic or click, whirr response [19]. As an example of the use of judgemental heuristics, one user commented that "Hotmail does this a lot, you just wait awhile and it works again". The Internet (and specifically the web and web browsers) have conditioned users to act this way: Guessing is cheap, if you get it right it's very quick, and if you

² Applying the ostrich algorithm is a natural human reaction to things that make us uneasy. When a security researcher demonstrated to his parents that the lock on the front door of their house could be picked in a matter of seconds and offered relatively easy unauthorised entry to their house, their reaction was to ask him not to inform his of this again.

don't get it right you just click the back button and try again. This technique has been christened "information foraging" by HCI researchers [20], but is more commonly known as "maximum benefit for minimum effort", or by somewhat more negative label of "laziness" (in this case not in the usual negative sense, it's merely optimising the expenditure of effort).

In a similar case, this time with a government site used to pay multi-thousand dollar property taxes, users ignored the large red cross and warning text that the certificate was invalid shown in Figure 4 for over two months before a security expert notified the site administrators that they needed to fix the certificate. In yet another example, a major US credit union's certificate was invalid for over a year without anyone noticing.

Security

Our site is hosted on a secure server where software encrypts the credit card number into our rates reconciliation system. You can enter your credit card number on a secure form and transmit the form over the internet to a secure server without risk of an intermediary obtaining your credit card information. Your credit card details are temporarily stored on the secure server until your payment is completed and confirmed. After your payment is complete, these details are transferred to an offline database, using a secure transfer mechanism, and deleted from the site. At no stage are your credit card details held in a complete form at the offline site, but rather held in a truncated form for reconciliation purposes only.



Figure 4: This certificate warning didn't stop users from making multi-thousand-dollar payments via the site

These real-life examples, taken from major banking sites and a large government site, indicate that certificates, when deployed into a high-false-positive environment, are completely ineffective in performing their intended task of preventing man-in-the-middle attacks.

SSH fares little better than SSL, with the majority of users accepting SSH server keys without checking them. This occurs because, although SSH users are in general more security-aware than the typical web user, the SSH key verification mechanism requires that the user stop whatever they're trying to do and verify from memory a long string of hex digits (the key fingerprint) displayed by the client software. A relatively straightforward attack, for the exceptional occasion where the user is actually verifying the fingerprint, is to generate random keys until one of them has a fingerprint whose first few hex digits are close enough to the real thing to pass muster [21]

SSL Certificates: Indistinguishable from Placebo

The security model used with SSL server certificates might be called honesty-box security: In some countries newspapers and similar low-value items are sold on the street by having a box full of newspapers next to a coin box (the honesty box) into which people are trusted to put the correct coins before taking out a paper. Of course they can also put in a coin and take out all the papers, or put in a washer and take out a paper, but most people are honest and so most of the time it works. SSL's certificate usage is similar. If you use a \$495 certificate, people will come to your site. If you use a \$9.95 certificate, people will come to your site. If you use a \$0 self-signed certificate, people will come to your site. If you use an expired or invalid certificate, people will come to your site. If you're a US financial institution and use no certificate at all but put up a message reassuring users that everything is OK (see Figure 5), people will come to your site. In medical terms, the effects of this "security" are indistinguishable from placebo.



ONLINE SECURITY

Browser security indicators

You may notice when you are on our home page that some familiar indicators do not appear in your browser to confirm the entire page is secure. Those indicators include the small "lock" icon in the bottom right corner of the browser frame and the "s" in the Web address bar (for example, "https").

To provide the fastest access to our home page, we have made signing in to Online Services secure without making the entire page secure. Again, please be assured that your ID and password are secure.

Figure 5: Who needs SSL when you can just use a disclaimer?

In fact the real situation is even worse than this. There has in the past been plenty of anecdotal evidence of the ineffectiveness of SSL certificates, an example being the annual SecuritySpace survey, which reported that 58% of all SSL server certificates in use today are invalid without having any apparent effect on users of the sites [22]. However, it wasn't until mid-2005, ten years after their introduction, that a rigorous study of their actual effectiveness was performed. This study, carried out with computer-literate senior-year computer science students (who one would expect would be more aware of the issues than the typical user) confirmed the anecdotal evidence that invalid SSL certificates had no effect whatsoever on users visiting a site. Security expert Perry Metzger has summed this up, tongue-in-cheek, as "PKI is like real security, only without the security part".

It gets worse though. In one part of the study, users were directed to a site that used no SSL at all, at which point several of the users who had been quite happy to use the site with an invalid certificate now refused to use it because of the lack of SSL. Users assumed that the mere existence of a certificate (even if it was invalid) meant that it was safe to use the site, while they were more reluctant to use a site that didn't use SSL or certificates. This is quite understandable — no-one worries about an expired safety certificate in an elevator because all it signifies is that the owner forgot to get a new one, not that the elevator will crash into the basement and kill its occupants the next time it's used. In fact for the vast majority of elevator users the most that they'll ever do is register that some form of framed paperwork is present. Whether it's a currently valid safety certificate or an old supermarket till printout doesn't matter.

This real-world conditioning carries across to the virtual world. To quote the study, "the actual security of existing browsers is appalling when the 'human in the loop' is considered. Because most users dismiss certificate verification error messages, SSL provides little real protection against man-in-the-middle attacks. Users actually behaved less insecurely when interacting with the site that was *not* SSL-secured" [23]. The astonishing result of this research is that not only is the use of SSL certificates in browsers indistinguishable from placebo, it's actually *worse* than placebo because users are happy to hand over sensitive information to a site just because it has a certificate. If a medicine were to act in this way, it would be withdrawn from sale.

Another example of the clash of certificate theory with reality was reported by a security appliance vendor. Their products ship with a pre-generated self-signed certificate that ensures that they're secure out of the box without the user having to perform any additional certificate setup. Because it's a self-signed certificate, the user gets a certificate warning dialog from the browser each time they connect to the appliance, which in effect lets them know that the security is active. However, if they replace the self-signed certificate with an "official" CA-issued one, the browser warning goes away. Having lost the comforting SSL browser warning dialog, users were assuming that SSL was no longer in effect and complained to the vendor [24].

Again, users treated the (at least from a PKI theory point of view) less secure self-signed certificate setup as being more secure than the official CA-issued one.

A similar problem occurred during an experiment into the use of S/MIME signed email. When signed messaging was enabled, users experienced arcane PKI warning dialogs, requests to insert crypto cards, X.509 certificate displays, and all manner of other crypto complexity that they didn't much understand. This caused much apprehension among users, the exact opposite of the reassurance that signed email is supposed to provide. The conclusion reached was to "sign your messages only to people who understand the concept. Until more usable mechanisms are integrated into popular email clients, signatures using S/MIME should remain in the domain of 'power users'" [25]. Since a vanishingly small percentage of users really understand signed email, the actual message of the study is "Don't use signed email".

This result is very disturbing to security people. I've experienced this shock effect a number of times at conferences when I've mentioned the indistinguishable-from-placebo nature of SSL's PKI. Security people were stunned to hear that it basically doesn't work, and didn't know seem to know what to do with the information. A similar phenomenon has occurred with researchers in other fields as well. Inattention blindness, which is covered later on, was filed away by psychologists for over a quarter of a century after its discovery in 1970 because it was disturbing enough that no-one quite knew how to deal with it [26].

This situation isn't helped by the fact that even if PKI worked, obtaining bogus certificates from legitimate CA's isn't that hard. For example researcher David Mazieres was able to obtain a \$350 Verisign certificate for a nonexistent business by providing a Doing Business As (DBA) license [27], which requires little more than payment of the US\$10-\$50 filing fee. In case you're wondering why a DBA (referred to as a "trading as" license in the UK) has so little apparent security, it's deliberately designed this way to allow small-scale businesses such as a single person to operate without the overhead of creating a full business entity. DBAs were never intended to be a security measure, they were designed to make operating a small independent business easier (their effectiveness is indicated by the fact that the US alone had more than 20 million sole proprietorships and general partnerships recorded for the 2004 tax year). \$9.95 certificates are even less rigorous, simply verifying the ability to obtain a reply from an email address. How much checking do users expect the CA to do for all of \$9.95?

The User is Trusting... What?

CAs are often presented as "trusted third parties", but as security researcher Scott Rea has pointed out they're really just plain "third parties" because the user has no basis for trusting them [25], and for the large number of unknown CAs hardcoded into common applications they're explicitly untrusted third parties because the user doesn't even know who they are. Consider the dialog shown in Figure 6, in which the user is being told that they've chosen to trust a certain CA. Most users have no idea what a CA is, and they most certainly never chose to trust any of them. It's not even possible to determine who or what it is that they're so blindly trusting. The certificate, when the 'View' button is clicked, is issued by something claiming to be "Digi-SSL Xp" (whatever that is), and that in turn is issued by "UTN-USERFirst-Hardware" (ditto). In other words the user is being informed that they're trusting an unknown entity which is in turn being vouched for by another unknown entity. To paraphrase Douglas Adams, "This must be some strange new use of the word 'trust' with which I wasn't previously familiar".



Figure 6: Who are these people and why am I trusting them?

A contributing factor in the SSL certificate problem is the fact that the security warnings presented to the user that are produced by certificates often come with no supporting context. Danish science writer Tor Nørretranders calls this shared context between communicating parties “exformation” [28]. In the case of certificates there’s no certificate-related exformation shared between the programmer and the user. Even at the best of times users have little chance of effectively evaluating security risk [29] (even experts find this extraordinarily difficult, which is why it’s almost impossible to obtain computer security insurance), and the complete lack of context provided for the warning makes this even more difficult. Since web browsers implicitly and invisibly trust a large number of CAs, and by extension a vast number of certificates, users have no exformation that allows them to reason about certificates when an error message mentioning one appears. One user survey found that many users assumed that it represented some form of notice on the wall of the establishment, like a health inspection notice in a restaurant or a Better Business Bureau certificate, a piece of paper that indicates nothing more than that the owner has paid for it (which is indeed the case for most SSL certificates).

Similarly, the introduction of so-called high-assurance or extended validation (EV) certificates that allow CAs to charge more for them than standard ones is simply a case of rounding up twice the usual number of suspects — presumably somebody’s going to be impressed by it, but the effect on phishing will be minimal since it’s not fixing any problem that the phishers are exploiting. Indeed, cynics would say that this was exactly the problem that certificates and CAs were supposed to solve in the first place, and that “high-assurance” certificates are just a way of charging a second time for an existing service. A few years ago certificates still cost several hundred dollars, but now that you can get them for \$9.95 the big commercial CAs have had to reinvent themselves by defining a new standard and convincing the market to go back to the prices paid in the good old days. When you consider certificates using a purely financial perspective then from a large-company mindset (“cost is no object”) this may make some sort of sense but from an Internet mindset (“anything that costs is bypassed”), it’s simply not going to work. Not everyone can issue or afford these extra-cost certificates, and not everyone is allowed to apply for them — the 20 million sole proprietorships and general partnerships mentioned earlier are automatically excluded, for example. High-assurance certificates are a revenue model rather than a solution for users’ problems, with the end result being the creation of barriers to entry rather than the solution of any particular security problem.

Predictably, when the effectiveness of EV certificates was tested once Internet Explorer with its EV support had been around for a few months, they were found to have no effect on security [30]. One usability researcher’s rather pithy summary of

the situation is that “the EV approach is to do more of what we have already discovered doesn’t work” [31]. As with the 2005 study on the effectiveness of browser SSL indicators which found that users actually behaved less insecurely when SSL was absent, this study also produced a surprising result: Users who had received training in browser EV security behaved less securely than ones who hadn’t! The reason for this was that the browser documentation talked about the use of (ineffective, see other parts of this section) phishing warnings, and users then relied on these rather than the certificate security indicators to assess a site. As a result they were far more likely to classify a fraudulent site as valid than users who had received no security training. This unexpected result emphasises the importance of post-release testing when you introduce new security features, which is covered in more detail later in the section on security testing.

In order for a certificate-differentiation mechanism to work the user would need to have a very deep understanding of CA brands (recall that the vast majority of users don’t even know what a CA is, let alone knowing CA names and brands), and know which of the 100-150 CA certificates hard-coded into web browsers are trustworthy and which aren’t. No-one, not even the most knowledgeable security expert, knows who most of these CAs really are. The CA brands are competing against multi-million dollar advertising campaigns from established brands like Nike and Coke — it’s no contest [32].

The problem with CA branding (and lack of brand recognition) was demonstrated in the study of user recognition of CA brands discussed in the next section in which, of the users who actually knew what a CA was (many didn’t), far more identified Visa as a trusted CA than Verisign, despite the fact that Verisign is the world’s largest CA and Visa isn’t a CA at all [12]. Combine this with the previously-described user response to certificates and you have a situation where a bogus CA with a well-known brand like Visa will be given more weight than a genuine CA like Verisign. After all, what user would doubt <https://www.visa.com>, certified by Visa’s own CA?

In practice almost everything trumps certificate-based SSL security indicators. One large-scale study found, for example, that if users were presented with two identical pages of which one was SSL-protected and had a complex URL, <https://www.-accountonline.com/View?docId=Index&siteId=AC&langId=EN> and the other wasn’t secured and had a simple URL, <http://www.-attuniversalcard.com>, people rated the unprotected version with the simple URL as considerably more trustworthy than the protected one with the complex URL [33] (the unsecured page — note the different domains that the two are hosted in, even though they’re the same page — has since been updated to redirect to the secured page). Other factors that usability researchers have found will trump SSL indicators include:

- The complexity of the web page. Using fancy graphics and Flash animation exploits the watermark fallacy, in which users translate the use of complex features in physical objects that’s used for anti-counterfeiting of items like banknotes and cheques into an indication of authenticity in the virtual world.
- Pseudo-personalisation such as displaying the first four digits of the user’s credit card number, for example 4828-****-****-****, to “prove” that you know them. The first four digits are identical across large numbers of users and therefore relatively easy to anticipate. For attacks targeting the user bases of individual banks, it’s even easier because prefixes for all cards from that bank will be identical. For example when phishers spammed (possible) customers of the Mountain America credit union in Salt Lake City, they were able to display the first five digits of the card as “proof” of legitimacy because all cards issued by the bank have the same prefix [34] (in addition they used a legitimate CA-issued certificate to authenticate their phishing site).
- Providing an independent verification channel for information such as a phone number to call. This exploits the “not-my-problem” fallacy, no-one actually calls the number since they assume that someone else will. In addition phishers have

already set up their own interactive voice response (IVR) systems using VoIP technology that mimic those of the target bank, so having a phone number to call is no guarantee of authenticity [35][36].

Alongside these tricks, there are myriad other ways that are being actively exploited by phishers. Any of these factors, or factors in combination, can trump SSL security in the eyes of the users.

Password Mismanagement

The start of this section touched on the poor implementation of password security by applications, pointing out that both SSH and SSL/TLS, protocols designed to secure (among other things) user passwords, will connect to anything claiming to be a server and then hand over the user's password in plaintext form without attempting to apply even the most basic protection mechanisms. However, the problem goes much further than this. Applications (particularly web browsers) have conditioned users into constantly entering passwords with no clear indication of who they're handing them over to. These password mechanisms are one of the many computer processes that are training users to become victims of phishing attacks.

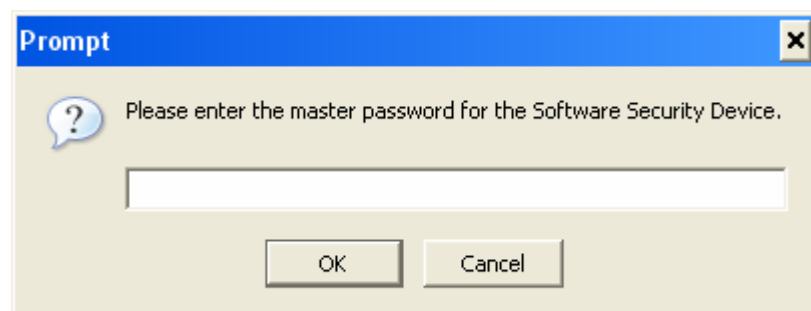


Figure 7: Gimme your password!

Consider the dialog in Figure 7, in this case a legitimate one generated by the Firefox browser for its own use. This dialog is an example of geek-speak at its finest. In order to understand what it's asking, you need to know that Netscape-derived browsers use the PKCS #11 crypto token interface internally to meet their cryptographic security requirements. PKCS #11 is an object-oriented interface for devices like smart cards, USB tokens, and PCMCIA crypto cards, but can also be used as a pure software API. When there's no hardware crypto token available, the browser uses an internal software emulation, a so-called PKCS #11 soft-token. In addition, the PKCS #11 device model works in terms of user sessions with the device. The default session type is a public session, which only allows restricted (or even no) access to objects on the device and to device functionality. In order to fully utilise the device, it's necessary to open a private session, which requires authenticating yourself with a PIN or password. What the dialog is asking for is the password that's required to open a private session with the internal PKCS #11 soft-token in order to gain access to the information needed to access a web site.

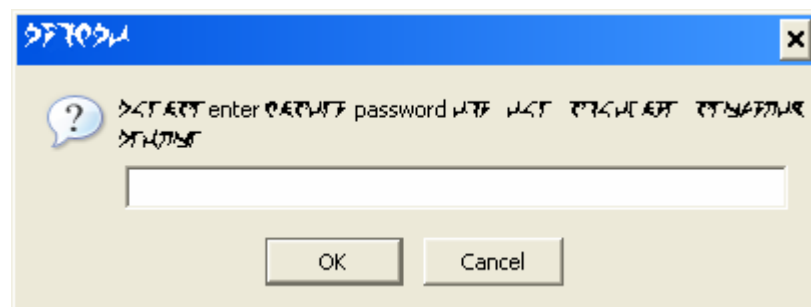


Figure 8: Password dialog as the user sees it

Possibly as much as one hundredth of one percent of users exposed to this dialog will understand that. For everyone else, it may as well be written in Klingon (see Figure

8). All they know is that whenever they fire up the browser and go to a web site that requires authentication, this garbled request for a password pops up. After an initial training period, their proficiency increases to the point where they're barely aware of what they're doing when they type in their password — it's become an automatic process of the kind described in the next chapter.

Other poorly-thought-out password management systems can be similarly problematic. The OpenID standard, a single-sign-on mechanism for web sites, goes to a great deal of trouble to remain authentication-provider neutral. The unfortunate result is what security practitioner Ben Laurie has termed “a standard that has to be the worst I've ever seen from a phishing point of view” [37] because it allows any web site to steal the credentials you use at any other web site. To do this, an attacker sets up a joke-of-the-day or animated-dancing-pigs or kitten-photos web page or some other site of the kind that people find absolutely critical for their daily lives, and uses OpenID to authenticate users. Instead of using your chosen OpenID provider to handle the authentication, the attacker sends you to an attacker-controlled provider that proxies the authentication to the real provider. In this way the attacker can use your credentials to empty your PayPal account while you're reading the joke of the day or looking at kitten pictures.

This is far worse than any standard phishing attack because instead of having to convince you to go to a fake PayPal site, the attacker can use any site at all to get at your PayPal credentials. What OpenID is doing is training users to follow links from random sites and then enter their passwords, exactly the behaviour that phishers want [38]. By declaring this problem “out of scope” for the specification [39], the developers of the OpenID standard get to pass the problem on to someone else. Other federated single-sign-on mechanisms like Internet2's Shibboleth exhibit similar flaws.

Future developments have the potential to make this situation even worse. If the biometrics vendors get their way, we'll be replacing login passwords with thumbprints. Instead of at least allowing for the possibility of one password per account, there'll be a single password (biometric trait) for all accounts, and once it's compromised there's no way to change it. Far more damaging though is the fact that biometrics makes it even easier to mindlessly authenticate yourself at every opportunity, handing out your biometric “password” to anything that asks for it. Articles proposing the use of biometrics as anti-phishing measures never even consider these issues, choosing to focus instead on the technical aspects of fingerprint scanning and related issues [40].

Other Languages, Other Cultures

Up until about fifteen years ago, it was assumed that there were universal maxims such as modes of conversation and politeness that crossed all cultural boundaries. This turned out to be largely an illusion, contributed to at least to some extent by the fact that most of the researchers who published on the subject came from an Anglo-Saxon, or at least European, cultural background.

Since then, the ensuing field of cross-cultural pragmatics, the study of how people interact across different cultures, has helped dispel this illusion. For example, the once-popular assumption that the “principles of politeness” are the same everywhere have been shown to be incorrect in ways ranging from minor variations such as English vs. eastern European hospitality rituals through to major differences such as cultures in which you don't thank someone who performs a service for you because if they didn't want you to accept the service they wouldn't have offered it, a practice that would seem extremely rude to anyone coming from a European cultural background.

Let's look at a simple example of how a security user interface can be affected by cross-cultural pragmatics issues. Imagine a fairly standard dialog that warns that something has gone slightly wrong somewhere and that if the user continues, their privacy may be compromised. Even the simple phrase “your privacy may be compromised” is a communications minefield. Firstly, the English term “privacy”

has no equivalent in any other European language. In fact the very concept of “privacy” reflects a very Anglo-Saxon cultural value of being able to create a wall around yourself when and as required. Even in English, privacy is a rather fuzzy concept, something that philosopher Isaiah Berlin calls a “negative liberty” which is defined by an intrusion of it rather than any innate property. Like the US Supreme Court’s (non-)definition of obscenity, people can’t explicitly define it, but know when they’ve lost it [41]. So in this case warning of a *loss* of privacy (rather than stating that taking a certain measure will increase privacy) is the appropriate way to communicate this concept to users — assuming that they come from an Anglo-Saxon cultural background, that is.

Next we have the phrase “may be”, a uniquely English way of avoiding the use of an imperative [42]. In English culture if you wanted to threaten someone, you might tell them that if they don’t take the requested action they might have a nasty accident. On the continent, you’d be more likely to inform them that they *will* have a nasty accident. Moving across to eastern Europe and Italy, you’d not only inform them of the impending accident but describe it in considerable and occasionally graphic detail.

The use of so-called whimperatives, extremely common in English culture, is almost unheard-of in other European languages [43]. A request like “Would you mind opening the window” (perhaps watered down even further with a side-order of “it’s a bit cold in here”) would, if you attempted to render it into a language like Polish, “Czy miałabyś ochotę ...”, sound quite bizarre — at best it would come across as an inquiry as to whether the addressee is capable of opening the window, but certainly not as a request.

Finally, we come to the word “compromise”, which in everyday English is mostly neutral or slightly positive, referring to mutual concessions made in order to reach agreement (there’s an old joke about a manager who wonders why security people are always worrying about compromise when everyone knows that compromise is a necessary requirement for running a business). In other languages the connotations are more negative, denoting weakness or a sell-out of values. Only in the specialised language of security-speak, however, is compromise an obviously negative term.

The fact that it’s taken four paragraphs just to explain the ramifications of the phrase “your privacy may be compromised” is a yardstick of how tricky the effective communication of security-relevant information can be. Even something as simple as the much-maligned “Are you sure?” dialog box can be problematic. In some cultures, particularly when offering hospitality, you never try to second-guess someone else’s wishes. A host will assume that the addressee should always have more, and any resistance by them can be safely disregarded (the authors of endless “Are you sure?” dialogs should probably take this attitude to heart). The common English question “Are you sure?” can thus sound quite odd in some cultures.

Japan has a cultural value called *enryo*, whose closest English approximation would be “restraint” or “reserve”. The typical way to express *enryo* is to avoid giving opinions and to sidestep choices. Again using the example of hospitality, the norm is for the host to serve the guest a succession of food and drink and for the guest to consume at least a part of every item, on the basis that to not do so would imply that the host had miscalculated the guest’s wishes. The host doesn’t ask, and the guest doesn’t request. When responding to a security-related dialog in which the user is required to respond to an uninvited and difficult-to-answer request, the best way to express *enryo* is to click ‘OK’. In a Japanese cultural context, the ‘OK’ button on such dialogs should really be replaced with one that states ‘*Nan-demo kaimasen*’, “Anything will be all right with me”. (In practice it’s not quite that bad, since the fact that the user is interacting with a machine rather than a human relaxes the *enryo* etiquette requirements).

So going beyond the better-known problems of security applications being localised for *xx-geek* by their developers, even speaking in plain English can be quite difficult when the message has to be accurately communicated across different languages and cultures. Some time ago I was working on an internationalised security application

and the person doing the Polish translation told me that in situations like this in which the correct interpretation of the application developer's intent is critical, he preferred to use the English version of the application (even though it wasn't his native language) because then he knew that he was talking directly with the developer, and not witnessing an attempt to render the meaning across a language and cultural barrier.

References

- [1] "Security Absurdity: The Complete, Unquestionable, And Total Failure of Information Security", Noam Eppel, <http://www.securityabsurdity.com/failure.php>.
- [2] "The Twelve Networking Truths", RFC 1925, Ross Callon, 1 April 1996.
- [3] "Requirements for IPsec Remote Access Scenarios", RFC 3457, Scott Kelly and Sankar Ramamoorthi, January 2003.
- [4] "Authentication Components: Engineering Experiences and Guidelines", Pasi Eronen and Jari Arkko, *Proceedings of the 12th International Workshop on Security Protocols (Protocols '04)*, Springer-Verlag Lecture Notes in Computer Science No.3957, April 2004, p.68.
- [5] "Straw poll on TLS SRP status", thread on ietf-tls mailing list, May-June 2007, <http://www1.ietf.org/mail-archive/web/tls/current/msg01667.html>.
- [6] "Man-in-the-Middle in Tunnelled Authentication Protocols", N. Asokan, Valtteri Niemi, and Kaisa Nyberg, Cryptology ePrint Archive, Report 2002/163, November 2002, <http://eprint.iacr.org/2002/163>.
- [7] "Man-in-the-Middle in Tunnelled Authentication Protocols", N. Asokan, Valtteri Niemi, and Kaisa Nyberg, *Proceedings of the 11th Security Protocols Workshop (Protocols '03)*, Springer-Verlag Lecture Notes in Computer Science No.3364, April 2003, p.29.
- [8] "The Compound Authentication Binding Problem", IETF draft `draft-puthenkulam-eap-binding-04`, Jose Puthenkulam, Victor Lortz, Ashwin Palekar, and Dan Simon, 27 October 2003.
- [9] "Computer Security in the Real World", Butler Lampson, keynote address at the 14th Usenix Security Symposium (Security'05), August 2005.
- [10] "An Analysis of the System Security Weaknesses of the US Navy Fleet Broadcasting System, 1967-1974, as exploited by CWO John Walker", Laura Heath, Master of Military Art and Science thesis, US Army Command and General Staff College, Ft. Leavenworth, Kansas, 2005.
- [11] "User-Centered Security", Mary Ellen Zurko, *Proceedings of the 1996 New Security Paradigms Workshop (NSPW'96)*, September 1996, p.27.
- [12] "Design Principles and Patterns for Computer Systems That Are Simultaneously Secure and Usable", Simson Garfinkel, PhD thesis, Massachusetts Institute of Technology, May 2005.
- [13] "Challenges in deploying low-latency anonymity (Draft)", Roger Dingledine, Nick Mathewson and Paul Syverson, 2005, <http://tor.eff.org/svn/trunk/doc/design-paper/challenges.pdf>.
- [14] "About Face 2.0: The Essentials of Interaction Design", Alan Cooper and Robert Reimann, John Wiley and Sons, 2003.
- [15] "Cabirn Fever", Peter Ferrie and Peter Szor, *Virus Bulletin*, August 2004, p.4.
- [16] "Security and Identification Indicators for Browsers against Spoofing and Phishing Attacks", Amir Herzberg and Ahmad Jbara, Cryptology ePrint Archive, <http://eprint.iacr.org/2004/>, 2004.
- [17] "Why Features Don't Matter Any More: The New Laws of Digital Technology", Andreas Pfeiffer, *ACM Ubiquity*, Vol.7, Issue 7 (February 2006), http://www.acm.org/ubiquity/views/v7i07_pfeiffer.html.
- [18] "Invalid banking cert spooks only one user in 300", Stephen Bell, ComputerWorld New Zealand, 16 May 2005, <http://www.computerworld.co.nz/news.nsf/NL/-FCC8B6B48B24CDF2CC2570020018FF73>.

- [19] “The heuristic-systematic model in its broader context”, Serena Chen and Shelly Chaiken, *Dual-Process Theories in Social Psychology*, Guilford Press, 1999, p.73.
- [20] “Information Foraging in Information Access Environments”, Peter Pirolli and Stuart Card, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (SIGCHI’95)*, May 1995, p.51.
- [21] “pattern recognition”, Dan Kaminsky, invited talk at Black Ops 2006 at the 20th Large Installation System Administration Conference (LISA’06), December 2006.
- [22] “Web Survey and Internet Research Reports by SecuritySpace”, http://www.securityspace.com/s_survey/data/, 2005. Note that the figures given in the survey results don’t add up to the value quoted in the text, since some certificates are invalid for multiple reasons and therefore appear in multiple categories.
- [23] “Hardening Web Browsers Against Man-in-the-Middle and Eavesdropping Attacks”, Haidong Xia and José Brustuloni, *Proceedings of the 14th international conference on the World Wide Web (WWW’05)*, May 2005, p.489.
- [24] Lucky Green, private communications, 10 December 2006.
- [25] “A Case (Study) For Usability in Secure Email Communication”, Apu Kapadia, *IEEE Security and Privacy*, **Vol.5, No.2** (March/April 2007), p.80.
- [26] “Sights unseen”, Siri Carpenter, *Monitor on Psychology*, **Vol.32, No.4** (April 2001), p.54.
- [27] “Self-certifying File System”, David Mazieres, PhD thesis, MIT, May 2000.
- [28] “The User Illusion: Cutting Consciousness Down to Size”, Tor Nørretranders, Penguin, 1999.
- [29] “Why Johnny Can’t Evaluate Security Risk”, George Cybenko, *IEEE Security and Privacy*, **Vol.4, No.1** (January/February 2006), p.5
- [30] “An Evaluation of Extended Validation and Picture-in-Picture Phishing Attacks”, Collin Jackson, Dan Simon, Desney Tan, and Adam Barth, *Proceedings of the Usable Security 2007 Conference (USEC’07)*, February 2007.
- [31] “Re: [hcisec] EV certificates and phishing”, James Donald <jamesd@echeque.com>, posting to the hcisec@yahoogroups.com mailing list, message-ID 45DD1784.5010606@echeque.com, 22 February 2007.
- [32] “Improving Authentication On The Internet”, Gervase Markham, <http://www.gerv.net/security/improving-authentication/>, 12 May 2005.
- [33] “The Human Factor in Phishing”, Markus Jakobsson, *Proceedings of the 6th National Forum on Privacy & Security of Consumer Information*, January 2007.
- [34] “The New Face of Phishing”, Brian Krebs, 13 February 2006, http://blog.washingtonpost.com/securityfix/2006/02/-the_new_face_of_phishing_1.html.
- [35] “Phishers try a phone hook”, Joris Evers, CNet News.com, 20 April 2006, http://news.com.com/Phishers+try+a+phone+hook/2100-7349_3-6066171.html.
- [36] “Phishers come calling on VoIP”, Joris Evers, CNet News.com, 10 July 2006, http://news.com.com/Phishers+come+calling+on+VoIP/-2100-7349_3-6092366.html.
- [37] “OpenID: Phishing Heaven”, Ben Laurie, 19 January 2007, <http://www.links.org/?p=187>.
- [38] “Phishing and OpenID: Bookmarks to the Rescue?”, Ka-Ping Yee, 20 January 2007, <http://usablesecurity.com/2007/01/20/phishing-and-openid/>.
- [39] “OpenID Authentication 2.0”, <http://openid.net/specs.bml>.
- [40] “A Touch of Money”, Anil Jain and Sharathchandra Pankanti, *IEEE Spectrum (INT)*, **Vol.43, No.7** (July 2006), p.14.
- [41] “Privacy and Freedom”, Alan Westin, Atheneum, 1967.
- [42] “Watching the English”, Kate Fox, Hodder & Stoughton Paperbacks, 2005.

- [43] “Cross-Cultural Pragmatics: The Semantics of Human Interaction (2nd ed)”,
Anna Wierzbicka, Walter de Gruyter, 2003.

The Psychology of Security Usability

Some of the problems mentioned in the previous chapter wouldn't be too surprising to cognitive psychologists, people who study the mental processes involved in how people understand, analyse, and solve problems. The field of psychology provides a great deal of insight into how people deal with security user interfaces. This chapter looks at how some of the human mental processes that are relevant to security work, and explores why security user interface elements often perform so poorly in the real world.

How Users Make Decisions

To help understand how we've got into this situation, it's useful to look at how the human decision-making process actually works. The standard economic decision-making model, also known as the Bayesian decision-making model, assumes that someone making a decision will carefully take all relevant information into account in order to come up with an optimal decision. Although this model has been tuned somewhat over time, for example by using Herbert Simon's concept of limited rationality to model decisions made from incomplete knowledge [1], it wasn't until the 1980s that the US Department of Defence helped dispel the illusion of the economic decision-making model when they sponsored research to try and find techniques for helping battlefield commanders make more effective decisions. This work showed that, contrary to expectations, people under pressure for a quick decision didn't weigh up the relative merits of a set of options and choose the most optimal one. They didn't even make a choice from a cut-down subset of options. Instead, they followed a process that's now called recognition-primed decision making (RPD) in which they generate options one at a time (without ever comparing any two), rejecting ones that don't work and going with the first one that does. Humans take this approach when we can't hold all of the necessary information in working memory, or can't retrieve the information needed to solve the problem from long-term memory, or can't apply standard problem-solving techniques within the given time limit. Whatever the case, once the scope of the problem exceeds the bounds on our rationality, we have to take shortcuts, and these shortcuts can lead to errors [2].

This approach to making decisions, called the singular evaluation approach, is something that you've probably encountered yourself in various forms. For example if you move house into a new area and know that you'll eventually need a plumber to install a sink for you, you have the luxury of being able to make a few inquiries about prices and work times, and perhaps look to neighbours for recommendations before you make your decision. On the other hand if your basement is under a metre of slowly-rising water, you'll go with the first plumber who answers their phone and can get there within 10 minutes. This is the singular evaluation approach.

Singular evaluation is used when the decision-maker is under pressure (a computer user wanting to get on with their job automatically falls into the time-pressure category, even if there's no overt external time pressure), when the conditions are dynamic (the situation may change by the time you perform a long detailed analysis), and the goals are ill-defined (most users have little grasp of the implications of security mechanisms and the actions associated with them) [3][4]. The reason why experts are better at this type of decision-making than the average person is that they're more likely to come up with a good option as their first choice than the typical person [5]. Research into how experts solve problems has also indicated that they tend to spend a considerable amount of time thinking about different representations of the problem and how to solve it, while novices simply go for the most obvious representation [6]. The experts were able to both cover more solution strategies and arrive at the final solution more quickly than the novices.

Psychological studies have shown that in the presence of external stimuli such as stress (or in this case the desire to get a job done, which is often the same as stress), people will focus on the least possible amount of evidence to help them make a quick

decision. Specifically, the external stimuli don't affect the way that we process information, but reduce our ability to gather information and the ability to use our working memory to sort out the information that we do have [7][8][9][10]. Thus flooding an expert with external stimuli eventually reduces their decision-making ability to that of a novice.

Psychologists distinguish between the two types of action taken in response to a situation as automatic vs. controlled processes [11][12]. Controlled processes are slow and costly in terms of the amount of mental effort required, but in compensation provide a great deal of flexibility in handling unexpected situations. Automatic processes in contrast are quick and have little mental overhead. While controlled processes are associated with deliberate action, automatic processes are essentially acting on autopilot. Because of this, automatic processing is inherently parallel (it's possible to handle multiple tasks at the same time) while controlled processing is strictly serial (you have to focus exclusively on the task at hand).

A good illustration of the difference between controlled and automatic actions is the difference between a novice and an experienced driver. A novice driver has to manually and consciously perform actions such as changing gears and checking the rear-view mirror, while for an experienced driver these actions occur automatically without any conscious effort. One characteristic of an automatic process is that it's triggered by a certain stimulus and, once begun, is very hard to stop, since there's no conscious effort involved in performing it. This makes automatic processes very hard to control: Present the right stimulus and the body reacts on its own. This is why people click away confirmation dialogs without thinking about it, or even being aware of what they're doing (lack of conscious awareness of an action is another characteristic of automatic processes).

The ability to sort out the relevant details from the noise is what makes it possible for humans to function. For example as you've been reading this you probably haven't noticed the sensation of your clothes on your skin until this sentence drew your attention to them. The entire human sensory and information-processing system acts as a series of filters to reduce the vast flow of incoming information to the small amount that's actually needed in order to function. Even the very early stages of perception involve filtering light and sound sensations down to a manageable level. Selective attention processes provide further filtering, giving us the ability to do things like pick out a single conversation in a crowded room, the so-called cocktail party phenomenon (or more formally the source separation problem) [13]. At the other end of the chain, forgetting discards non-meaningful or non-useful information.

Imagine if, instead of using singular evaluation, humans had to work through the implications of all possible facts at their disposal in order to come to a conclusion about everything they did. They would never get anything done. There exists a mental disorder called somatising catatonic conversion in which people do exactly this, over-analysing each situation until, like the 1960s operating system that spent 100% of its time scheduling its own operational processes, they become paralysed by the overhead of the analysis. Artificial intelligence researchers ran into exactly this problem, now called the frame problem, when they tried to recreate singular evaluation (or to use its more common name, "common sense") using computer software. The mechanistic approach resulted in programs that had to grind through millions of implications, putting all the relevant ones in a list of facts to consider, and then applying each one to the problem at hand to find an appropriate solution. Singular evaluation in humans isn't a bug, it's what makes it possible for us to function.

Usability researchers have already run into this issue when evaluating browser security indicators. When users were asked to carefully verify whether sites they were visiting were legitimate or not, the researchers had to abort the experiment after finding that users spent "absurd amounts of time" trying to verify site legitimacy [14]. On top of this, making users switch off singular evaluation lead to a false-positive rate of 63%, because when the users tried hard enough they would eventually find some reason somewhere to justify regarding the site as non-kosher. More worryingly, even after spending these absurd amounts of time trying to detect

problem sites, the users still failed to detect 36% of false sites using standard browser security indicators, no matter how much time they spend on the problem. As in the non-computer world, the use of singular evaluation is a basic requirement for users to be able to function, and a security user interface has to carefully take into account this human approach to problem-solving.

Consequences of the Human Decision-making Process

From a psychological perspective, judgemental heuristics in the form of automatic processes work well in most cases by saving users time, energy, and mental capacity when dealing with pointless popup dialogs, but suffer from the problem that attackers can take advantage of the click, whirr response that they engender to simulate undesirable (or desirable, from the attacker's point of view) behaviour from the user, tricking them into overriding security features in applications to make them vulnerable to attack. This aspect of user behaviour in response to SSL certificates is being exploited by phishers through the technique of "secure phishing", in which attackers convince users to hand over passwords and banking details to a site that must be OK because it has a certificate.

In 2005, the first year that records for this phenomenon were kept, over 450 such secure phishing attacks were discovered. These used a variety of techniques ranging from self-signed certificates and cross-site scripting and frame injection to insert content into real financial web sites through to the use of genuine certificates issued to sound-alike domains [15]. An example of the latter type of attack was the *visa-secure.com* one aimed at Visa cardholders. Since Visa itself uses soundalike domains such as *verifiedbyvisa.com* and *visabuxx.com* and the site was otherwise indistinguishable from a real Visa site, the phish proved to be extremely effective [16]. As Figure 9 shows, Visa isn't the only company with this problem.

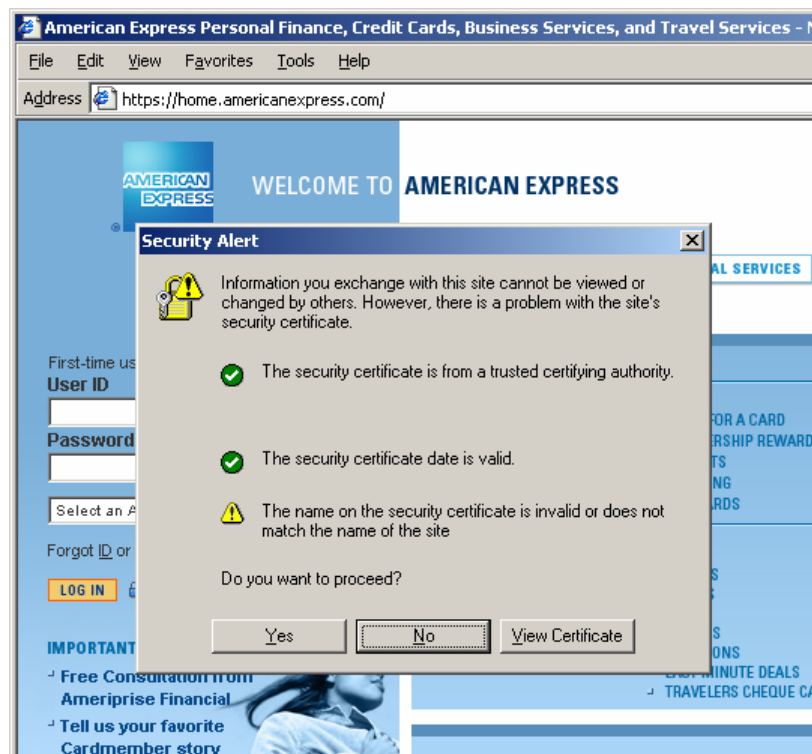


Figure 9: American Express certificate for a different site

The use of multiple completely unrelated domains is fairly common among financial institutions. Citibank for example uses alongside the obvious *citibank.com* six other unrelated domain names like *accountonline.com*, see XXX for one example of such a domain (with the wrong certificate). The domain *citibank-verify.4t.com* on the other hand is (or was) a phishing site, complete with a legitimate CA-issued certificate. Other domains in the "citibank" namespace alone

include citibank-america.com, citibank-credicard.com, citibank-credit-card.com, citibank-credit-cards.com, citibank-account-updating.com, citibank-creditcard.com, citibank-loans.com, citibank-login.com, citibank-online-security.com, citibank-secure.com, citibank-site.com, citibank-sucks.com, citibank-update.com, citibank-updateinfo.com, citibank-updating.com, citibankaccount.com, citibankaccountonline.com, citibankaccounts.com, citibankaccountsonline.com, and citibankbank.com, of which some are legitimate and some aren't. For example citibank-account-updating.com is owned by Ms. Evelyn Musa, ezayoweezay_halobye@yahoo.com. Another example of unrelated domain name usage is the Hanscom Federal Credit Union (serving the massive Hanscom air force base, a tempting target), which uses all of www.hfcu.org, locator.hfcu.org, ask.hfcu.org, calculators.hfcu.org, www.loans24.net, hfcu.mortgagewebcenter.com, secure.andra.com, secure.autofinancialgroup.com, hffo.cuna.org, www.cudlautosmart.com, www.carsmart.com, reorder.libertysite.com, www.ncua.gov, www.lpl.com, anytime.cuna.org, usa.visa.com, and www.mycardsecure.com. Although obfuscated names like hffo.cuna.org aren't (at least in this case) being run by Evelyn Musa in Nigeria, it's hard to see what relates something like "libertysite" to "Hanscom Federal Credit Union".

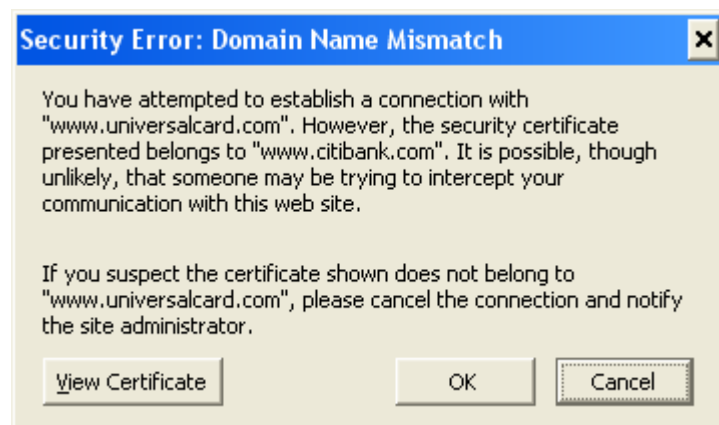


Figure 10: One of Citibank's many aliases

The reason why people are so bad at spotting these phishing attacks is that they're not very good at either generating testable hypotheses or designing tests that falsify hypotheses, a fact that con artists and salespeople have exploited for centuries, if not millennia [17][18]. Scientists on the other hand, a subgroup whose lives revolve around rationality and seeking the truth, know that good science consists of designing an experiment to try and demonstrate that a theory is wrong. For example a standard statistical technique consists of generating a null hypothesis as a sceptical reaction to the research hypothesis that the research is designed to investigate, and then proving it wrong. So if the research hypothesis postulates that "factor X is significant" then the null hypothesis would be "factor X is not significant", and the study or experiment would attempt to prove the null hypothesis wrong. This technique is used in statistics because sampling variation means that we can never prove a hypothesised value true (another set of samples might produce a slightly different result), but what we can do is determine whether there is evidence to rule out a hypothesised value. To put it more simply, it's much easier to tell someone that they're wrong than to tell them what the correct answer is.

The US Navy has addressed this inability to generate testable hypotheses in the reassessment of tactical decision making that occurred after the accidental shootdown of an Iranian civilian airliner in July 1988. Part of this reassessment included the

introduction of the so-called STEP cycle, which involves creating a Story (hypothesis), Testing the hypothesis, and then Evaluating the result [19]. In other words it makes the creation and application of testable hypotheses an explicit part of the tactical decision-making process.

An ability to focus on evidence that falsifies a hypothesis seems to be an innate aspect of geek nature. For example there's an interesting demonstration that's performed by the convenor of the New Security Paradigms Workshop to demonstrate that a group of geeks, when given a large amount of correct information and one item of incorrect information, will focus immediately on the one item that's wrong rather than the many items that are correct. The rest of the population however is far more likely to merely try to confirm their hypotheses, a dangerous approach since any amount of confirmatory evidence can be negated by a single piece of evidence that falsifies it.

One of the first investigations into the phenomenon of confirmation bias was carried out by Peter Wason using a type of task that has since gained fame (at least among psychologists) as the Wason selection task. In one common form, Wason's 2-4-6 task, subjects were given a sequence of three numbers such as { 2, 4, 6 } and asked to determine the rule that governed the sequence by generating a sequence of their own that they thought followed the rule. When they'd done this, they checked the accuracy of their prediction by asking the experimenter whether their estimation followed the actual rule. While the actual rule was a very simple "any ascending sequence", the subjects tried to come up with far more complex rules ("even numbers", { 4, 6, 8 }) and never really tried to disconfirm them ({ 4, 5, 6 }) [20].

The human tendency towards confirmation bias has been extensively explored by both philosophers and psychologists [21][22][23]. Psychologists have studied the phenomenon of humans cooking the facts to support the conclusions that they want to reach in great detail. For example when people are exposed to a test or evaluation that makes them look bad, they tend to seek out information that questions the validity of the test; conversely, when it makes them look good, they seek out information confirming its validity [24][25]. Psychologists call the practice of seeking out material that supports your opinions and avoiding material that challenges them dissonance-motivated selectivity. In one experiment to investigate this effect in which participants were asked to evaluate the effectiveness of capital punishment based on the outcomes of two studies, they chose whatever study produced the conclusion that matched their own personal beliefs on capital punishment and came up with detailed reasons for why the other study was flawed [26]. Subsequent studies showed that even trained scientists fell into this trap [27]. Like our physical immune system, we have a psychological immune system that allows us to feel good and cope with situations, and the above examples are instances of our psychological immune system at work [28].

An example of this inability to generate testable hypotheses was the alarming practice used by one user in a phishing study to determine whether a site was genuine or not: She entered her user name and password, and assumed that if the site allowed her in then it was the real thing, since only the genuine site would know her password (the same practice has been reported among other users) [29]. Hearing of practices like this makes security peoples' toes curl up. (Having said that though, if the security people had implemented password-based authentication properly in the first place then this would be a perfectly valid site-validity check).

Techies on the other hand are accustomed to living so far off the end of the bell curve that they can't see that this is a perfectly sensible site-legitimacy test for many users. Dismissing this issue with the comment that "well, computer programmers are just weird" (or at least have thought processes that are very different from those of the typical user) provides more than just a catchy sound bite though. If you look at the Myers-Briggs type indicator profile (MBTI, a widely-used psychometric for personality types), you'll find that a large proportion of programmers have *TJ traits (other traits in the profile like introvert vs. extrovert aren't relevant to this discussion). For example research has shown that *SF* personality types obtain less than half the score of the opposing-personality *NT* types in code reviewing

(debugging) tasks [30]. NT types are thought of as being “logical and ingenious”, with a particular aptitude for solving problems within their field of expertise. This same strong personality-type bias applies to the field of computer security as well, with security exhibiting a predominance of INTJ types [31]. This means that security people (and programmers in general) tend to have long attention spans, construct mental models in order to make sense of things, take time to order and process information before acting on it, and make decisions with their heads rather than their hearts [32][33][34]. Why does this make programmers weird? Because only 7% of the population have this particular personality profile³. In other words 93% of the users of the software that they’re creating have minds that handle the software very differently from the way that its creators do.

Hand-in-hand with the confirmation-bias problem is the disconfirmation bias problem, the fact that people are more likely to accept an invalid but plausible conclusion than a valid but implausible one [35]. In other words people will believe what they want to believe, and the beliefs themselves are often based on invalid reasoning. This is why people are ready to accept a site that looks and behaves exactly like their bank’s home page, but that’s hosted in eastern Europe — there must be a transient problem with the server, or the browser has got the URL wrong, or something similar.

This user conditioning presents a somewhat difficult problem. Psychologists have performed numerous studies over the years that examine people’s behaviour once they’ve become habituated into a particular type of behaviour and found that, once acquired, an (incorrect) click, whirr response is extremely difficult to change, with users resisting attempts to change their behaviour even in the face of overwhelming evidence that what they’re doing is wrong [36]. Gestalt psychologists call this phenomenon “Einstellung”, which can be translated as “set” or “fixity” but is better described using the more recent terminology of an inability to think outside the box. Any number of brain-teaser puzzles take advantage of the human tendency to become locked into a certain Einstellung/set from which it’s very hard to break free. For example one party game that exploits this involves having the organiser place a blanket or sheet over a participant and telling them that they have something that they don’t need and should hand over to the organiser. The participants typically hand over all manner of items (including, if it’s that sort of party, their clothing) before they realise that the unneeded item is the blanket that’s covering them — their Einstellung has blinded them to seeing this, since the blanket functions as a cover and thus doesn’t come into consideration as a potential discardable item.

Software vendors have in the past tried to work around users’ Einstellung, the tendency to stick with whatever works (even if it works really badly) by trying to “cure” them of the habit with techniques like a tip-of-the-day popup and, notoriously, the MS Office paperclip, but the success of these approaches has been mixed at best.

A standard theme in the psychological literature is the recognition that humans are primarily pattern recognisers (click, whirr) rather than analytical problem solvers, and will attempt to solve any problem by repeatedly applying context-specific pattern recognition to find a solution before they fall back to tedious analysis and optimisation. The psychological model for this process is the generic error modelling system (GEMS), in which someone faced with a problem to solve first tries repeated applications of a rule-based approach (“if (situation) then (action)”) before falling back to a knowledge-based approach that requires analysing the problem space and formulating an appropriate course of action. This fallback step is only performed with extreme reluctance, with the user trying higher and higher levels of abstraction in order to try and find some rule that fits before finally giving up and dropping back to a knowledge-based approach [37].

It’s therefore important to not dismiss the click, whirr response as simply grumbling about lazy users. It’s not even grumbling about lazy users with psychological

³ When I was a student, a sociologist gave one of the Computer Science years an MBTI test. The results were a singularity way off the end of the bell curve. I have no idea what she did with the results, although I’m sure the term “anomalous” appeared in her report.

justification for the grumbling. This is a statement of fact, an immutable law of nature that you can't ignore, avoid, or "educate" users out of. Click, whirr is not the exception, it's the environment, and you need to design your user interface to work in this environment if you want it to work at all.

Going beyond the genetically-acquired resistance to some types of security measures, security policies often expect us to behave in ways that are contrary to deeply-ingrained social conditioning. For example when we pass through a door, social etiquette demands that we hold it open for anyone following us. Security demands that we slam it in their face to prevent tailgating. Security policies at workplaces often require that we behave in ways that are perceived to be, as one study of user's reactions puts it, "paranoid" and "anal" [38]. Going beyond the usual indifference to security measures, security requirements that conflict with social norms can meet with active resistance from users, who are unlikely to want to aspire to an image of being an anal-retentive paranoid.

Security and Rationality

As psychologist James Alcock reminds us, our brains evolved to increase our chances for survival and reproduction, not to automatically seek the truth [39]. Quick and dirty techniques that more or less work serve evolutionary goals better than purely rational ones that require more time and effort [40].

For example a phishing study found that users were able to rationalise almost any kind of site-misdirection with reasons like `www.ssl-yahoo.com` being a "subdirectory" of Yahoo!, `sign.travelocity.com.zaga-zaga.us` being an outsourcing site for `travelocity.com`, the company running the site having to register a different name from its brand because the name was already in use by someone else, other sites using IP addresses instead of domain names so this IP-address-only site must be OK, other sites using redirection to a different site so this one must be OK, and other similar rationalisations, many taken from real-world experience with legitimate sites [41].

This occurs because humans are very good at rationalising away inconsistencies. In various experiments carried out on medical patients who had had the physical connection between their brain hemispheres severed in order to treat severe epileptic attacks (in medical terms a corpus callosotomy, in layman's terms a split brain), the left brain hemisphere was able to rationalise away behaviour initiated by the right hemisphere even though it had no idea what the other half of the brain was doing or why it was doing it [42]. This is a specialised instance of a phenomenon called illusory correlation in which people believe that two variables are statistically related even though there's no real connection. In one early experiment into the phenomenon, researchers took pictures drawn by people in a psychiatric institution and matched them at random to various psychiatric symptoms. Subjects who examined the randomly-labelled pictures reported all manner of special features in the various drawings that were indicative of the symptom [43].

This remarkable ability to fill in the gaps isn't limited to cognitive processes. Humans have a blind spot on the back of their retinas where the optic nerve attaches to the eyeball, which means that there are no visual receptors there to register an image. We never notice this in practice though because our brains invent something from the surrounding image details and use it to fill in the gap.

A similar phenomenon occurs with hearing. Researchers have carried out experiments where they partially obliterated a word with a cough and then used it in various sentences. Depending on the surrounding context, participants "heard" completely different words because their minds had filled in the obliterated detail so smoothly that they genuinely believed that they'd heard what their minds were telling them [44][45].

Security at Layers 8 and 9

Users are in general unmotivated and will often choose the path of least resistance even if they know that it's less secure. In other words, security is very rarely the user's priority, and if it gets in the way they'll avoid it. For example, students at Dartmouth College in the US preferred using passwords on public PCs even though far more (theoretically) secure USB tokens had been made freely available by the college, because passwords were more convenient [46]. In general, users dislike being forced to use a particular interface, with one Gartner group survey finding that although users claimed that they wanted more security, when it came down to it they really wanted to stick with passwords rather than going to more (theoretically) secure solutions like smart cards and RSA keys [47].

This has led to some ingenious efforts to make this style of token more convenient for everyday use. The SecurID fob camera, in which a user got tired of having to have a SecurID token on him and “solved” the usability problem by placing it under a webcam that he could access via any web browser, is one such example [48]. More recently this approach has been extended with OCR software, removing the human from the loop [49]. Extending this a step further, the ultimate user-friendly authentication token would be one with a built-in web server that allows its output to be automatically retrieved by anything needing authentication information. Although this is meant as a tongue-in-cheek observation, this is more or less the approach used by single-sign-on initiatives like OpenID, with security consequences that have already been discussed.

It's not that users are universally unmotivated, it's that they're unmotivated to comply with security measures that they don't understand — passwords and smart cards provide the same function, so why use the more complex one when the simpler one will do just as well? Most users are in fact reasonably security-conscious *if they understand the need for the security measures* [50]. As the section on theoretical vs. effective security pointed out, users need to be able to understand the need for a security measure in order to apply it appropriately.

Consider the case of 802.11 (in)security. After a German court in Hamburg found the owner of an open (insecure) 802.11 LAN responsible for its misuse by a third party (this was in response to a music industry lawsuit, so the legal angle is somewhat skewed), one user complained in a letter to a computer magazine that “My WLAN is open [insecure] in order to make it useful. Everyone who's used a WLAN knows this [...] misuse of a WLAN requires a considerable amount of criminal energy against which I can't defend myself, even if I use encryption” [51]. The magazine editors responded that one mouse click was all the criminal energy that it took to misuse an open 802.11 access point. This comment, coming from a typical user, indicates that they both have no idea how easy it is to compromise an open 802.11 LAN, and no idea that using encryption (at least in the form of WPA, not the broken WEP) could improve the situation. In other words they didn't want to apply security measures because they had no idea that they were either necessary or useful.

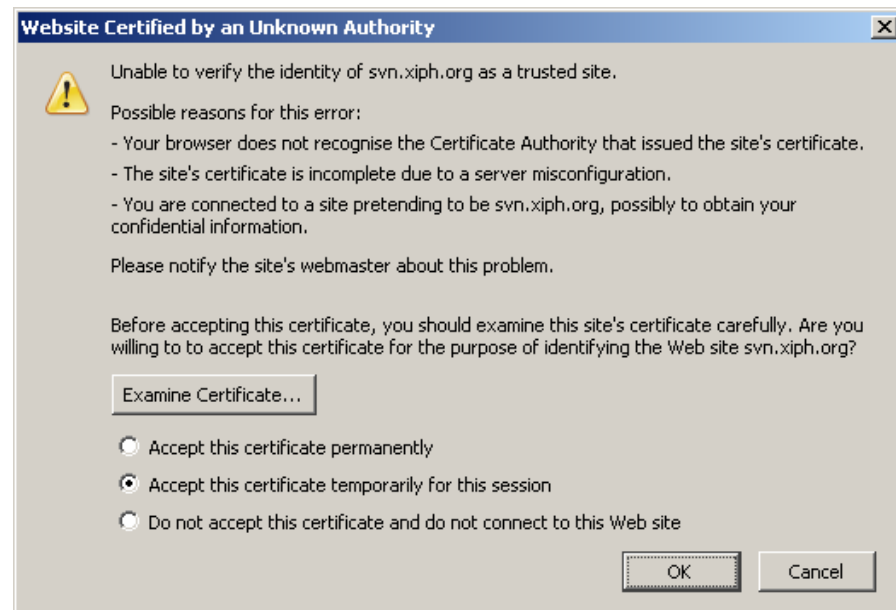


Figure 11: What the developers wrote

Similarly, users won't pay much attention to a security user interface feature unless it's a part of the critical action sequence, or in plain English an integral part of what they're doing. This is why almost no-one bothers to check site certificates on web sites, because it's not an essential part of what they're trying to accomplish, which is to use the web site. If a user is trying to perform task A and an unexpected dialog box B pops up (Figure 11), they aren't going to stop and carefully consider B. Instead, they're going to find the quickest way to get rid of B so that they can get back to doing their intended task A (Figure 12).



Figure 12: What the user sees

This is reflected in studies of the effectiveness of security user interface elements in web browsers. Carried out on a cross-section of university-educated computer users who were aware in advance (via the study's consent form) that they were taking part in a security usability evaluation study, it found that all of the standard browser security indicators were incapable of effectively communicating the security states to the user: 65% ignored the padlock, 59% paid no attention to the `https://` in the address bar, 77% didn't notice the Firefox address bar SSL indicator (and of the few who did notice it, only two users actually understood its significance), and when presented with an invalid-certificate warning dialog, 68% immediately clicked 'OK' without reading the dialog. Of the total number of users in the study, just one single user was able to explain what they'd just done when they clicked on the dialog [29].

Another study found that not a single user checked the certificate when deciding whether a site was secure or not [52]. Other studies that examined user's abilities to detect non-SSL-protected or spoofed sites found similar results: browser security indicators simply don't work, with one study of experienced computer users finding that only 18% of them could correctly identify an unprotected (no SSL vs. SSL-protected) site, concluding that "current browser indicators are not sufficient for security" [14].

An extreme example of the click, whirr response occurs with EULAs (End-User License Agreements for software), which no-one ever reads because all that they do is stall progress when setting up an application. Usability researchers have performed an experiment in which they expended considerable effort to make the EULA easier to read, but found that this didn't help because users still didn't read it [53]. Spyware and malware developers take advantage of this fact when they install their malware on a PC with the user's "permission". Probably the best approach to this problem is the EULalyzer, a scanner that scans EULAs for trigger words and phrases and alerts the user if any are present [54]. The fact that EULAs have become an arms race between vendors' lawyers and users is an indication of just how dysfunctional this mechanism really is.

Copyright notices at the start of a videotape or DVD run into the same problem as EULAs, with users either fast-forwarding through them on their VCRs or ignoring them after film studios forced DVD player manufacturers to disable fast-forward while the copyright notice was being displayed. Film enthusiasts will go so far as to re-master DVDs just to get rid of the annoying messages that interrupt their enjoyment of the film that they've bought. Users want to see a film (or run an application), and reading a legal notice is just an impediment to doing this. For example in the EULA study, typical user feedback was "No matter what you do, eventually I'm going to ignore it and install the software anyway".

This phenomenon is known to user interface developers as the "dancing bunnies problem", users will do whatever it takes to see the dancing bunnies that an email message is telling them about [55]. In one phishing study, nearly half of the users who fell victim to phishing sites said that they were concentrating on getting their job done rather than monitoring security indicators, with several noting that although they noticed some of the security warnings, they had to take some risks in order to get the job done [56]. Something similar happened during usability testing of a password-manager plugin for the Firefox browser, users simply gave up trying to use the password manager rather than looking to the documentation for help [57].

Microsoft has encountered this type of problem in its automatic security update system for Windows, which automatically downloads security updates without requiring the process to be initiated by the user, since most users never bother to do so. However, before silently installing updates, Windows tells the user what's about to happen. Microsoft found that considerable numbers of users were simply clicking 'Cancel' or the window-close control whenever it popped up because all they wanted was for the dialog to go away [58]. Once habituation had set in, this became an automatic action for any popups that appeared. Apart from the usual problem of user reactions to such dialogs, an extra contributing factor in this case would have been the fact that many Windows machines are so riddled with adware popups that users treated the security update dialog as just another piece of noise to be clicked away.



Figure 13: Desktop noise to be clicked away

Another situation where the click, whirr response occurs is with the copy of the Norton/Symantec security software that seems to come standard with any computer purchased from a large vendor like Dell, Gateway, or HP (the software vendors pay the computer companies up to US\$3 per desktop icon to get their products into the customer's focus, helping to subsidise the cost of the computer). Since the software is sold on a subscription basis it expires after a year leaving the computer unprotected, doubly so because it deactivates the Windows firewall by its presence. The results, as illustrated in Figure 13, are predictable: "a large proportion of these [virus-infected] systems had some form of Norton AV installed, and EVERY SINGLE ONE had a virus subscription which had lapsed. Entirely useless in protecting those computers" [59]. Like Windows Update, the Symantec nag screen habituates people into dismissing it without thinking, even more so because it's demanding time and money from the user rather than merely asking permission to install. Although this is more a business-model issue than a security usability one, it's worth noting at this point that using the subscription model to sell security software may be wonderful for the bottom line, but it's terrible for security.

One minor aid in trying to fix this problem is to remove the window-close control on the dialog box, providing a roadblock to muscle memory for users who have got into the habit of automatically clicking close to get rid of any pop-ups (even without this motivation, putting close boxes on dialogs is an interface design blooper because it's not clear whether clicking the close control represents an 'OK' or 'Cancel' action for the dialog). The additional step of making the dialog modal forces the user to pay attention to it. For a while in the 90s, modal dialogs were regarded as Evil, and so application developers went to great lengths to avoid them. As a result, far too many applications allow users to pile up a stack of (ignored) non-modal dialogs while ploughing ahead in an unsafe manner.

Unfortunately, this isn't possible in all circumstances. For example in extensive usability testing, Microsoft found that so many users were becoming trapped by badly-designed wizards created by third-party vendors that they had to remove the ability to disable the Cancel button and Close controls on wizards in order to protect users against poorly-designed applications [60].

A better approach to the problem, used by Apple in OS X, is to launch a full-blown application (in this case Software Update) in an attempt to garner more respect from the user. Apple also distinguishes security updates from general software updates, allowing users to apply only critical fixes and leave their system otherwise untouched, since many users are reluctant to make changes for fear of "breaking something".

If you redesign your application to get rid of unnecessary warning dialogs, you need to be careful how the replacement functionality works. For example the Firefox browser developers (and as a follow-on effect some developers of Firefox extensions) made a conscious effort to deprecate warning dialogs in place of notification ribbons that appear at the top or bottom border of the window to inform users that the browser or extension has blocked some potentially malicious action. Unfortunately the implementation of the ribbon sometimes fails to follow through on the optimised design since it merely provides a shortcut to the usual dialog-based interface. For example the ribbon that Firefox displays when it blocks a popup or prevents the installation of an extension leads to the full edit-site-permissions dialog in the browser's options menu. As a result, if the user wants to allow a one-off install of a component, they have to add the site as a trusted site, add the component, navigate down through the browser menus to the trusted-site dialog (which they may not even know exists, since it's only presented in response to clicking on the ribbon), remember which site they've just added, and remove it again. Apart from being a pain to do (users will invariably leave a site permanently in the trusted-sites list rather than go through the rigmarole of removing it again), this also leads to a race-condition attack in which a site installs a harmless extension and then, in the time it takes to turn site installs off again, installs a more malicious one. Alternatively, a malicious site can simply rely on the fact that for most users it'll be too much bother to remove the site again once they've added it, leaving them open to future malicious content from the site. A better approach would have been to allow the site's action on a one-off basis for just that action and no other, something that's already done by some of the many threat-blocking plugins that exist for Firefox.

The 'Simon Says' Problem

Related to this issue is what usability researcher Ka-Ping Yee has called the "Simon Says problem". In the children's game of the same name, users are expected to do what a leader tells them when they precede the order with "Simon says...", but to change their behaviour in the absence of the "Simon says" phrase. In other words users are expected to react to the *absence* of a stimulus rather than its presence, something that anyone who's ever played the game can confirm is very difficult. This problem is well-known to social psychologists, who note that it's one of the things that differentiate novices from experts — an expert will notice the absence of a particular cue while a novice won't, because they don't know what's supposed to happen and therefore don't appreciate the significance of something when it doesn't happen.

Psychologists have known about the inability of humans to react to the absence of stimuli for some time. In one experiment carried out more than a quarter of a century ago, participants were shown sets of trigrams (groups of three letters) and told that one of them was special. After seeing 34 sets of trigrams on average, they were able to figure out that the special feature in the trigram was that it contained the letter T. When this condition was reversed and the special trigram lacked the letter T, no-one was ever able to figure this out, no matter how many trigrams they saw [61]. In other words they were totally unable to detect the absence of a certain stimulus. Unfortunately this lack of something happening is exactly what web browsers expect users to respond to: a tiny padlock indicates that SSL security is in effect, but the *absence* of a padlock indicates that there's a problem.

Another contributing factor towards the Simon Says problem is the fact that people find negative information far more difficult to process than positive information [62][63]. This problem is well known among educational psychologists, who advise educators against using negative wording in teaching because information is learned as a series of positively-worded truths, not a collection of non-facts and false statements [64]. Consider the following propositional calculus problem:

If today is not Wednesday then it is not a public holiday.
Today is not a public holiday.

Is today not Wednesday? Research has shown that people find negative-information problems like this much harder to evaluate than positive-information ones ("If today

is Wednesday ...”), and are far more likely to get it wrong. Now compare this to the problem presented by browser security indicators, “If the padlock is not showing then the security is not present”. This is that very problem form that psychological research tells us is the hardest for people to deal with!

Contributing to the problem is the fact that the invisibly secured (via SSL) web browser looks almost identical to the completely unsecured one, making it easy for the user to overlook. In technical terms, the Hamming weight of the security indicators is close to zero. This has been confirmed in numerous studies. For example one study, which went to some trouble to be as realistic as possible by having users use their own accounts and passwords and giving them browser security training beforehand, report a 100% failure rate for browser HTTPS indicators — not one user noticed that they were missing [65]. Another example of an indicator with insufficient Hamming weight is the small warning strip that was added to Internet Explorer 6 SP2, with one usability test on experienced users and developers finding that no-one had noticed its presence [66]. Another test that examined the usability of password managers found that no-one noticed the fact that the password manager changed the background colour of password fields to indicate that the password had been secured [57].

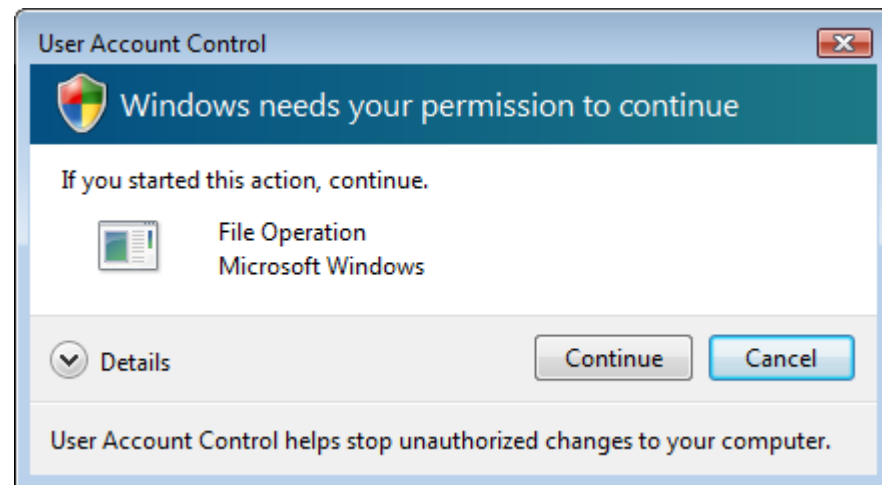


Figure 14: Vista UAC dialog

Another (informal) evaluation of Windows Vista’s (much-maligned) User Account Control (UAC) dialog, of which an example is shown in Figure 14, found that not one user noticed that the UAC dialog title had different colours in different situations [67], let alone knowing what it was that the different colours signified. Neither the official Microsoft overview of UAC in Microsoft Technet [68], nor popular alternative information sources like Wikipedia [69] even document the existence of these colour differences, let alone indicating what they mean. It requires digging deep down into an extremely long and geeky discussion of UAC [70] to find that a red title means that the application is blocked by Windows Group Policy, blue/green means that it’s a Vista administrative application, grey means that it’s an Authenticode signed application, and yellow means that it’s not signed. Bonus points if you can explain the significance of those distinctions.

This problem isn’t confined solely to security indicators. In one user test, the status bar on the spreadsheet application being tested would flash the message “There is a \$50 bill taped to the bottom of your chair. Take it!”. After a full day of user testing, not one user had claimed the bill [71]. A better-known example of the phenomenon, which has been used in a number of pop-psychology TV programs, was demonstrated by 2004 Ig Nobel prize winners Daniel Simons and Christopher Chabris in a 1999 experiment in which test subjects were asked to observe a video of a people playing basketball in front of three elevator doors. Halfway through the video, a tall woman carrying an umbrella or a person dressed in a gorilla suit (both obviously non-players) walked across the scene. Only 54% of the test subjects noticed [72].

This amazing phenomenon, in which people are unable to perceive unexpected objects, is known as inattention blindness. One very common situation in which this occurs is on the road, where drivers are looking for other cars and (on some streets) pedestrians, but are unable to register the presence of unexpected objects. Cyclists and motorbike riders were all too familiar with this problem decades before it even had a name because they found that they were more or less invisible to the drivers that they shared the roads with. A simple change to a motorbike such as mounting a pair of driving lights relatively far apart on a bike can greatly improve your “visibility” to drivers (at the expense of making your bike look really ugly) because now you match the visual pattern “car” rather than the invisible “not-a-car”. The first major work to explore this area concluded that “there is no conscious perception without attention” [73]. If people aren’t specifically looking for something like a security indicator then most of them won’t see it when it appears.

Over several million years of human evolution, we have learned to focus our attention on what’s important to us (things like imminent danger) and filter out irrelevant details. Human perception therefore acts to focus us on important details and prevents us from being distracted by irrelevant (or irrelevant-seeming) noise [74]. Over time, humans have learned to instinctively recognise obvious danger indicators like snakes, flashing red lights, and used-car salesmen, and can react automatically to them without having to stop and think about it. Psychologists have found that subjects who have never even seen something like a snake before are still instinctively afraid of it the first time that they’re shown one. Having your application flash up a photo of a cobra about to strike probably isn’t a good idea though.

On the other hand people pay scant attention to the lack of a padlock because it’s both unobvious and because it’s never been something that’s associated with danger. After all, why would a computer allow them to simply go ahead and perform a dangerous operation? Would someone build a house in which the power was carried by exposed copper wiring along the walls, with a little lightning-bolt icon down at ground level to warn users of the danger of electrocution? If they did, how long would they stay in business?

Even the more obvious indicators like the security toolbars that are available as various forms of browser plugin have little additional value when it comes to securing users (and that’s assuming that the toolbars themselves aren’t the source of security holes [75]). A study of the effectiveness of a range of these toolbars on university-educated users who had been informed in advance that they were taking part in a phishing study (informed consent is an ethical requirement in studies on human subjects) found that an average of 39% of users were fooled by phishing sites across the entire range of toolbars [76]. Without this advance warning the figures would be far worse, both because users wouldn’t specifically be on the lookout for phishing attacks and more importantly because most users wouldn’t notice the toolbars and if they did would have had little idea what they signified.

Browser security indicators

You may notice when you are on our home page that some familiar indicators do not appear in your browser to confirm the entire page is secure. Those indicators include the small "lock" icon in the bottom right corner of the browser frame and the "s" in the Web address bar (for example, "https").

To provide the fastest access to our home page for all of our millions of customers and other visitors, we have made signing in to Online Banking secure without making the entire page secure. Again, please be assured that your ID and passcode are secure and that only Bank of America has access to them.



Figure 15: Conditioning users to become victims of phishing attacks

Problems in motivating people to think about security also occur at the service provider side. Most US banking sites are still using completely insecure, unprotected logins to online banking services because they want to put advertising on their home pages (low-interest home loans, pre-approved credit cards, and so on) and using SSL to secure them would make their pages load more slowly than their competitors'. This practice has been widely decried by security experts for years and has even been warned about by browser vendors [77] without having any noticeable effect on the banks' security practices.

Browsers will actually warn users of this problem, but since the warning pops up whenever they enter any information of any kind into their browser, and includes an enabled-by-default "Don't display this warning again" setting (see the earlier discussion of this issue), the warning is long since disabled by the time the user gets to their banking page [78].

Even more frighteningly, US financial institutions are actively training users to become future victims of phishing attacks through messages such as the ones shown in Figure 15 and Figure 16 (this practice is depressingly common, these two examples are representative of a widespread practice). More recently, they've come up with a new twist on this by training users to ignore HTTPS indicators in favour of easily-spoofed (and completely ineffective) site images, a practice covered in more detail in the section on usability testing. This illustrates that not only end-users but also large organisations like financial institutions completely misunderstand the nature of SSL's certificate-based security model and what it's supposed to achieve.

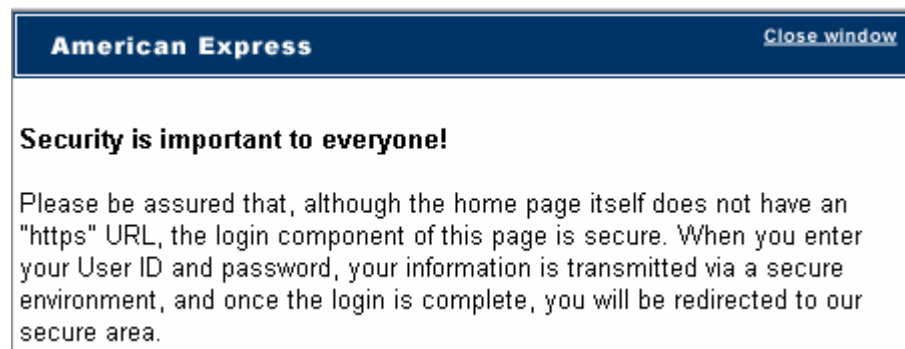


Figure 16: More user conditioning

In contrast, the use of un-secured online banking logins is almost unheard of outside the US, when banks are more conscious of customer security. In some countries there were concerted efforts by all banks to ensure that they had a single, consistent, secure interface to all of their online banking services, although even there it

occasionally lead to intense debate over whether security should be allowed to override advertising potential. When you're planning your security measures, you should be aware of the conflicting requirements that business and politics will throw up, often requiring solutions at the business or political rather than the technological level.

User Education, and Why it Doesn't Work

Don't rely on user education to try and solve problems with your security user interface. Computer security is simply too complicated, and the motivation for most users to learn its intricacies too low, for this strategy to ever work. Nobody wants to read instruction manuals, even if they're in the form of pop-up dialogs. Studies of real-world users have shown that they just aren't interested in having to figure out how an application works in order to use it. Furthermore, many concepts in computer security are just too complex for anyone but a small subset of hardcore geeks to understand. For example one usability study in which technology-savvy university students were given 2-3 page explanations of PKI technology (as it applied to SSL) found that none of them could understand it, and that was after reading a long explanation of how it worked, a point that the typical user would never even get to [79]. Before the PGP fans leap on this as another example of X.509's unusability, it should be mentioned that PGP, which a mere 10% of users could understand, fared little better.

These results have been confirmed again and again by experiments and studies across the globe. For example one two-year trial in Italy, which tried to carefully explain the security principles involved to its users, received feedback like "please remove all these comments about digital certificates etc., just write in the first page 'protected by 128bit SSL' as everybody else does" [80].

This lack of desire and inability to understand applies even more to something where the benefits are as nebulous as providing security, as opposed to something concrete like removing red-eye from a photograph. When confronted with a user interface, people tend to scan some of the text and then click on the first reasonable option, a technique called satisficing that allows users to find a solution that both satisfies and suffices (this is a variation of the singular evaluation approach that we encountered earlier). As a result, they don't stop to try and figure out how things work, they just muddle through [81]. The French have formalised this process under the name "le système D", where the D stands for "se débrouiller", meaning "to muddle through".

In addition to applying système D, users don't really appear to mind how many times they click (at least up to a point), as long as each click is an unambiguous, mindless choice [82]. People don't make optimal choices, they satisfice, and only resort to reading instructions after they've failed at several attempts to muddle through.

Attackers will then take advantage of this complexity, lack of user understanding, and user satisficing, to sidestep the security measures. For example when users, after several years of effort, finally learned that clicking on random email attachments was dangerous, attackers made sure that the messages appeared to come from colleagues, friends, trading partners, or family. For example AOL reported that in 2005 six of the top ten spam subject lines fell into this category [83], completely defeating the "Don't click on attachments from someone you don't know" conditioning. In addition to this problem, a modern electronic office simply can't function without users clicking on attachments from colleagues and trading partners, rendering years of user education effort mostly useless.

A better use of the time and effort required for user education would have been to concentrate on making the types of documents that are sent as attachments purely passive and unable to cause any action on the destination machine. A generalisation of this problem is that we have Turing machines everywhere — in the pursuit of extensibility, everything from Word documents to web site URLs has been turned into a programming language (there's even a standards group that manages the creation of such embedded Turing machines [84]).

Since many of these embedded Turing machines don't look anything like programming languages, it's very difficult to disable or even detect their use. A better alternative to trying to screen them would be to only allow them to be run in a special least-privileges context from which they couldn't cause any damage, or a variety of other basic security measures dating back to the 1960s and 70s. For example most operating systems provide a means of dropping privileges, allowing the attachment to be viewed in a context in which it's incapable of causing any damage.

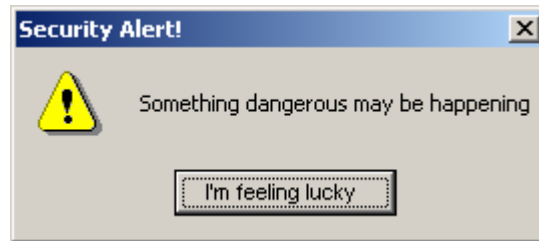


Figure 17: A typical security dialog translated into plain language

Another reason why user education doesn't work is that it's often used as a catch-all for problems that are too hard for the security application developer to solve: "If a problem is too complicated to solve easily, we'll make it a user education issue, and then it's someone else's problem". Any dialog that asks a question phrased something like "There may or may not be something dangerous ahead, do you want to continue?" is an example of an instance where the application developer has simply given up (see Figure 17). Interaction designer Alan Cooper calls this "uninformed consent" [71] — all the power of the application's security mechanisms is now being controlled by a single user judgement call. By offloading this responsibility, the user will still fall head-first down the mine-shaft, but now it's their fault and not the developer's.

HCI researchers label this use of dialogs warn-and-continue (WC), acknowledging the fact that the majority of users will dismiss the dialog and continue anyway. The user's handling of such confirmation dialogs has been characterised as "Yes, yes, yes, yes, oh dear" [85]. While dropping security decisions into a WC may satisfy the application developer, it does little to protect the user. This "not-my-problem" approach to handling responsibility for security decisions was illustrated in one study into the effectiveness of browser security which found that "users expect the browser to make such trust decisions correctly; however browser vendors do not accept this responsibility, and expect users to make the ultimate trust decision" [14]. As a result, no-one took responsibility for (in this case) trusting keys and certificates, since both sides assumed that it was the other side's problem and that they therefore didn't have to concern themselves with it. Psychology professor James Reason, whose specialty is the breakdown of complex technological systems, calls such design flaws latent pathogens, problems that aren't discovered until the user has fallen victim to them [86].

Attacks against the user interface are getting better and better as attackers gain more experience in this area. As these attacks evolve, they're tested in the world's largest usability testing lab (the real world), with ones that succeed being developed further and ones that fail being dropped (compare this to general-purpose software, where buggy and hard-to-use software often persists for years because the same evolutionary pressures don't exist). Usability researchers have actually found that their work makes them much better at attacking users, because by studying security usability they're able to easily defeat the (often totally inadequate) security user interface in applications. Just as spammers have employed professional linguists to help them to get around spam filters and phishers have employed psychology graduates to help them scam victims, so it's only a matter of time before attackers use user interface research against poorly-designed security applications. As one study into the effectiveness of phishing puts it, "None of these [papers proposing security mechanisms] consider that these indicators of trust may be spoofed and that the very guidelines that are developed for legitimate organisations can also be adopted by phishers" [29]. Don't assume that some sort of user education can make a complex

user interface provide security — it'll only work until the bad guys use its complexity against it, or a new crop of non-educated (for that particular interface) users appears.

Only a small number of real-world evaluations of the effectiveness of user education have been performed to date, and the outcomes have been discouraging. In one evaluation of the effectiveness of trying to educate users about phishing, researchers discovered that the education attempts had made no difference in users' ability to detect phishing email. What it did do was scare them into rejecting more phishing emails, but also rejecting proportionately more non-phishing emails (the same thing happened in the false-web-site detection tests discussed earlier). The ratio of rejected phishing emails to non-phishing emails was identical before and after the "education", the only thing that had changed was their fear-based rejection threshold for any email at all [87]. While fear-based marketing has long been a staple of the security industry (see the discussion of people's fears of losing something in the next section for why this is so effective), this may be the first experiment that reveals that in some cases fear is the sole effect of trying to inform people of security issues.

Other education attempts have fared even worse. In the EV certificate evaluation discussed earlier, users actually performed worse after they'd been "educated" because they were inadvertently being trained to rely on the wrong security indicators, and as other earlier discussions have pointed out, US banks have a proud tradition of mis-educating users into insecure behaviour.

A more succinct summary of the fallacy of user education as a solution to the problem has been offered by anti-virus researcher Vesselin Bontchev: "If user education was going to work, it would have worked by now" [88].

References

- [1] "Models of Man : Social and Rational", Herbert Simon, John Wiley and Sons, 1957.
- [2] "Instant Notes in Cognitive Psychology", Jackie Andrade and Jon May, Garland Science/BIOS Scientific Publishers, 2004.
- [3] "Sources of Power: How People Make Decisions", Gary Klein, MIT Press, 1998.
- [4] "Die Logik des Mißlingens. Strategisches Denken in komplexen Situationen", Dietrich Dörner, rowohlt Verlag, 2003.
- [5] "Characteristics of Skilled Option Generation in Chess", Gary Klein, S. Wolf, Laura Militello, and Carolyn Zsombok, *Organizational Behavior and Human Decision Processes*, **Vol.62, No.1** (April 1995), p.63.
- [6] "Problem Solving", Alan Lesgold, "The Psychology of Human Thought", Cambridge University Press, 1988, p.188.
- [7] "Environmental load and the allocation of attention", Sheldon Cohen, *Advances in Environmental Psychology: Vol I — The Urban Environment*: John Wiley & Sons, 1978, p.1.
- [8] "Decision making under stress: scanning of alternatives under controllable and uncontrollable threats", Giora Keinan, *Journal of Personality and Social Psychology*, **Vol.52, No.3** (March 1987), p.639.
- [9] "'Information Load' and Consumers", Debra Scammon, *Journal of Consumer Research: An Interdisciplinary Quarterly*, Vol.4, Issue 3, 1977, p.148.
- [10] "On Leaping to Conclusions When Feeling Tired: Mental Fatigue Effects on Impressional Primacy", Donna Webster, Linda Richter, and Arie Kruglanski, *Journal of Experimental Social Psychology*, **Vol.32, No.2** (March 1996), p.181.
- [11] "Controlled and Automatic Human Information Processing: 1. Detection, Search, and Attention", Walter Schneider and Richard Shiffrin, *Psychological Review*, **Vol.84, No.1** (January 1977), p.1.
- [12] "Controlled & automatic processing: behavior, theory, and biological mechanisms", Walter Schneider and Jason Chein, *Cognitive Science*, **Vol.27, No.3** (May/June 2003), p.525.

- [13] "Some experiments on the recognition of speech with one and two ears", E.C.Cherry, *Journal of the Acoustic Society of America*, **Vol.25, No.5** (May 1953), p.975.
- [14] "Security and Identification Indicators for Browsers against Spoofing and Phishing Attacks", Amir Herzberg and Ahmad Jbara, Cryptology ePrint Archive, <http://eprint.iacr.org/2004/>, 2004.
- [15] "More than 450 Phishing Attacks Used SSL in 2005", Rich Miller, 28 December 2005, http://news.netcraft.com/archives/2005/12/28/-more_than_450_phishing_attacks_used_ssl_in_2005.html.
- [16] "Cardholders targetted by Phishing attack using visa-secure.com", Paul Mutton, 8 October 2005, http://news.netcraft.com/archives/2004/10/08/-cardholders_targetted_by_phishing_attack_using_visasecurecom.html.
- [17] "Judgement under uncertainty: Heuristics and biases", Amos Tversky and Daniel Kahneman, *Science*, **Vol.185, Issue 4157** (27 September 1974), p.1124.
- [18] "Judgment under Uncertainty: Heuristics and Biases", Daniel Kahneman, Paul Slovic, and Amos Tversky, Cambridge University Press, 1982.
- [19] "Critical Thinking Skills in Tactical Decision Making: A Model and A Training Strategy", Marvin Cohen, Jared Freeman, and Bryan Thompson, in "Making Decisions Under Stress: Implications for Individual and Team Training", American Psychological Association (APA), 1998, p.155.
- [20] "On the failure to eliminate hypotheses in a conceptual task", Peter Wason, *Quarterly Journal of Experimental Psychology*, **Vol.12, No.4** (1960) p.129.
- [21] "Thinking and Reasoning", Philip Johnson-Laird and Peter Wason, Penguin, 1968.
- [22] "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises", Raymond Nickerson, *Review of General Psychology*, **Vol.2, Issue 2** (June 1998), p.175.
- [23] "The Cambridge Handbook of Thinking and Reasoning", Keith Holyoak and Robert Morrison (eds), Cambridge University Press, 2005.
- [24] "Recent Research on Selective Exposure to Information", Dieter Frey, *Advances in Experimental Social Psychology*, **Vol.19**, 1986, Academic Press, p.41.
- [25] "Selection of Information after Receiving more or Less Reliable Self-Threatening Information", Dieter Frey and Dagmar Stahlberg, *Personality and Social Psychology Bulletin*, **Vol.12, No.4** (December 1986), p.434.
- [26] "Biased Assimilation and Attitude Polarization: The effects of Prior Theories on Subsequently Considered Evidence", Charles Lord, Lee Ross, and Mark Lepper, *Journal of Personality and Social Psychology*, **Vol.37, No.11** (November 1979), p.2098.
- [27] "The Influence of Prior Beliefs on Scientific Judgments of Evidence Quality" Jonathan Koehler, *Organizational Behavior and Human Decision Processes*, **Vol.56, Issue 1** (October 1993), p.28.
- [28] "Psychological Defense: Contemporary Theory and Research", D.Paulhus, B.Fridhandler, and S.Hayes, *Handbook of Personality Psychology*, Academic Press, p.543-579.
- [29] "Why Phishing Works", Rachna Dhamija, J.D.Tygar, and Marti Hearst, *Proceedings of the Conference on Human Factors in Computing Systems (CHI'06)*, April 2006, p.581.
- [30] "Does Personality Matter? An Analysis of Code-Review Ability", Alessandra Devito da Cunha and David Greathead, *Communications of the ACM*, **Vol.50, No.5** (May 2007), p.109.
- [31] "Defender Personality Traits", Tara Whalen and Carrie Gates, Dalhousie University Technical Report CS-2006-01, 10 January 2006.
- [32] "A Guide to the Development and Use of the Myers-Briggs Type Indicator", Isabel Briggs Myers and Mary McCaulley, Consulting Psychologists Press, 1985.
- [33] "Essentials of Myers-Briggs Type Indicator Assessment", Naomi Quenk, Wiley 1999.
- [34] "Psychology (7th ed)", David Myers, Worth Publishers, 2004.

- [35] "On the Conflict Between Logic and Belief in Syllogistic Reasoning", J.Evans, J.Barston, and P.Pollard, *Memory and Cognition*, **Vol.11, No.3** (May 1983), p.295.
- [36] "Influence: Science and Practice", Robert Cialdini, Allyn and Bacon, 2001.
- [37] "Human Error", James Reason, Cambridge University Press, 1990.
- [38] "Pretty good persuasion: : a first step towards effective password security in the real world", Dirk Weirich and Angela Sasse, *Proceedings of the 2001 New Security Paradigms Workshop (NSPW'01)*, September 2001, p.137.
- [39] "The Belief Engine", James Alcock, *The Skeptical Enquirer*, **Vol.19, No.3** (May/June 1995), p.255.
- [40] "Irrationality: Why We Don't Think Straight!", N.Sutherland, Rutgers University Press, 1994.
- [41] "Do Security Toolbars Actually Prevent Phishing Attacks", Min Wu, Robert Miller, and Simson Garfinkel, *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI'06)*, April 2006, p.601.
- [42] "The Social Brain: Discovering the Networks of the Mind", Michael Gazzaniga, Basic Books, 1987.
- [43] "Genesis of popular but erroneous psychodiagnostic observations", Loren Chapman and Jean Chapman, *Journal of Abnormal Psychology*, **Vol.72, No.3** (June 1967), p.193.
- [44] "Perceptual Restoration of Missing Speech Sounds", Richard Warren, *Science*, **Volume 167, Issue 3917** (23 January 1970), p.392.
- [45] "Auditory Perception: A New Analysis and Synthesis (2nd ed)", Richard Warren, Cambridge University Press, 1999.
- [46] "The TIPPI Point: Toward Trustworthy Interface", Sara Sinclair and Sean Smith, *IEEE Security and Privacy*, **Vol.3, No.4** (July/August 2005), p.68.
- [47] "Gartner: Consumers Dissatisfied with Online Security", Paul Roberts, PC World, December 2004.
- [48] "FobCam", <http://fob.webhop.net/>.
- [49] "Zufall unter Beobachtung", Michael Schilli, Linux Magazine, May 2007, p.98.
- [50] "Users are not the enemy", Anne Adams and Martina Sasse, *Communications of the ACM*, **Vol.42, No.12** (December 1999), p.41.
- [51] "Unverhältnismäßiges Urteil", Ulf Kersing, *c't Magazin für Computertechnik*, 2 October 2006, p.11.
- [52] "Gathering Evidence: Use of Visual Security Cues in Web Browsers", Tara Whalen and Kori Inkpen, *Proceedings of the 2005 Conference on Graphics Interface*, 2005, p.137.
- [53] "Stopping Spyware at the Gate: A User Study of Privacy, Notice and Spyware", Nathaniel Good, Rachna Dhamija, Jens Grossklags, David Thaw, Steven Aronowitz, Deirdre Mulligan, and Joseph Konstan, *Proceedings of the 2005 Symposium on Usable Privacy and Security*, July 2005, p.43.
- [54] "EULalyzer", Javacool Software, <http://www.javacoolsoftware.com/eulalyzer.html>.
- [55] "Beware of the dancing bunnies", Larry Osterman, 12 July 2005, <http://blogs.msdn.com/larryosterman/archive/2005/07/12/438284.aspx>.
- [56] "Do Security Toolbars Actually Prevent Phishing Attacks", Min Wu, Robert Miller, and Simson Garfinkel, *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI'06)*, April 2006, p.601.
- [57] "A Usability Study and Critique of Two Password Managers", Sonia Chasson, Paul van Oorschot, and Robert Biddle, *Proceedings of the 15th Usenix Security Symposium (Security'06)*, August 2006, p.1.
- [58] "The default answer to every dialog box is 'Cancel'", Raymond Chen, <http://blogs.msdn.com/oldnewthing/archive/2003/09/01/54734.aspx>, 1 September 2003.
- [59] "Norton must die!", 'GFree', <http://ask.slashdot.org/comments.pl?sid=205872&cid=16791238>, 10 November 2006.

- [60] "Why can't I disable the Cancel button in a wizard?", Raymond Chen, <http://blogs.msdn.com/oldnewthing/archive/2006/02/24/538655.aspx>, 24 February 2006.
- [61] "The Feature-Positive Effect in Adult Human Subjects", Joseph Newman, William Wolff, and Eliot Hearst, *Journal of Experimental Psychology: Human Learning and Memory*, **Vol.6, No.5** (September 1980), p.630.
- [62] "Thinking and Reasoning", Alan Garnham and Jane Oakhill, Blackwell Publishing, 1994.
- [63] "Thought and Knowledge: An Introduction to Critical Thinking (4th ed)", Diane Halpern, Lawrence Erlbaum Associates, 2002.
- [64] "Handbook of Classroom Assessment: Learning, Achievement, and Adjustment", Gary Phye (ed), Academic Press Educational Psychology Series, 1996.
- [65] "The Emperor's New Security Indicators", Stuart Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer, *IEEE Symposium on Security and Privacy*, May 2007, to appear.
- [66] "Better Website Identification and Extended Validation Certificates in IE7 and Other Browsers", Rob Franco, 21 November 2005, <http://blogs.msdn.com/ie/archive/2005/11/21/495507.aspx>.
- [67] "Tricking Vista's UAC To Hide Malware", "kdawson", 26 February 2007, <http://it.slashdot.org/article.pl?sid=07/02/26/-0253206>.
- [68] "User Account Control Overview", 7 February 2007, <http://www.microsoft.com/technet/windowsvista/-security/uacppr.msp>.
- [69] "User Account Control", http://en.wikipedia.org/wiki/User_Account_Control.
- [70] "Understanding and Configuring User Account Control in Windows Vista", <http://www.microsoft.com/technet/windowsvista/-library/00d04415-2b2f-422c-b70e-b18ff918c281.msp>.
- [71] "The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How To Restore The Sanity", Alan Cooper, Sams, 1999.
- [72] "Gorillas in our midst: sustained inattention blindness for dynamic events", Dan Simons and Christopher Chabris, *Perception*, **Vol.28** (1999), p.1059.
- [73] "Inattention Blindness", Arien Mack and Irvin Rock, MIT Press, 1998.
- [74] "How the Mind Works", Steven Pinker, W.W.Norton and Company, 1997.
- [75] "A Remote Vulnerability in Firefox Extensions", Christopher Soghoian, 30 May 2007, <http://paranoia.dubfire.net/2007/05/remote-vulnerability-in-firefox.html>.
- [76] "A Usability Study and Critique of Two Password Managers", Sonia Chasson, Paul van Oorschot, and Robert Biddle, *Proceedings of the 15th Usenix Security Symposium (Security'06)*, August 2006, p.1.
- [77] "TLS and SSL in the real world", Eric Lawrence, 20 April 2005, <http://blogs.msdn.com/ie/archive/2005/04/20/-410240.aspx>.
- [78] "SSL without a PKI", Steve Myers, in "Phishing and Countermeasures", Markus Jakobsson and Steven Myers (eds), John Wiley and Sons, 2007.
- [79] "Making Security Usable", Alma Whitten, PhD thesis, Carnegie Mellon University, May 2004.
- [80] "Re: Intuitive cryptography that's also practical and secure", Andrea Pasquinucci, posting to the cryptography@metzdowd.com mailing list, message-ID 20070130203352.GA17174@old.at.home, 30 January 2007.
- [81] "Models of Man: Social and Rational", Herbert Simon, Wiley and Sons, 1957.
- [82] "Don't Make Me Think : A Common Sense Approach to Web Usability", Steve Krug, New Riders Press, 2005.
- [83] "AOL Names Top Spam Subjects For 2005", Antone Gonsalves, Information Week TechWeb News, 28 December 2005, <http://www.informationweek.com/news/showArticle.jhtml?articleID=175701011>.

- [84] Microformats, http://microformats.org/wiki/Main_Page.
- [85] “Design Rules Based on Analyses of Human Error”, Donald Norman, *Communications of the ACM*, **Vol.26, No.4** (April 1983), p.255.
- [86] “Human Error”, James Reason, Cambridge University Press, 1990.
- [87] “Phishing IQ Tests Measure Fear, Not Ability”, Vivek Anandpara, Andrew Dingman, Markus Jakobsson, Debin Liu, and Heather Roinestad, *Usable Security 2007 (USEC'07)*, February 2007, <http://usablesecurity.org/program.html>.
- [88] Vesselin Bontchev, remarks during the “Where have all the OUTBREAKS gone” panel session, Association of anti Virus Asia Researchers (AVAR) 2006 conference, Auckland, New Zealand, December 2006.

Security Usability Design

Now that we've looked at all of the problems that need to be solved (or at least addressed) in designing a security user interface, we can move on to the security usability design process. The following sections look at various user interface design issues and ways of addressing some of the problems mentioned in the previous chapter.

Ease of Use

Users hate configuring things, especially complex security technology that they don't understand. One usability study of a PKI found that a group of highly technical users, most with PhDs in computer science, took over two hours to set up a certificate for their own use, and rated it as the most difficult computer task they'd ever been asked to perform [1]. Even more, when they'd finished they had no idea what they'd just done to their computers, with several commenting that had something gone wrong they would have been unable to perform even basic troubleshooting, a problem that had never encountered before.

In practice, security experts are *terrible* at estimating how long a task will take for a typical user. In the PKI usability study, other security researchers who reviewed the paper had trouble believing the empirical results obtained because it couldn't possibly take users that long to obtain and configure a certificate ("I'm sorry but your facts just don't support our theory"). The researchers who set up the study had themselves managed to complete the task in two-and-a-half minutes. The test users (who, as has already been mentioned, had PhDs in computer science and were given screenshot-by-screenshot paint-by-numbers instructions showing them what to do) took two hours and twenty minutes. A more typical user, without a PhD and paint-by-numbers instructions to guide them, has no hope of ever completing this task.

On the other hand the equipment vendors (who have direct contact with end users) were under no illusions about the usability of PKI, expressing surprise that anyone would take on the complexity of a PKI rather than just going with user names and passwords. The assumption by the security experts was that if they could do it in ten minutes then anyone could do it in ten minutes, when in fact a typical user may still not be able to do it after ten hours. This is because users aren't interested in finding out how something works, they just want to use it to do their job. This is very hard for techies, who are very interested in how things work, to understand [2].

Consumer research has revealed that the average user of a consumer electronics device such as a VCR or cell phone will struggle with it for twenty minutes before giving up [3]. Even the best-designed, simplest security mechanism requires more effort to use than not using any security at all, and once we get to obscure technologies like certificates, for which the perceived benefits are far less obvious than for cell phones and VCRs, the user's level of patience drops correspondingly (even the two-and-a-half minutes required by seasoned experts is probably too long for this task).

To avoid problems like this, it should be immediately obvious to a user how the basic security features of your application work. Unlike other applications like web browsers, word processors, and photo editors, users don't spend hours a day inside the security portions of applications, and don't have the time investment to memorise how to use them. Your application should auto-configure itself as much as possible, leaving only a minimal set of familiar operations for the user. For example a network server can automatically generate a self-signed certificate on installation and use that to secure communications to it, avoiding the complexity and expense of obtaining a certificate from an external CA. An email-application can automatically obtain a certificate from a local CA if there's one available (for example an in-house one if the software is being used in an organisation) whenever a new email address is set up. Even if you consider this to be a lowering of *theoretical* security, it's raising its *effective* security because now it'll actually be used.

On the client side, your application can use cryptlib's plug-and-play PKI facility to automatically locate and communicate with a CA server [4], requiring that the user enter nothing more than a name and password to authenticate themselves (this process takes less than a minute, and doesn't require a PhD in computer science to understand). For embedded devices, the operation can occur automatically when the device is configured at the time of manufacture.

Since all users are quite used to entering passwords, your application can use the traditional user name and password (tunnelled over a secure channel such as SSL/TLS or SSH) rather than more complex mechanisms like PKI, which in most cases is just an awkward form of user name and password (the user name and password unlock the private key, which is then used to authenticate the user). Many users choose poor passwords, so protocols like TLS' password-based authentication (TLS-PSK), which never transmit the password even over the secured link, should be preferred to ones that do. TLS-PSK used in this manner is automatically part of the critical action sequence.

An additional benefit of TLS' password-based authentication is that it performs mutual authentication of both parties, identifying not only the client to the server but also the server to the client, without any of the expense, overhead, or complexity of certificates and a PKI. Whereas PKI protects names (which isn't very useful), TLS-PSK protects relationships (which is). Interestingly, RSA Data Security, the company that created Verisign, has recently advocated exactly this method of authentication in place of certificates [5]. Of course users don't know (or care) about the fact that they're performing mutual authentication, all they care about is that they have a verified secure channel to the other party, and all they know about is that they're entering their password as usual.

A final benefit of TLS-PSK is that it allows the server to perform password-quality checks and reject poor, easy-to-guess passwords ("What's your dog's maiden name?"). With certificates there's no such control, since the server only sees the client's certificate and has no idea of the strength of the password that's being used to protect it on the client machine. A survey of SSH public-key authentication found that nearly two thirds of all private keys weren't just poorly protected, they used *no protection at all*. As far as the server was concerned the clients were using (hopefully strong) public-key-based authentication, when the private keys were actually being held on disk as unprotected plaintext files [6]. Furthermore, SSH's **known-host** mechanism would tell an attacker who gains access to a client key file exactly which systems they could compromise using the unprotected key.

You can obtain invisible TLS-PSK-type beneficial effects through the use of other security mechanisms that double up an operation that the user wants to accomplish with the security mechanism. Perhaps the best-known of these is the use of an ignition key in a car. Drivers don't use their car keys as a security measure, they use them to tell the car when to start and stop. However, by doing so they're also getting security at a cost so low that no-one notices.

A more overt piggybacking of security on usability was the design of the common fill device for the KW-26 teletype link encryptor, which was keyed using a punched card supported at the end by pins. To prevent the same card from being re-used, it was cut in half when the card reader door was opened [7]. Since it was supported only at the two ends by the pins, it wasn't possible to use it any more. This meant that the normal process of using the device guaranteed (as a side-effect) that the same key was never re-used, enforcing with the simplest mechanical measures something that no amount of military discipline had been able to achieve during the previous world war. Being able to double up the standard use of an item with a security mechanism in this manner unfortunately occurs only rarely, but when it does happen it's extraordinarily effective.

(In practice the KW-26 mechanism wasn't quite as effective as its designers had hoped. Since distributing the cards ended up costing \$50-100 a pop due to the security requirements involved, and there were potentially a dozen or more devices to re-key, mistakes were quite costly. Users discovered that it was possible, with a bit

of patience, to tape the segments back together again in such a way that they could be re-used, contrary to the designers' intentions. This is the sort of problem that a post-delivery review, discussed in the section on usability testing, would have turned up).

Automation vs. Explicitness

When you're planning the level of automation that you want to provide for users, consider the relative tradeoffs between making things invisible and automated vs. obvious but obtrusive. Users will act to minimise or eliminate monotonous computer tasks if they can, since humans tend to dislike repetitive tasks and will take shortcuts wherever possible. The more that users have to perform operations like signing and encryption, the more they want shortcuts to doing so, which means either making it mostly (or completely) automated, with a concomitant drop in security, or having them avoid signing/encrypting altogether. So a mechanism that requires the use of a smart card and PIN or biometrics will inevitably end up being rarely-used, while one that automatically processes anything that floats by will be. You'll need to decide where the best trade-off point lies — see the section on theoretical vs. effective security above for more guidance on this.

There are however cases where obtrusive security measures are warranted, such as when the user is being asked to make important security decisions. In situations like this, the user should be required to explicitly authorise an action before the action can proceed. In other words any security-relevant action that's taken should represent a conscious expression of the will of the user. Silently signing a message behind the user's back is not only bad practice (it's the equivalent of having them sign a contract without reading it), but is also unlikely to stand up in a court of law, thus voiding the reason usually given for signing a document.

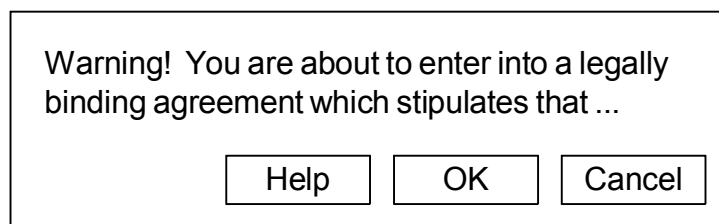
If the user is being asked to make a security-relevant decision of this kind, make sure that the action of proceeding really does represent an informed, conscious decision on their part. Clear the mouse and keyboard buffers to make sure that a keystroke or mouse click still present from earlier on doesn't get accepted as a response for the current decision. Don't assign any buttons as the default action, since something as trivial as bumping the space bar will, with most GUIs, trigger the default action and cause the user to inadvertently sign the document (in this case the secure default is to do nothing, rather than allowing the user to accidentally create a signature). If necessary, consult with a lawyer about requirements for the wording and presentation of requests for security-related decisions that may end up being challenged in court.

Making sure that the input that your user interface is getting was directly triggered by one of the interface elements is an important security measure. If you don't apply measures like this, you make yourself vulnerable to a variety of presentation attacks in which an attacker redirects user input elsewhere to perform various malicious actions. Consider a case where a web page asks the user to type in some scrambled letters, a standard CAPTCHA/reverse Turing test used to prevent automated misuse of the page by bots. The letters that the user is asked to type are "xyz". When the user types the 'x', the web page tries to install a malicious ActiveX control. Just as they type the 'y', the browser pops up a warning dialog asking the user whether they want to run the ActiveX control, with a Yes/No button to click. The input focus is now on the warning dialog rather than the web page, which receives the user's typed 'y' and instantly disappears again as the browser installs the malicious ActiveX control. This attack, which was first noticed by the Firefox browser developers [8][9][10] but also affected Internet Explorer [11] is somewhat unusual in that it's more effective against skilled users, whose reaction time to unexpected stimuli is far slower than their typing speed.

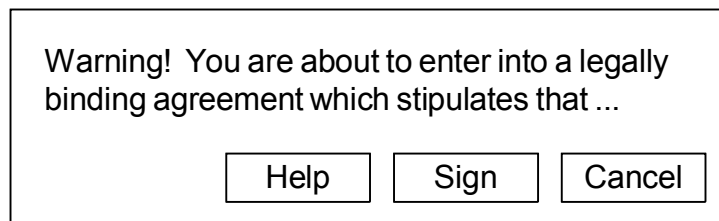
This type of attack isn't limited solely to the keyboard. Since dialogs pop up at known locations, it's possible to use enqueued mouse clicks in a similar way to enqueued keystrokes, having users double-click on something and then popping up a dialog under the location of the second click, or forcing them to click away a series of popups with something critical hidden at the bottom of the stack. On most systems this occurs so quickly that the user won't even be aware that it's happened [12].

The Firefox solution to this problem was to clear the input queue and insert a time delay into the XPI extension installation button, hopefully giving users time to react to the dialog before taking any action [13]. Unfortunately users weren't aware of why the delay was there and perceived it as a nagware tactic, in some cases altering their browser configuration to reduce the delay to zero [14][15]. There's even an XPI plugin to remove the XPI plugin-install delay [16]. A "Why is this button greyed out" tooltip would have helped here.

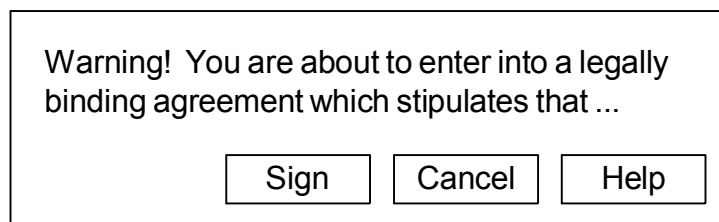
Apple's solution to the problem was to force users to use a mouse click to acknowledge an install dialog, and to add a second "Are you sure?" dialog to confirm this. While this isn't useful against user conditioning to click OK on any dialog that pops up, it does insert enough of a speed bump that users can't be tricked into installing something without even knowing that they've done it, or at least no more so than a standard click, whirr response would allow anyway. As the second attack variant described above indicates, just the mouse-only requirement by itself isn't a practical defence against this type of attack, and has the added drawback of making the dialog inaccessible to non-mouse users.



Consider the signature dialog above, which represents the first attempt at an appropriate warning dialog in a digital signature application. When challenging this in court, J.P.Shyster (the famous defence lawyer) claims that his client, dear sweet Granny Smith, was merely acknowledging a warning when she clicked OK, and had no idea that she was entering into a legally binding contract. The sixty-year-old judge with a liberal arts degree and a jury of people whose VCRs all blink '12:00' agree with him, and the contract is declared void.



So the application designers try again, and having learned their lesson come up with the dialog above. This time it's obvious that a signature is being generated. However, now J.P.Shyster points out that the buttons are placed in a non-standard manner (the 'Sign' button is where the 'Cancel' button would normally be) by obviously incompetent programmers, and produces a string of expert witnesses and copies of GUI design guidelines to back up his argument. The judge peers at the dialog through his trifocals and agrees, and the case is again dismissed.



The designers try again, at the third attempt coming up with the dialog above. This time, J.P.Shyster argues that Granny Smith was again merely being presented with a warning that she was about to enter into an agreement, and that there was no indication in the dialog that she was assenting to the agreement the instant she clicked

‘Sign’. The judge, who’s getting a bit tired of this and just wants to get back to his golf game, agrees, and the case is yet again dismissed.

By clicking 'Sign' below I acknowledge that I am entering into a legally binding agreement ...

The application designers’ fourth attempt is shown above. J.P.Shyster has since moved on to a successful career in politics, so this time the design isn’t tested in court. This does, however, show how tricky it is to get even a basic security dialog right, or at least capable of standing up to hostile analysis in court (a skilled lawyer will be able to find ambiguity in a “No smoking” sign). More dangerous than the most devious phisher, more dangerous even than a government intelligence agency, a hostile expert witness is the most formidable attack type that any security application will ever have to face.

Safe Defaults

Your application should provide sensible security defaults, and in particular ensure that the default/most obvious action is the safest one. In other words if the user chooses to click “OK” for every action (as most users will do), they should be kept from harming themselves or others. Remember that if you present the user with a dialog box that asks “A possible security problem has been detected, do you want to continue [Yes/No]”, what the user will read is “Do you want this message to go away [Yes/No]” (or more directly “Do you want to continue doing your job [Yes/No]”, see Figure 18). Ensuring that the Yes option is the safe one helps prevent the user from harming themselves (and others) when they click it automatically.

A possible security problem has been detected, do you want to continue?

Do you want this warning to go away?

Figure 18: What the developer wrote (above); what the user sees (below)

One simple way to test your application is to run it and click OK (or whatever the default action is) on every single security-related dialog that pops up (usability testing has shown that there are actually users who’ll behave in exactly this manner). Is the result still secure?

Now run the same exercise again, but this time consider that each dialog that’s thrown up has been triggered by a hostile attack rather than just a dry test-run. In other words the “Are you sure you want to open this document (default ‘Yes’)” question is sitting in front of an Internet worm and not a Word document of last week’s sales figures. Now, is your application still secure? A great many applications will fail even this simple security usability test.

cryptlib already enforces this secure-by-default rule by always choosing safe settings for security options, algorithms, and mechanisms, but you should carefully check

your application to ensure that any actions that it takes (either implicitly, or explicitly when the user chooses the default action in response to a query) are the safest ones. The use of safe defaults is also preferable to endless dialogs asking users to confirm every action that needs to be taken, which rapidly becomes annoying and trains users to dismiss them without reading them.

One way avoiding the “Click OK to make this message go away” problem is to change the question from a basic yes/no one to a multi-choice one, which makes user satisficing much more difficult. In one real-world test, about a third of users fell prey to attacks when the system used a simple yes/no check for a security property such as a verification code or key fingerprint, but this dropped to zero when users were asked to choose the correct verification code from a selection of five (one of which was “None of the above”) [17]. The reason for this was that users either didn’t think about the yes/no question at all, or applied judgemental heuristics to rationalise any irregularities away as being transient errors, while the need to choose the correct value from a selection of several actually forced them to think about the problem.

The shareware WinZip program uses a similar technique to force users to stop and think for the message that it displays when an unregistered copy is run, swapping the buttons around so that users are actually forced to stop and read the text and think about what they’re doing rather than automatically clicking ‘Cancel’ without thinking about it. Similarly, the immigration form used by New Zealand Customs swaps some of the yes/no questions so that it’s not possible to simply check every box in the same column without reading the questions (this is a particularly evil thing to do to a bunch of half-asleep people who have just come off the 12-hour flight that it takes to get there).

Another technique that you can use is to disable (grey out) the button that invokes the dangerous action for a set amount of time to force users to take notice of the dialog. If you do this, make the greyed-out button display a countdown timer to let users know that they can eventually continue with the action, but have to pause for a short time first (hopefully they’ll read and think about the dialog while they’re waiting). The Firefox browser uses this trick when browser plugins are installed, although in the case of Firefox it was actually added for an entirely different reason which was obscure enough that it was only revealed when a Firefox developer posted an analysis of the design rationale behind it [18]. Although this is borrowing an annoying technique from nagware, it may be the only way that you can get users to consider the consequences of their actions rather than just ploughing blindly ahead. Obviously you should restrict the use of this technique to exceptional error conditions rather than something that the user encounters every time that they want to use your application.

Techniques such as this, which present a roadblock to muscle memory, help ensure that users pay proper attention when they’re making security-relevant decisions. Another muscle memory roadblock, already mentioned earlier, is removing the window-close control on dialog boxes. There also exist various other safety measures that you can adopt for actions that have potentially dangerous consequences. For example Apple’s user interface guidelines recommend spacing buttons for dangerous actions at least 24 pixels away from other buttons, twice the normal distance of 12 pixels [19].

Another way of enforcing the use of safe defaults is to require extra effort from the user to do things the unsafe way, and to make it extremely obvious that this is a bad way to do things. The technical term for this type of mechanism, which prevents (or at least makes unlikely) some type of mistake, is a forcing function [20]. Forcing functions are used in a wide variety of applications to dissuade users from taking unwise steps. For example the programming language Oberon requires that users who want to perform potentially dangerous type casts import a pseudo-module called `SYSTEM` that provides the required casting functions. The presence of this import in the header of any module that uses it is meant to indicate, like the fleur-de-lis brand on a criminal, that unsavoury things are taking place here and that this is something you may want to avoid contact with.

An example of a security-related forcing function occurs in the MySQL database replication system, which has a master server controlling several networked slave machines. The replication system user starts the slave with `start slave`, which automatically uses SSL to protect all communications with the master. To run without this protection, the user has to explicitly say `start slave without security`, which both requires more effort to do and is something that will give most users an uneasy feeling. Contrast this with many popular mail clients, which perform all of their communication with the host in the clear unless the user remembers to check the “Use SSL” box buried three levels down in a configuration dialog or include the `ssl` option on the command-line. As one assessment of the Thunderbird email client software puts it, “This system is *only usable by computer experts*. The only reason I was able to ‘quickly’ sort this out was because I knew (as an experienced cryptoplumber) exactly what it was trying to do. I know that TLS requires a cert over the other end, and there is a potential client-side cert. But without that knowledge, a user would be lost [...] It took longer to do the setting up of some security options than it takes to download, install, and initiate an encrypted VoIP call over Skype with someone who has *never used Skype before* [21]

Requirements and Anti-requirements

One way to analyse potential problem areas is to create a set of anti-requirements to parallel the more usual design requirements. In other words, what *shouldn't* your user interface allow the user to do? Should they really be able to disable all of the security features of your software via the user interface (see Figure 19)? There are in fact a whole raft of viruses and worms that disable Office and Outlook security via OLE automation, and no Internet worm would be complete without including facilities to disable virus checkers and personal firewalls. This functionality is so widespread that at one point it was possible to scan for certain malware by checking not so much for the malware itself but merely the presence of code to turn off protection features.

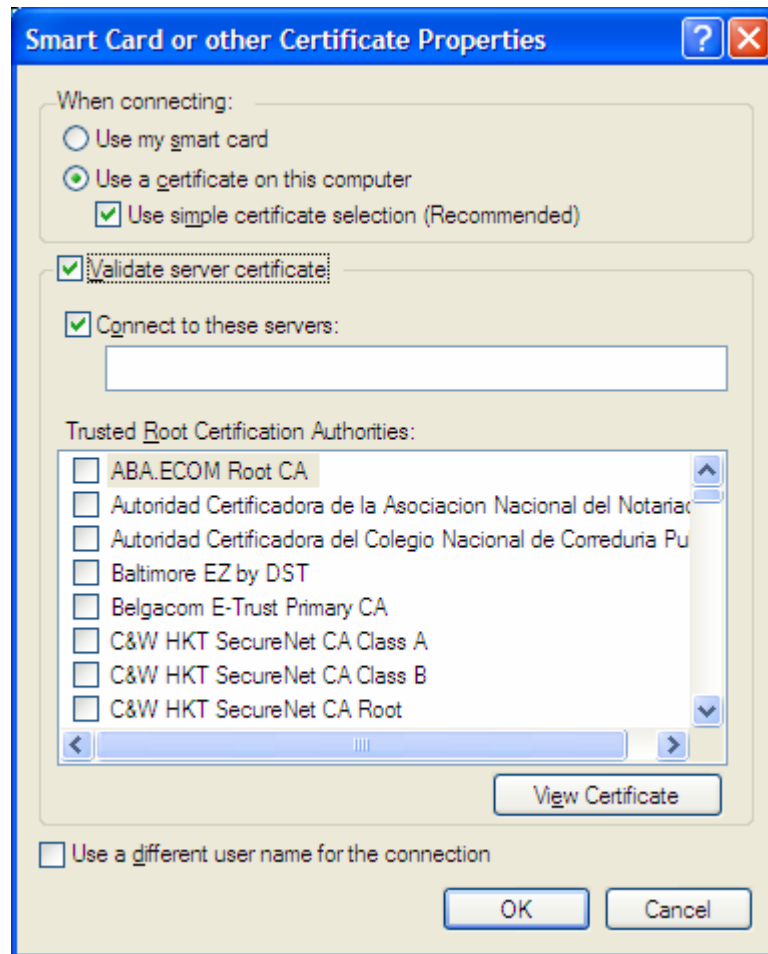


Figure 19: Would you buy a car that had a ‘disable the brakes’ option?

Just because malware commonly takes advantage of such capabilities, don’t assume that these actions will be taken only by malware. Many vendor manuals and websites contain step-by-step instructions (including screenshots) telling users how to disable various Windows security features in order to make some piece of badly-written software run, since it’s easier to turn off the safety checks than to fix the software. So create a list of anti-requirements — things that your user interface should on no account allow the user to do — and then make sure that they are in fact prevented from doing them.

Another way to analyse potential problems in the user interface is to apply the `bCanUseTheDamnThing` test (if you’re not familiar with Hungarian notation, the `b` prefix indicates that this is a boolean variable and the rest of the variable name should be self-explanatory). This comes from an early PKI application in which the developers realised that neither programmers nor users were even remotely interested in things such as whether an X.509 certificate’s policy-mapping-inhibit constraint on a mapped policy derived from an implicit initial policy set had triggered or not, all that they cared about was `bCanUseTheDamnThing`. Far too many security user interfaces (and at a lower level programming libraries) present the user or developer with a smorgasbord of choices and then expect them to be able to mentally map this selection onto `bCanUseTheDamnThing` themselves. As the previous section showed, users will invariably map a confusing choice that they’re presented with to `bCanUseTheDamnThing = TRUE` because they don’t understand what they’re being asked to decide but they do understand that a value of `TRUE` will produce the results they desire.

The `bCanUseTheDamnThing` test is a very important one in designing usable security interfaces. If the final stage of your interface algorithm consists of “the user maps our explanation of the problem to `bCanUseTheDamnThing`” then it’s a sign

that your interface design is incomplete, since it's offloading the final (and probably most important) step onto the user rather than handling it itself. Lack of attention to `bCanUseTheDamnThing` shows up again and again in post-mortem analyses of industrial accidents and aircraft crashes: by the time the operators have checked the 800 dials and lights to try and discover where the problem lies, the reactor has already gone critical. It's traditional to blame such faults on "human error", although the humans who made the mistake are really the ones who designed latent pathogens into the interface and not the operators.

Interaction with other Systems

Secure systems don't exist in a vacuum, but need to interact not only with users but also with other, possibly insecure systems. What assumptions is your design making about these other systems? Which ones does it trust? Given Robert Morris Sr.'s definition of a trusted system as "one that can violate your security policy", what happens if that trust is violated, either deliberately (it's compromised by an attacker) or accidentally (it's running buggy software)? For example a number of SSH implementations assumed that when the other system had successfully completed an SSH handshake this constituted proof that it would only behave in a friendly manner, and were completely vulnerable to any malicious action taken by the other system. On a similar note, there's more spam coming from compromised "good" systems than "bad" ones. Trust but verify — a digitally signed virus is still a virus, even if it has a valid digital signature attached.

Going beyond the obvious "trust nobody" approach, your application can also provide different levels of functionality under different conditions. Just as many file servers will allow read-only access or access to a limited subset of files under a low level of user authentication and more extensive access or write/update access only under a higher level of authentication, so your application can change its functionality based on how safe (or unsafe) it considers the situation to be. So instead of simply disallowing all communications to a server whose authentication key has changed (or, more likely, connecting anyway to avoid user complaints), you can run in a "safe mode" that disallows uploads of (potentially sensitive) data to the possibly-compromised server and is more cautious about information coming from the server than usual.

The reason for being cautious about uploads as well as downloads is that setting up a fake server is a very easy way to acquire large amounts of sensitive information with the direct cooperation of the user. For example if an attacker knows that a potential victim is mirroring their data via SSH to a network server, a simple trick like ARP spoofing will allow them to substitute their own server and have the victim hand over their sensitive files to the fake server. Having the client software distrust the server and disallow uploads when its key changes helps prevent this type of attack.

Be careful what you tell the other system when a problem occurs — it could be controlled by an attacker who'll use the information against you. For example some SSH implementations send quite detailed diagnostic information to the other side, which is great for debugging the implementation, but also rather dangerous because it's providing an attacker with a deep insight into the operation of the local system. If you're going to provide detailed diagnostics of this kind, make it a special debug option and turn it off by default. Better yet, make it something that has to be explicitly enabled for each new operation, to prevent it from being accidentally left enabled after the problem is diagnosed (debugging modes, once enabled, are invariably left on "just in case", and then forgotten about).

Security systems can also display emergent properties unanticipated by their original designers when they interact, often creating new vulnerabilities in the process. Consider what happens when you connect a PC with a personal firewall to an 802.11 access point. An attacker can steal the PC's IP and MAC address and use the access point, since the personal firewall will see the attacker's packets as a port scan and silently drop them. Without the personal firewall security system in place, the attacker's connections would be reset by the PC's IP stack. It's only the modification of the two security systems' designed behaviours that occurs when they interact that

makes it possible for two systems with the same IP and MAC addresses to share the connection. So as well as thinking about the interaction of security systems in the traditional “us vs. them” scenario, you should also consider what happens when they interact constructively to produce an unwanted effect.

Conversely, be very careful with how you handle any information from the remote system. Run it through a filter to strip out any special non-printable characters and information before you display it to the user, and present it in a context where it's very clear to the user that the information is coming from another system (and is therefore potentially controlled by a hostile party) and not from your application. Consider the install dialog in Figure 20. The attacker has chosen a description for their program that looks like instructions from the application to the user.

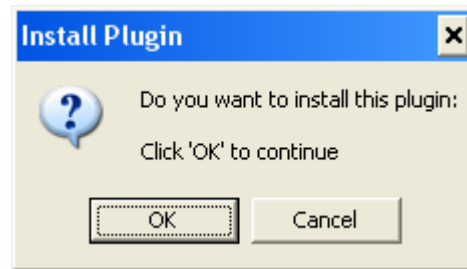


Figure 20: Spoofed plugin install dialog

Since the dialog doesn't make a clear distinction between information from the application and information from the untrusted source, it's easy for an attacker to mislead the user. Such attacks have already been used in the past in conjunction with Internet Explorer, with developers of malicious ActiveX controls giving them misleading names and descriptions that appear to be instructions from the browser.

Matching Users' Mental Models

In order to be understandable to users, it's essential that your application match the user's mental model of how something should work and that it follow the flow of the users' conception of how a task should be performed. If you don't do this, users will find it very difficult to accomplish what they want to do when they sit down in front of your application. In most cases users will already have some form of mental model of what your software is doing, either from the real world or from using similar software (admittedly the accuracy of their model will vary from good through to bizarre, but there'll be some sort of conception there). Before you begin, you should try and discover your users' mental models of what your application is doing and follow them as much as possible, because an application that tries to impose an unfamiliar conceptual model on its users instead of building on the knowledge and experience that the users already have is bound to run into difficulties. This is why (for example) photo-management applications go to a great deal of programming effort to look like photo albums even if it means significant extra work for the application developers, because that's what users are familiar with.

Consider the process of generating a public/private key pair. If you're sitting at a Unix command line, you fire up a copy of `gpg` or `openssl`, feed it a long string of command-line options, optionally get prompted for further pieces of input, and at the end of the process have a public/private key pair stored somewhere as indicated by one of the command-line options.

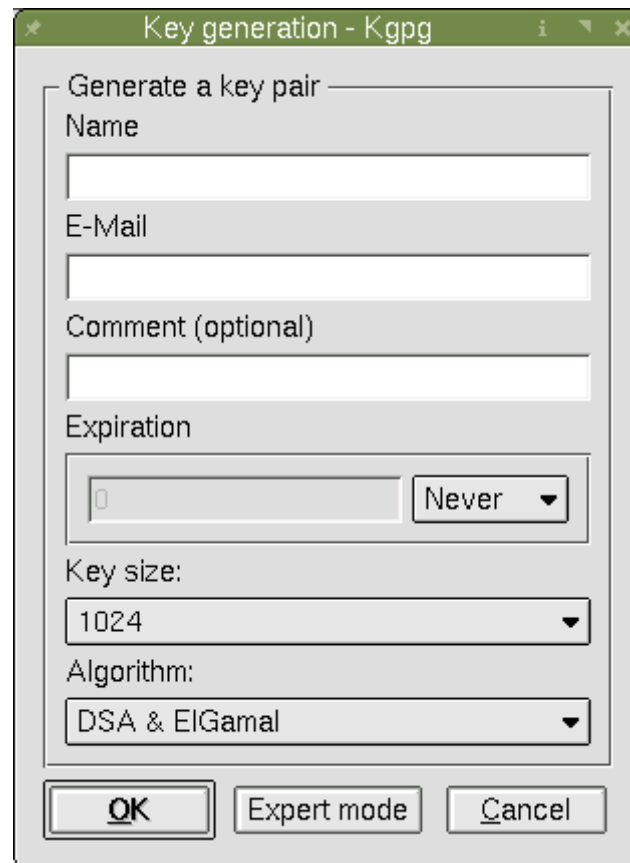


Figure 21: KPGP key generation dialog

This command-line interface-style design has been carried over to equivalent graphical interfaces that are used to perform the same operation. `-k keysize` has become a drop-down combo box. `-a algorithm` is a set of checkboxes, and so on, with Figure 21 being an example of this type of design (note the oxymoronic ‘Expert mode’ option, which leads to an even more complex interface, dropping the user to a command prompt). Overall, it’s just a big graphical CLI-equivalent, with each command-line option replaced by a GUI element, often spread over several screens for good measure (one large public CA requires that users fill out *eleven pages* of such information in order to be allowed to generate their public/private key pair, making the process more a test of the user’s pain tolerance than anything useful). These interfaces violate the prime directive of user interface design: Focus on the users and their tasks, not on the technology [22].

The problem with this style of interface, which follows a design style known as task-directed design, is that while it may cater quite well to people moving over from the command-line interface it’s very difficult to comprehend for the average user without this level of background knowledge, who will find it a considerable struggle to accomplish their desired goal of generating a key to protect their email or web browsing. What’s a key pair? Why do I need two keys instead of just one? What’s a good key size? What are Elgamal and DSA? What’s the significance of an expiry date? Why is an email application asking me for my email address when it already knows it? What should I put in the comment field? Doesn’t the computer already know my name, since I logged on using it? This dialog should be taken outside and shot.

Interfaces designed by engineers tend to end up looking like something from Terry Gilliam’s “Brazil”, all exposed plumbing and wires. To an engineer, the inner workings of a complex device are a thing of beauty and not something to be hidden away. In the software world, this results in a user interface that has a button for every function, a field for every data input, a dialog box for every code module. To a programmer, such a model is definitively accurate. Interaction with the user occurs in

perfect conformity with the internal logic of the software. Users provide input when it's convenient for the application to accept it, not when it's convenient for the user to provide it. This problem is exemplified by the Windows Vista UAC dialog discussed in a previous chapter. Informing the user that an application has been blocked because of a Windows Group Policy administrative setting may be convenient for the programmer, but it provides essentially zero information to the user (the manifold shortcomings of the UAC dialog have provided fertile ground for user interface designers ever since it was released).

The reason why task-directed design is so popular (apart from the fact that it closely matches programmers' mental models) is that as security properties are very abstract and quite hard to understand, it's easier for application developers to present a bunch of individual task controls rather than trying to come up with something that achieves a broader security goal. However, wonderful though your application may be, to the majority of users it's merely a means to an end, not an end itself. Rather than focusing on the nuts and bolts of the key generation process, the interface should instead focus on the activity that the user is trying to perform, and concentrate on making this task as easy as possible. Microsoft has espoused this user interface design principle in the form of Activity-Based Planning, which instead of giving the user a pile of atomic operations and forcing them to hunt through menus and dialogs to piece all the bits and pieces together to achieve their intended goal, creates a list of things that a user might want to do (see the section on pre-implementation testing further on) and then builds the user interface around those tasks.

Activity-Based Planning

Activity-based planning matches users' natural ways of thinking about their activities. Consider the difference in usability between a car designed with the goal of letting the user control the fuel system, camshafts, cooling system, ignition system, turbochargers, and so on, and one designed with goal of making the car go from A to B in the most expedient manner possible. Outside of a few hardcore petrol-heads, no-one would be able to use the former type of car. In fact, people pay car manufacturers significant amounts of money to ensure that the manufacturer spends even more significant amounts of money to keep all of this low-level detail as far away from them as possible. The vast majority of car owners see a car as simply a tool for achieving a goal like getting from A to B, and will spend only the minimal effort required by law (and sometimes not even that) to learn its intricacies.

A similar thing occurs with security applications. Users focus on goals such as "I want my medical records kept private" or "I want to be sure that the person/organisation that I'm talking to really is who they claim to be", rather than focusing on technology such as "I want to use an X.509 certificate in conjunction with triple-DES encryption to secure my communications". Your application should therefore present the task involving security in terms of the users' goals rather than of the underlying security technology, and in terms that the users can understand (most users won't speak security jargon). This both makes it possible for users to understand what it is they're doing, and encourages them to make use of the security mechanisms that are available.

The actual goals of the user often come as a considerable surprise to security people (there's more on this in the section on security usability testing). For example security researchers have been pushing for voter verifiable paper trails (VVPAT) as a safety mechanism for electronic voting machines in the face of a seemingly never-ending stream of reports about the machines' unreliability and insecurity. However, when voting machines with VVPAT capabilities were tested on voters, they completely ignored the paper record, and had less confidence in the VVPAT-enabled devices than in the purely electronic ones, despite extensive and ongoing publicity about their unreliability [23]. A two-year study carried out in Italy ran into the same issues, receiving user comments like "this receipt stuff and checking the votes are dangerous, please give only the totals at the end and no receipts" [24]. This indicates that the users of the equipment (the voters) had very different goals to the security

people who were designing them (or at least trying to fix up the designs of existing devices).

A useful trick to use when you're creating the test for your user interface is to pretend that you're looking over the user's shoulder explaining how to accomplish the task to them, because this tends to lead naturally towards a goal-oriented workflow. If your application is telling the user what to do, use the second person: "Choose the key that you want to use for encryption". If the user is telling the application what to do, use the first person: "Use this key for encryption".

Using the key generation example from earlier, the two activities mentioned were generating a key to protect email, and generating a key to protect web browsing (in other words, for an SSL web server). This leads naturally to an interface in which the user is first asked which of the two tasks they want to accomplish, and once they've made their choice, asked for their name and email address (for the email protection key) or their web server address (for the SSL/web browsing key). Obviously if the key generation is integrated into an existing application, you'd skip this step and go straight to the actual key generation stage — most users will be performing key generation as a side-effect of running a standard application, not because they like playing key administrator with a key management program.

A better option when you're performing the key generation for an application-specific purpose is to try to determine the details automatically, for example by reading the user's name and email address information from the user's mail application configuration and merely asking them to confirm the details. Under Windows you can use CDO (Collaboration Data Objects) to query the `CdoPR_GIVEN_NAME`, `CdoPR_SURNAME`, and `CdoPR_EMAIL` fields of the `CurrentUser` object. Under OS X you can use the `ABAddressBook` class of the `AddressBook` framework to query the "Me" (current) user's `kABFirstNameProperty`, `kABLastNameProperty`, and `kABEmailProperty` and use them to automatically populate the dialog fields. OS X is particularly good in this regard, asking for your address book data the first time that you log in, after which applications automatically use the address book information instead of asking for it again and again in each application. The Opera web browser tries to fix this problem from the opposite end with its Magic Wand feature, which initially records user details and then template-matches them to fields in web pages, providing a browser-based equivalent to the OS X address book, at least for web-based forms.

Conversely, Linux and the *BSDs seem to have no facility for such centralised user information management, requiring that you manually enter the same information over and over again for each application that needs it. One thing that computers are really good at is managing data, so the user shouldn't be required to manually re-enter information that the computer already knows. This is one of the tenets of Macintosh user interface design, the user should never have to tell the machine anything that it already knows or can deduce for itself.

Another benefit of pre-filling in fields is that, even if the information isn't quite what the user wanted and they have to manually correct it, it still provides them with a template to guide them, the equivalent of a default choice in dialog box buttons that provides just-in-time instructions to help them figure out how to complete a field. Again, see the section on pre-implementation testing for a discussion of how to work out details such as where to store the generated key.

There are three additional considerations that you need to take into account when you're using Activity-Based Planning to design your user interface. First, you need to be careful to plan the activities correctly, so that you cover the majority of typical use cases and don't alienate users by forcing them down paths that they don't want to take, or having to try and mentally reverse-engineer the flow to try and guess which path they have to take to get to their desired goal (think of a typical top-level phone menu, for which there are usually several initial choices that might lead to any desired goal). If you have, for example, a key generation wizard that involves more than three or four steps then it's a sign that a redesign is in order.



Figure 22: A portion of the GPA key generation wizard

GPA, an application from the same family as KPGP, used to use an almost identical key generation dialog as KPGP, but in more recent versions has switched to using a wizard-style interface, of which one screen is shown in Figure 22. Unfortunately this new interface is merely the earlier (incomprehensible) dialog cut up into lots of little pieces and presented to the user a step at a time, adding Chinese water torture to the sins of its predecessor.

The second additional consideration is that you should always provide an opt-out capability to accommodate users who don't want to perform an action that matches one of your pre-generated ones. This would be handled in the key-generation interface by the addition of a third option to generate some other (user-defined) type of key, the equivalent of the "Press 0 to talk to an operator" option in a phone menu.

Taking advantage of extensive research by educational psychologists, the dialog uses a conversational rather than formal style. When the user's brain encounters this style of speech rather than the more formal lecturing style used in many dialogs, it thinks that it's in a conversation and therefore has to pay more attention to hold up its end. In other words at some level your brain pays more attention when it's being talked with rather than talked at [25].

Finally, you should provide a facility to select an alternative interface, usually presented as an expert or advanced mode, for users who prefer the nuts-and-bolts style interface in which they can specify every little detail themselves (dropping to the command-line, however, is not a good way to do this). Although the subgroup of users who prefer this level of configurability for their applications is relatively small, it tends to be a rather vocal minority who will complain loudly about the inability to specify their favourite obscure algorithm or select some peculiar key size (this level of flexibility can actually represent a security risk, since it's possible to fingerprint users of privacy applications if they choose unusual combinations of algorithms and key sizes, so that even if their identity is hidden they can be tracked based on their algorithm choice).

The need to handle special cases is somewhat unfortunate since a standard user interface design rule is to optimise your design for the top 80% of users (the so-called "80 percent rule"). The 80% rule works almost everywhere, but there are always special cases where you need to take extra care. An example of such a case is word processors, which will be reviewed by journalists who use the software in very different ways than the average user. So if you want to get a positive review for your word processor, you have to make sure that features used by journalists like an article word count are easy to use. Similarly, when you're designing a security user interface, it's the 1-2% of users who are security experts (self-appointed or otherwise) who will complain the most when your 80 percent solution doesn't cater to their particular requirements.

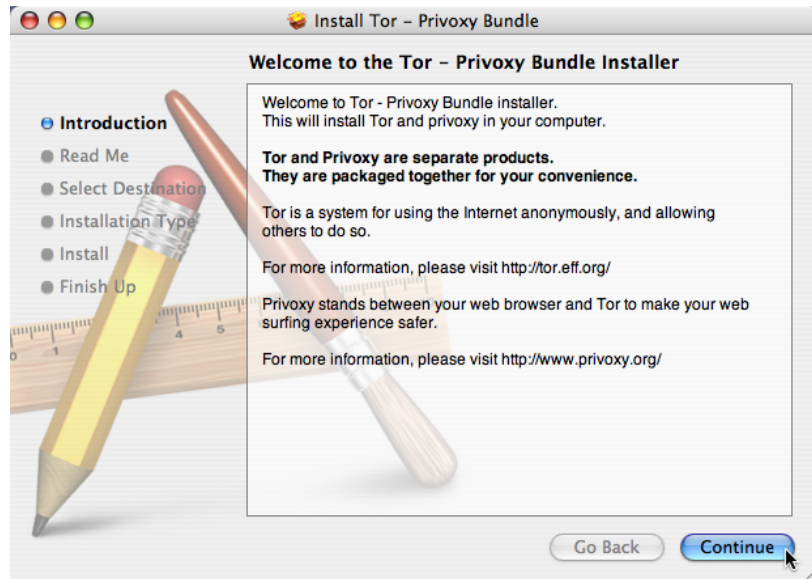


Figure 23: OS X Wizard interface

If you're going to provide an expert-mode style interface, remember to make the simplest, most straightforward interface configuration the default one, since studies have shown that casual users don't customise their interfaces (typically for fear of "breaking something") even when a configuration capability is available.

Design Example: Key Generation

Let's look at a simple design exercise for activity-based planning, in this case involving the task of user key generation for a mail encryption program. The first page of the wizard, shown in Figure 24, explains what's about to happen, and gives the user the choice of using information obtained automatically from the default mail application, or of entering the details themselves.

Create Key - Step 1 of 2

To communicate securely with others, you need to create an encryption key. This key will be labelled as belonging to Bob Sample with the email address `bob@sample.com`.

If you'd like to change your key settings, click 'Change Details', otherwise click 'Create Key'.

Create Key
Change Details
Cancel

Figure 24: Key creation wizard, part 1

There are several things to note about this dialog. The most obvious one is the contents of the title bar, which gives the operation being performed as "Create Key" rather than "Generate Key" or "Generate Key Pair". This is because users *create* documents or *create* images, they don't generate them, so it makes sense that they should also create a key as well. In addition what they're creating is a key, not a key pair — most users will have no idea what a key pair is or why they need two of them. Finally, the title bar indicates their progress through the wizard process, removing the uncertainty over whether they're going to be subject to any Chinese water torture to get their key. OS X Assistants (the equivalent of Windows' wizards, shown in Figure 23) display a list of steps on the left-hand side of the dialog box, including the progress indicator as a standard part of the dialog.

The other point to note is the default setting 'Create Key', and the fact that it's worded as an imperative verb rather than a passive affirmation. This is because the caption for a button should describe the action that the button initiates rather than being a generic affirmation like 'OK', which makes obvious the action that the user is about to perform. In addition, by being the default action it allows the majority of users who simply hit Enter without reading the dialog text to Do The Right Thing.

Finally, note the absence of a window-close control, preventing the user from automatically getting rid of the dialog and then wondering why the application is complaining about the absence of a key.

Create Key - Step 2 of 2	
Your key has been created and saved.	
In order for others to communicate securely with you, you need to publish your key information. Would you like to do this now?	
<input type="button" value="Publish Key"/>	<input type="button" value="Don't Publish Key"/>

Figure 25: Key creation wizard, part 2

The next step, shown in Figure 25, informs the user that their key has been created and safely stored for future use. Again, the default action publishes the key for others to look up. If the user chooses not to publish the key, they're led to a more expert-mode style dialog that warns them that they'll have to arrange key distribution themselves, and perhaps gives them the option of exporting it in text format to mail to others or post to a web page.

Create Key - Done
Your new key is now ready for use.
<input type="button" value="Finish"/>

Figure 26: Key creation wizard, step 3

The final step, shown in Figure 26, completes the wizard and lets the user know that their key is now ready for use (although completion pages for wizards are in general frowned upon, in this case the use is permissible in order to make explicit the fact that the previous action, which would otherwise be invisible to users, has completed successfully). In the worst case, all that the user has to do is hit Enter three times without bothering to stop and read the dialog, and everything will be set up for them.

One possible extra step that isn't shown here is the processing of some form of password or PIN to protect the newly-generated key. This is somewhat situation-specific and may or may not be necessary. For example the key might be stored in a USB security token or smart card that's already been enabled via a PIN, or protected by a master password that the user entered when the application started.

An interesting phenomenon occurs when users are exposed to this style of simple-but-powerful interface. In a usability test of streamlined scanner software, every one of the test users commented that it was the "most powerful" that they'd tried, even though it had fewer features than the competition. What made it powerful was the effective power realised by the user, not the feature count. A side-effect of this "powerful" user interface was that it generated a radically smaller number of tech support calls than was normal for a product of this type [26]. This confirms

interaction designer Alan Cooper’s paraphrasing of architect Mies van der Rohe’s dictum “Less is more” into the user interface design principle “No matter how cool your user interface is, less of it would be better” [27] (a remark that’s particularly applicable to skinnable interfaces).

Use of Familiar Metaphors

Many users are reluctant to activate security measures because the difficulty of configuring them is greater than any perceived benefits. Using a metaphor that’s familiar to the user can help significantly in overcoming this reluctance to deal with security issues. For example most users are familiar with the use of keys as security tools, making a key-like device an ideal mechanism for propagating security parameters from one system to another. One of the most usable computer security devices ever created, the Datakey, is shown in Figure 27. To use the Datakey, you insert it into the reader and turn it to the right until it clicks into place, just like a standard key. To stop using it, you turn it back to the left and remove it from the reader.

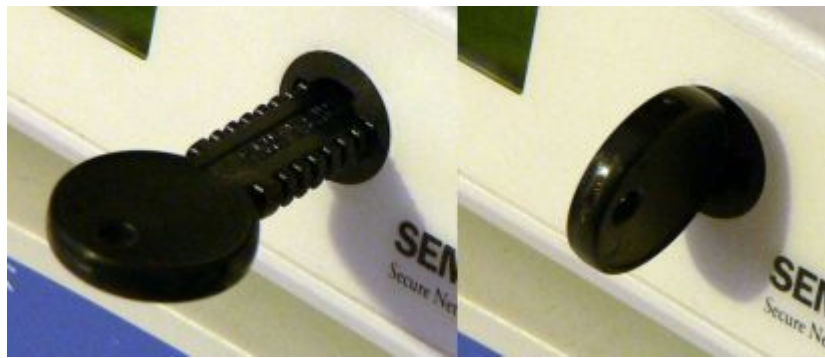


Figure 27: A Datakey being used to key a VPN box

Instead of using a conventional key, the device used to initialise security parameters across devices is a USB memory key that the user takes to each device that’s being initialised. This mechanism is used in Microsoft’s Windows Network Smart Key (WNSK), in which Windows stores WiFi/802.11 encryption keys and other configuration details onto a standard USB memory key, which is then inserted into the other wireless devices that need to be configured.

Since USB keys can store amounts of information that would be impossible for humans to carry from one device to another (the typical WNSK file size is around 100KB), it’s possible to fully automate the setup using full-strength security parameters and configuration information that would be impossible for humans to manage. In addition to the automated setup process, for compatibility with non-WNSK systems it’s also possible to print out the configuration parameters, although the manual data entry process is rather painful. Using the familiar metaphor of inserting a key into an object in order to provide security greatly increases the chances that it’ll actually be used, since it requires almost no effort on the part of the user.

This type of security mechanism is known as a location-limited channel, one in which the user’s credentials are proven by the fact that they have physical access to the device(s) being configured. This is a generalisation of an older technique from the field of secure transaction processing called geographic entitlement, in which users were only allowed to initiate a transaction from a fixed location like a secure terminal room in a brokerage house, for which access required passing through various physical security controls [28]. If the threat model involves attackers coming in over a network, such a location-limited channel is more secure than any fancy (and complex) use of devices such as smart cards and certificates, since the one thing that a network attacker can’t do is plug a physical token into the device that’s being configured.

A similar type of mechanism, which is often combined with a location-limited channel, is a time-limited channel in which two devices have to complete a secure initialisation within a very small time window. An example of such a mechanism is one in which the user simultaneously presses secure initialisation buttons on both devices being configured. The device being initialised would then assume that anything that responded at that exact point in time would be its intended peer device.

An additional countermeasure against a rogue device trying to insert itself into the channel is to check whether more than one response is received (one from the legitimate device and one from the rogue one) within the given time window, and reset the process if this type of tampering is detected. Like tamper-evident seals on food containers, this is a simple, effective measure that stops all but the most determined attacker. This mechanism combines both location-limited channels (the user is demonstrating their authorisation by being able to activate the secure initialisation process) and a time-limited channel (the setup process has to be carried out within a precise time window in order to be successful).

This type of secure initialisation mechanism has already been adopted by some vendors of 802.11 wireless devices who are trying to combat the low level of adoption of secure wireless setups, although unfortunately since there's no industry standard for this they all do it differently. An example of this is the use of location-limited and time-limited channels in Broadcom's SecureEasySetup, which is used for secure initialisation of 802.11 WPA devices via a secure-setup pushbutton or an equivalent mechanism like a mouse click on a PC dialog [29][30]. Since Broadcom are an 802.11 chipset vendor, anyone using their chipsets has the possibility to employ this type of simple security setup. Although only minimal technical details have been published [31], the Broadcom design appears to be an exact implementation of the type of channel described above. This is a good example of effective (rather than theoretically perfect) security design. As David Cohen, a senior product manager at Broadcom, puts it, "The global problem we're trying to solve is over 80 percent of the networks out there are wide open. Hackers are going to jump on these open networks. We want to bring that number down".

A further extension of the location-limited channel concept provides a secure key exchange between two portable devices with wireless interfaces. This mechanism relies for its security on the fact that when transmitting over an open medium, an opponent can't tell which of the two devices sent a particular message, but the devices themselves can. To establish a shared secret, the devices are held together and shaken while they perform the key exchange, with key bits being determined by which of the two devices sent a particular message. Since they're moving around, an attacker can't distinguish one device from the other via signal strength measurements [32]. This is an extremely simple and effective technique that works with an out-of-the-box unmodified wireless device, providing a high level of security while being extremely easy to use.

These types of security mechanisms provide both the ease of use that's necessary in order to ensure that they're actually used, and a high degree of security from outside attackers, since only an authorised user with physical access to the system is capable of performing the initialisation steps.

Note though that you have to exercise a little bit of care when you're designing your location-limited channel. The Bluetooth folks, for example, allowed *anyone* (not just authorised users) to perform this secure initialisation (forced re-pairing in Bluetooth terminology), leading to the sport of bluejacking, in which a hostile party hijacks someone else's Bluetooth device. A good rule of thumb for these types of security measures is to look at what Bluetooth does for its "security" and then make sure that you don't do anything like it.

References

- [1] “In Search of Usable Security: Five Lessons from the Field”, Dirk Balfanz, Glenn Durfee, Rebecca Grinter, and D.K. Smetters, *IEEE Security and Privacy*, **Vol.2, No.5** (September/October 2004), p.19.
- [2] “Leading Geeks: How to Manage and Lead the People Who Deliver Technology”, Paul Glen, David Maister, and Warren Bennis, Jossey-Bass, 2002.
- [3] “Scientist: Complexity causes 50% of product returns”, Reuters, 6 March 2006, <http://www.computerworld.com/hardwaretopics/hardware/story/-0,10801,109254,00.html>.
- [4] “Plug-and-Play PKI: A PKI Your Mother Can Use”, Peter Gutmann, *Proceedings of the 12th Usenix Security Symposium*, August 2003, p.45.
- [5] “Trends and Attitudes in Information Security — An RSA Security e-Book”, RSA Data Security, 2005.
- [6] “Inoculating SSH Against Address Harvesting”, Stuart Schechter, Jaeyon Jung, Will Stockwell, and Cynthia McLain, *Proceedings of the 13th Annual Network and Distributed System Security Symposium (NDSS’06)*, February 2006.
- [7] “Securing Record Communications: The TSEC/KW-26”, Melville Klein, Center for Cryptologic History, National Security Agency, 2003.
- [8] “Race conditions in security dialogs”, Jesse Ruderman, <http://www.squarefree.com/2004/07/01/race-conditions-in-security-dialogs/>, 1 July 2004.
- [9] “Mozilla XPInstall Dialog Box Security Issue”, Secunia Advisory SA11999, <http://secunia.com/advisories/11999/>, 5 July 2004.
- [10] “Race conditions in security dialogs”, Jesse Ruderman, <http://archives.neohapsis.com/archives/fulldisclosure/2004-07/0264.html>, 7 July 2004.
- [11] “Microsoft Internet Explorer Keyboard Shortcut Processing Vulnerability”, Secunia Research, http://secunia.com/secunia_research/2005-7/advisory/, 13 December 2005.
- [12] “Internet Explorer Suppressed “Download Dialog” Vulnerability”, Secunia Research, http://secunia.com/secunia_research/2005-21/advisory/, 13 December 2005.
- [13] “Bugzilla Bug 162020: pop up XPInstall/security dialog when user is about to click”, https://bugzilla.mozilla.org/show_bug.cgi?query_format=specific&order=relevance+desc&bug_status=__open__&id=162020.
- [14] “Disable Extension Install Delay (Firefox)”, [http://kb.mozillazine.org/Disable_Extension_Install_Delay_\(Firefox\)](http://kb.mozillazine.org/Disable_Extension_Install_Delay_(Firefox)).
- [15] “MR Tech Disable XPI Install Delay”, Mel Reyes, <https://addons.mozilla.org/firefox/775/>, 20 Apr 2006.
- [16] “MR Tech Disable XPI Install Delay”, 23 March 2007, <https://addons.mozilla.org/firefox/775/>.
- [17] “Users are not dependable — how to make security indicators to better protect them”, Min Wu, *Proceedings of the First Workshop on Trustworthy Interfaces for Passwords and Personal Information*, June 2005.
- [18] “Race conditions in security dialogs”, Jesse Ruderman, <http://www.squarefree.com/2004/07/01/race-conditions-in-security-dialogs/>, 1 July 2004.
- [19] “Apple Human Interface Guidelines”, Apple Computer Inc, November 2005.
- [20] “Programmers are People, Too”, Ken Arnold, *ACM Queue*, **Vol.3, No.5** (June 2005), p.54.
- [21] “Case Study: Thunderbird’s brittle security as proof of Iang’s 3rd Hypothesis in secure design: there is only one mode, and it’s secure”, Ian Grigg, 23 July 2006, <http://financialcryptography.com/mt/archives/000755.html>.
- [22] “GUI Bloopers: Don’ts and Do’s for Software Developers and Web Designers”, Jeff Johnson, Morgan Kaufmann, 2000.

- [23] “The Importance of Usability Testing of Voting Systems”, Paul Herson, Richard Niemi, Michael Hanmer, Benjamin Bederson, Frederick Conrad, and Michael Traugott, *Proceedings of the 2006 Usenix/Accurate Electronic Voting Technology Workshop*, August 2006.
- [24] “Re: Intuitive cryptography that’s also practical and secure”, Andrea Pasquinucci, posting to the cryptography@metzdowd.com mailing list, message-ID 20070130203352.GA17174@old.at.home, 30 January 2997.
- [25] “The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places”, Byron Reeves and Clifford Nass, Cambridge University Press, 1996.
- [26] “The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How To Restore The Sanity”, Alan Cooper, Sams, 1999.
- [27] “About Face 2.0: The Essentials of Interaction Design”, Alan Cooper and Robert Reimann, Wiley, 2003.
- [28] “Principles of Transaction Processing”, Philip Bernstein and Eric Newcomer, Morgan Kaufman, 1997.
- [29] “Another Take on Simple Security”, Glenn Fleishman, 6 January 2005, <http://wifinetnews.com/archives/004659.html>.
- [30] “Under the Hood with Broadcom SecureEasySetup”, Glenn Fleishman, 12 January 2005, <http://wifinetnews.com/archives/004685.html>.
- [31] “Securing Home WiFi Networks: A Simple Solution Can Save Your Identity”, Broadcom white paper Wireless-WP200-x, 18 May 2005.
- [32] “Shake them Up!: A movement-based pairing protocol for CPU-constrained devices”, Claude Castelluccia and Pars Muta, *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys ’05)*, June 2005, p.51.

Security User Interaction

An important part of the security usability design process is how to interact with users of the security application in a meaningful manner. The following section looks at various user interaction issues and discusses some solutions to user communications problems.

Speaking the User's Language

When interacting with a user, particularly over a topic as complex as computer security, it's important to speak their language. To evaluate a message presented by a security user interface, users have to be both motivated and able to do so. Developers who spend their lives immersed in the technology that they're creating often find it difficult to step back and view it from a non-technical user's point of view, with the result that the user interface that they create assumes a high degree of technical knowledge in the end user. An example of the type of problem that this leads to is the typical jargon-filled error message produced by most software. Geeks love to describe the problem, when they should instead be focusing on the solution. While the maximum amount of detail about the error may help other geeks diagnose the problem, it does little more than intimidate the average user.

The easiest way to determine how to speak the user's language when your application communicates with them is to ask the users what they'd expect to see in the interface. Studies of users have shown however that there are so many different ways to describe the same sorts of things that using the results from just one or two users would invariably lead to difficulties when other users expecting different terms or different ways of explaining concepts use the interface.

A better alternative is to let users vote on terminology chosen from a list of user-suggested texts and then select the option that garners the most votes. A real-world evaluation of this approach found that users of the interface with the highest-polling terminology made between *two and five times* less mistakes than when they used the same interface with the original technical terminology, the interface style that's currently found in most security applications.

The same study found that after prolonged use, error rates were about the same for both interfaces, indicating that, given enough time, users can eventually learn more or less anything... until an anomalous condition occurs, at which point they'll be completely lost with the technical interface.

In a similar vein, consider getting your user manuals written by non-security people to ensure that the people writing the documentation use the same terminology and have the same mindset as those using it. You can always let the security people nitpick the text for accuracy after it's finished.

Effective Communication with Users

In addition to speaking the user's language, you also need to figure out how to effectively communicate your message to them and turn the click, whirr response into controlled responding in which users react based on an actual analysis of the information that they've been given. Previous sections have pointed out a number of examples of ineffective user communication which would imply that this is tricky area to get right, however in this case we're lucky to have an entire field of research (with the accompanying industries of advertising and politics) dedicated to the effective communication of messages. For example social psychologists have determined that a request made with an accompanying explanation is far more likely to elicit an appropriate response than the request on its own [1]. So telling the user that something has gone wrong and that continuing with their current course of action is dangerous "because it may allow criminals to steal money from your bank account" is far more effective than just the generic warning by itself.

The text of this message also takes advantage of another interesting result from psychology research: People are more motivated by the fear of losing something than

the thought of gaining something [2][3]. For example doctors' letters warning smokers of the number of years of life that they'd lose by not giving up smoking have been found to be more effective than ones that describe the number of extra years they'd have if they do kick the habit [4]. This has interesting ramifications. Depending on whether you frame an issue as a gain or a loss, you can completely change people's answers to a question about it. The theory behind this was developed by Nobel prize-winners Daniel Kahneman and Amos Tversky under the name Prospect Theory [5][6]. In Kahneman and Tversky's original experiment, subjects were asked to choose between a sure gain of \$500 and a 50% chance of gaining \$1000 / 50% chance of gaining nothing. The majority (84%) chose the sure gain of \$500. However, when the problem was phrased in reverse, with subjects being told they would be given \$1000 with a sure loss of \$500 or a 50% chance of losing the entire \$1000 / 50% chance of losing nothing, only 31% chose the sure loss, even though it represented the exact same thing as the first set of choices.

One real-world example of the deleterious effects of this can be seen in a study of the working habits New York taxi drivers [7] which found that many of the drivers would set themselves a given earning target each day and quit once they'd reached their target (setting targets can be very motivating when performing boring or tedious activities, which is why it's so popular with people on things like exercise programs). However, while this can be a great motivator when there's nothing to be gained or lost (except for weight in an exercise program), it doesn't work so well when there's more at stake than this. In the case of the taxi drivers, what they were doing was quitting early when they were making good money, and working longer hours when they were earning little. If instead they had worked longer hours on good days and quit early on bad days, their earnings would have increased by 15%. Simply working the same hours each day would have increased their income by 8% (this result is directly contrary to supply-side economics, which argues that if you increase wages, people will work more in order to earn more).

Taking advantage of the findings from Prospect Theory, the previous message was worded as a warning about theft from a bank account rather than a bland reassurance that doing this would keep the user's funds safe. As the discussion of the rather nebulous term "privacy" in the previous chapter showed, some fundamental concepts related to security, and users' views of security, can in fact only be defined in terms of what users will lose rather than anything that they'll gain.

An additional important result from psychology research is the finding that if recipients of such a fear appeal aren't given an obvious way of coping then they'll just bury their heads in the sand and try and avoid the problem [8]. So as well as describing the consequences of incorrect action, your message has to provide at least one clear, unambiguous, and specific means of dealing with the problem. The canonical "Something bad may be happening, do you want to continue?" is the very antithesis of what extensive psychological research tells us we should be telling the user.

Another result from psychology research (although it's not used in the previous message example) is that users are more motivated to think about a message if it's presented as a question rather than an assertion. The standard "Something bad may be happening, do you want to continue?" message is an assertion dressed up as a question. "Do you want to connect to the site even though it may allow criminals to steal money from your bank account?" is a question that provides users with appropriate food for thought for the decision that they're about to make. A button labelled "Don't access the site" then provides the required clear, specific means of dealing with the problem.

A further psychological result that you can take advantage of is the phenomenon of social validation, the tendency to do something just because other people (either an authority figure or a significant number of others) have done it before you. This technique is well-understood and widely used in the advertising and entertainment industries through tricks such as salting donation boxes and collection trays with a small amount of seed money, the use of claque (paid enthusiastic audience members) in theatres, and the use of laugh tracks in TV shows. The latter is a good example of

applying psychology to actual rather than claimed human behaviour: both performers and the audience dislike laugh tracks, but entertainment companies keep using them for the simple reason that they work, increasing viewer ratings for the show that they're used with. This is because the laugh track, even though it's obviously fake, triggers the appropriate click, whirr response in the audience and provides social validation of the content. Laugh tracks are the MSG of comedy. Even though audience members, if asked, will claim that it doesn't affect them, real-world experience indicates otherwise. The same applies for many of the other results of psychology research mentioned above — you can scoff at them, but that won't change the fact that they work when applied in the field.

You can use social validation in your user interface to guide users in their decision-making. For example when you're asking the user to make a security-related decision, you can prompt them that “most users would do xyz” or “for most users, xyz is the best action”, where xyz is the safest and most appropriate choice. This both provides them with guidance on what to do (which is particularly important in the common case where the user won't understand what it is that they're being asked) and gently pushes them in the direction of making the right choice, both now and in the future where this additional guidance may not be available.

If you'd like to find out more about this field, some good starting points are *Influence: Science and Practice* by Robert Cialdini, *Persuasion: Psychological Insights and Perspectives* by Timothy Brooks and Melanie Green, and *Age of Propaganda: Everyday Use and Abuse of Persuasion* by Anthony Pratkanis and Elliot Aronson (if the phishers ever latch onto books like this, we'll be in serious trouble).

As with the user interface safety test that was described in the section on safe defaults, there's a relatively simple litmus test that you can apply to the messages that you present to users. Look at each message that you're displaying to warn users of a security condition and see if they deal with the responses “Why?” and “So what?”. For example you may be telling the user that “The server's identification has changed since the last time that you connected to it”. So what? “This may be a fake server pretending to be the real thing, or it could just mean that the server software has been reinstalled”. So what? “If it's a fake server then any information that you provide to it may be misused by criminals. Are you sure that you really want to continue?”. Finally the user knows why they're being shown the dialog! The “Why?” and “So what?” tests may not apply to all dialogs (usually only one applies to any particular dialog), but if the dialog message fails the test then it's a good indication that you need to redesign it.

Design Example: Connecting to a Server whose Key has Changed

Let's look at a design exercise for speaking the user's language in which a server's key (which is usually tied to its identity) has changed when the user connects to it. Many applications will present the user with either too little information (“The key has changed, continue?”), too much information (a pile of incomprehensible X.509 technobabble, in one PKI usability study *not a single user* was able to make any sense of the certificate information that Windows displayed to them [9]), or the wrong kind of information (“The sky is falling, run away”).



Figure 28: Internet Explorer certificate warning dialog

The standard certificate dialog used by Internet Explorer is shown in Figure 28. The typical user's response to this particularly potent latent pathogen will be something like "What the &*&#@*! is that supposed to mean?", and this is the improved version — earlier versions were even more incomprehensible (recognising the nature of this type of question, pre-release versions of Windows '95 used the text "In order to demonstrate our superior intellect, we will now ask you a question you cannot answer" as a filler where future text was to be added [10]). A few rare users may click on "View Certificate", but then they'll have no idea what they're supposed to be looking for there. In any case this additional step is completely pointless since if the certificate's contents can't be verified there's no point in examining them as the certificate's creators could have put anything they wanted in there.

In addition, users have no idea what the certifying authority (CA) that's mentioned in the dialog is. In one PKI usability study carried out with experienced computer users, 81% identified VeriSlim as a trusted CA (VeriSlim doesn't exist), 84% identified Visa as a trusted CA (Visa is a credit card company, not a CA), and no-one identified Saunalahden as a trusted CA (Saunalahden is a trusted CA located in Finland) [9]. 22% of these experienced users didn't even know what a CA was, and an informal survey of typical (non-experienced) users was unable to turn up anyone who knew what a CA was. In situations like this, applying judgemental heuristics (in other words guessing) makes perfect sense, since it's completely unclear which option is best. Since (from the user's point of view) the best option is to get rid of the dialog so that they can get on with their work, they'll take whatever action is required to make it go away.

Finally, the dialog author has made no attempt to distinguish between different security conditions — a day-old expired certificate is more benign than a year-old expired certificate, which in turn is more benign than a certificate belonging to another domain or issued by an unknown CA. The dialog doesn't even bother to filter out things that aren't a problem ("the certificate date is valid") from things that are ("the name on the certificate is invalid"). This is simply a convenient (to the application developer) one-size-fits-all dialog that indicates that something isn't quite right somewhere, and would the user like to ignore this and continue anyway. The only good thing that can be said about this dialog is that the default action is not 'Yes', requiring that the user at least move the mouse to dismiss it.

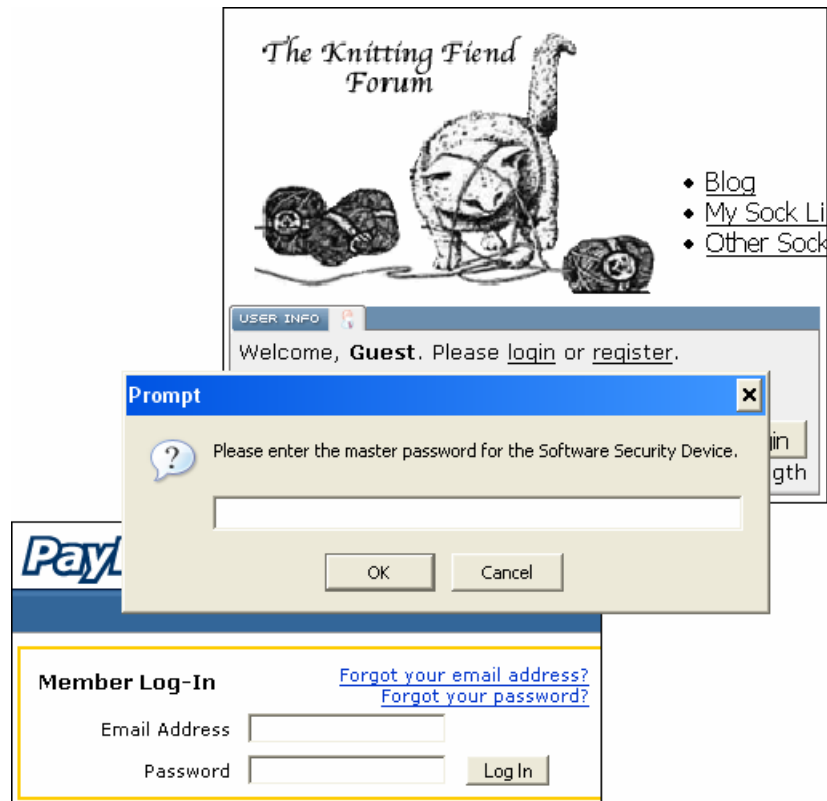


Figure 29: One-size-fits-all password entry

The inability to distinguish between different security levels is endemic to other browsers as well. For example Firefox uses a single master password to protect all secrets in the system, whether it's the password for the Knitting Pattern Weekly or the password for your online bank account. As shown in Figure 29, users end up either over-protecting something of little to no value (a long, complex master password used to protect access to Knitting Pattern Weekly) or under-protecting something of considerable value (a short, easy-to-type master password used to protect access to your PayPal account).

A better trade-off would have been to break the master-password control mechanism into two or even three levels, one with no master password at all for the large number of sites that require nuisance signups before they'll allow you to participate, one with a relatively easy-to-type master password for moderate-value sites, and a high-security one with a more complex master password that has to be re-entered on each use, for high-value sites such as online bank account access. Having to explicitly enter the high-value master password both makes users more aware of the consequences of their actions and ensures that a master password-enabled action performed some arbitrary amount of time in the past can't be exploited later when the user browses to a completely unrelated (and possibly malicious) site. Finally, since the browser now knows (via the use of the standard vs. high-value master password selection) that the user is performing a high-value transaction, it can apply additional safety checks such as more stringent filtering of what information gets sent where. Doing this for every site visited would "break" a lot of sites, but by taking advantage of the inside knowledge of which sites are considered important by the user, the browser can only apply the potentially site-breaking extra security measures to the cases where it really matters.

Internet Explorer 7 finally appears to be taking some steps towards fixing the incomprehensible certificate warning problem, although it remains to be seen how effective these measures will really be. For example one of the measures consists of warning users when they visit suspected phishing sites, even though an AOL UK user survey found that 84% of users didn't know what phishing was and were therefore unlikely to get anything from the warning. Another potential problem with the

proposed changes is a profusion of URL-bar colour-codes for web sites and colour-code differences between browsers — Firefox uses yellow for SSL-secured sites while MSIE 7 uses it to indicate a suspected phishing site, so that Firefox users will think an MSIE 7 phishing site is secured with SSL while MSIE 7 users will think that SSL-secured sites are phishing sites (this colour-change booby trap is a bit like changing the meaning of red and green traffic lights in different cities). Finally, there are plans to display the certificate issuer name in the URL bar alternating with the certificate subject name, a proposal that has the potential to equal the `<blink>` tag in annoyance value (displaying it as a tooltip would be a better idea), as well as being more or less meaningless to most users.

Look at the problem from the point of view of the user. They're connecting to a server that they've connected to many times in the past and that they need to get to now in order to do their job. Their natural inclination will be to do whatever it takes to get rid of the warning and connect anyway, making it another instance of the "Do you want this message to go away" problem presented earlier.

Your user interface should therefore explain the problem to them, for example "The server's identification has changed since the last time that you connected to it. This may be a fake server pretending to be the real thing, or it could just mean that the server software has been reinstalled. If it's a fake server rather than any information that you provide to it may be misused by criminals. Are you sure that you really want to continue?". Depending on the severity of the consequences of connecting to a fake server, you can then allow them to connect anyway, connect in a reduced-functionality "safe" mode such as one that disallows uploads of (potentially sensitive) data to the possibly-compromised server and is more cautious about information coming from the server than usual, or perhaps even require that they first verify the server's authenticity by checking it with the administrator who runs it. If you like, you can also include an "Advanced" option that displays the usual X.509 gobbledegook.

An alternative approach, which is somewhat more drastic but also far more effective, is to treat a key or certificate verification failure in the same way as a standard network server error. If the user is expecting to talk to a server in a secure manner and the security fails, then that's a fatal error, not just a one-click speed-bump. This approach has already been adopted by some newer network clients such as Linux's `xsupplicant` and Windows XP's PEAP (Protected Extensible Authentication Protocol) client. This is a failure condition that users will instinctively understand, and that shifts the burden from the user to the server administrators. Users no longer have to make the judgement call, it's now up to the server administrators to get their security right. In an indistinguishable-from-placebo environment this is probably the only safe way to handle key/certificate verification errors.

There's also a third alternative that runs the middle ground between these two extremes, which provides a mechanism for allowing the user to safely accept a new key or certificate. Instead of allowing the user to blindly click 'OK' to ignore the error condition, you can require that they enter an authorisation code for the new key or certificate that they can only obtain from the server administrator or certificate owner, forcing them to verify the key before they enable its use. The "authorisation code" is a short string of six to eight characters that's used to calculate an HMAC (hashed Message Authentication Code, a cryptographic checksum that incorporates an encryption key so that only someone else who has the key can recreate the checksum) of the new key or certificate, with the first two characters being used as the HMAC key and the remaining characters being the (truncated) HMAC result, as illustrated in Figure 30. For example if you set the first two characters to "ab" then computing HMAC("ab", key-or-certificate) will produce a unique HMAC value for that key or certificate. Taking the base64 encoding of the last few bytes of the HMAC value (say, "cdefg") produces the six-character authorisation code "abcdefg". When the user enters this value, their application performs the same calculation and only permits the use of the key or certificate if the calculated values match.

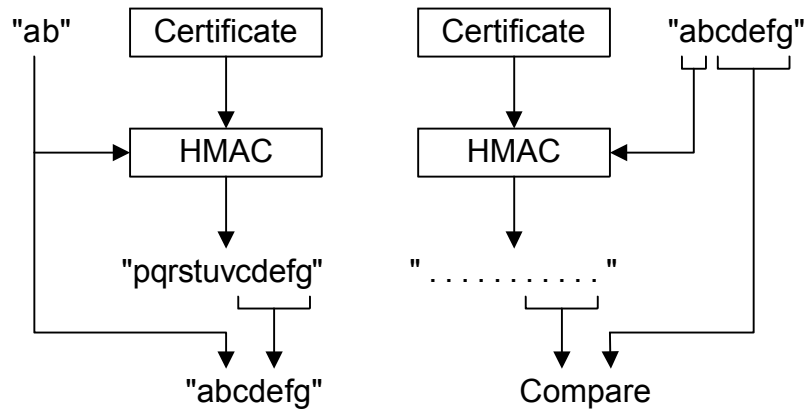


Figure 30: Generation of a certificate authorisation code

Obviously this use of an HMAC as a salted hash isn't terribly secure, but it doesn't have to be — what it's doing is raising the bar for an attacker, changing the level of effort from the trivial (sending out phishing/spam email) to nontrivial (impersonating an interactive communication between the key/certificate owner and the user). A determined attacker can still do this, but their job has suddenly become a whole lot harder since they now have to control the authorisation side-channel as well. Incidentally, the reason for using a keyed hash (the HMAC) rather than a standard hash is that most software already displays a hash of the key to the user, usually labelled as a fingerprint or thumbprint. If they copied this value across to the authorisation check, the user could bypass the separate side-channel that they'd otherwise be forced to use.

One thing that SSH does which SSL/TLS should really copy is keep a record of whether a trusted domain (that is, a server using SSH or SSL/TLS) has been visited before, and as an extension how many times it's been visited before (neither SSH nor SSL/TLS currently do the latter). With this information at hand the application can change its behaviour depending on whether this is a first visit, an infrequent visit, or a frequent visit. For example if the user frequently visits

<https://www.paypal.com> but is now visiting <https://www.paypai.com> for the first time, the application can warn that this is a potentially suspicious site that the user doesn't normally visit. This has been shown to significantly increase a user's ability to detect spoofed web sites [11]. Because SSL use is infrequent and is normally only applied to sites where the user has to enter valuable information such as credit card details, you can take advantage of the fact that the users themselves will be telling you when to be careful.

If you implement this measure you need to be careful to mask the list of hosts visited to avoid both privacy concerns and the ability of an attacker who gains access to the list to perform address-harvesting of the list of known/trusted hosts, a particular problem with SSH's `known-hosts` mechanism. Various workarounds for this problem are possible [12], the simplest of which is to store a MAC of the host name rather than the actual name. Since all that we're interested in is a presence-check, a comparison of the MAC value will serve just as well as a comparison of the full name.

Design Example: Inability to Connect to a Required Server

A variation of the problem in the previous design example occurs when you can't connect to the other system at all, perhaps because it's down, or has been taken offline by a DDoS attack, or because of a network outage. Consider for example the use of OCSP, a somewhat awkward online CRL query protocol, in combination with a web browser. The user visits a couple of sites with OCSP enabled, and everything works fine (although somewhat slowly, because of the extra OCSP overhead). Then they switch to a disconnected LAN, or a temporary network outage affects access to the OCSP server, or some similar problem occurs. Suddenly their browser is complaining whenever they try to access SSL sites (such problems are already being

reported with OCSP-enabled browsers like FireFox [13][14]). When they disable OCSP, everything works again, so obviously there was a problem with OCSP. As a result, they leave it disabled, and don't run into any more problems accessing SSL servers.

The failure pattern that we see here is that this is a feature that provides no directly visible benefit to the user while at the same time visibly reducing reliability. Since it's possible to turn it off and it's not necessary to turn it on again, it ends up disabled. The survivability of such a "feature" is therefore quite low.

What the addition of the extra security features has done is make the system considerably more brittle, reducing its reliability to the lowest common denominator of the web server and the OCSP server. While we've learned to make web servers extremely reliable, we haven't yet done the same for OCSP servers, and it's unlikely that there'll ever be much evolutionary pressure to give them the same level of reliability and performance that web servers enjoy. In fact things seem to be going very much in the opposite direction: since the OCSP protocol is inherently non-scalable, a recent performance "enhancement" was to remove protection against man-in-the-middle attacks, making it possible for a server (or an attacker) to replay an old response instead of having to generate a new one that reflects the true state of the certificate [15].

Exactly such a lowest-common-denominator reliability problem has already occurred with the Windows 2000 PKI implementation. Microsoft hardcoded the URL for a Verisign CRL server into their software, so that attempts to find a CRL for any certificate (no matter who the CA actually was) went to this particular Verisign server. When Verisign reorganised their servers, the URL ceased to function. As a result, any attempt to fetch a CRL resulted in Windows groping around blindly on the net for a minute, after which it timed out and continued normally.

In practice this wasn't such a big problem because CRL checking was turned off by default so almost no-one noticed, but anyone who did navigate down through all the configuration dialogs to enable it quickly learned to turn it off again. Another example is found in some JCE implementations, in which the JVM checks a digital signature on the provider when it's instantiated. This process involves some form of network access, with the results being the same as for the Windows CRL check — the JVM gropes around for awhile and then times out and continues anyway. All the user notices of this is the fact that the application stalls for quite some time every time it starts (one Java developer referred to this process as "being held captive to some brain-dead agenda" [16]).

This is another example of the Simon Says problem. From the certificate (or site) owner's point of view, it's in their best interests *not* to use OCSP, since this reduces the chances of site visitors being scared away by error messages when there's a problem with the OCSP server. The nasty misfeature of this mechanism is that it's only when you *enable* the use of OCSP that users start seeing indications of trouble — if you just go ahead and use the certificate without trying to contact the OCSP server, everything seems to work OK.

To determine how to fix this (or whether it needs fixing at all), it's instructive to perform a cost/benefit analysis of the use of OCSP with SSL servers. First of all, it's necessary to realise that OCSP can't prevent most type of phishing attacks. Since OCSP was designed to be fully compatible with CRLs and can only return a negative response, it can't be used to obtain the status of a forged or self-signed certificate. For example when fed a freshly-issued certificate and asked "Is this a valid certificate", it can't say "Yes" (a CRL can only answer "revoked"), and when fed an Excel spreadsheet it can't say "No" (the spreadsheet won't be present in any CRL). More seriously, CRLs and OCSP are incapable of dealing with a manufactured-certificate attack in which an attacker issues a certificate claiming to be from a legitimate CA — since the legitimate CA never issued it, it won't be in its CRL, therefore a blacklist-based system can't report the certificate as invalid. Finally, when used with soundalike certificates in secure phishing attacks, the certificate will be reported as not-revoked (valid) by OCSP (since it was issued by a legitimate CA)

until such time as the phish is discovered, at which point the site will be shut down by the hosting ISP, making it mostly irrelevant whether its certificate is revoked or not.

The result of this analysis is that there's no real benefit to the use of OCSP with SSL servers, but considerable drawbacks in the form of adverse user reaction if there's a problem with the OCSP server. The same problem affected the NSA-designed system mentioned earlier, in which the users' overriding concern was availability and not confidentiality/security.

Looking beyond the problems inherent in the use of the OCSP mechanism, we can use the X.509 CRL reason codes used by OCSP to try and determine whether revocation checking is even necessary. Going through each of the reason codes, we find that "key compromise" is unlikely to be useful unless the attacker helpfully informs the server administrator that they've stolen their key, "affiliation changed" is handled by obtaining a new certificate for the changed server URL, "superseded" is handled in the same way, and "cessation of operation" is handled by shutting down the server. In none of these cases is revocation of much use.

No doubt some readers are getting ready to jump up and down claiming that removing a feature in this manner isn't really an example of security user interface design. However, as the analysis shows, it's of little to no benefit, but potentially a significant impediment. The reason why OCSP was used in this design example is because such cases of redundancy⁴ only seem to occur in the PKI world. Outside of PKI, they're eliminated by normal Darwinian processes, but these don't seem to apply to PKI. So this is an example of a user interface design process that removes features in order to increase usability instead of adding or changing them.

Use of Visual Cues

The use of colour can play an important role in alerting users to safe/unsafe situations. Mozilla-based web browsers updated their SSL indication mechanism from the original easily-overlooked tiny padlock at the bottom of the screen to changing the background colour of the browser's location bar when SSL is active and the certificate is verified, as shown in Figure 31 (if you're seeing this on a black-and-white printout, the real thing has a yellow background). Changing the background colour or border of the object that the user is looking at or working with is an extremely effective way of communicating security information to them, since they don't have to remember to explicitly look elsewhere to try and find the information. The colour change also makes it very explicit that something special has occurred with the object that's being highlighted (one usability study found that the number of users who were able to avoid a security problem doubled when different colours were used to explicitly highlight security properties).

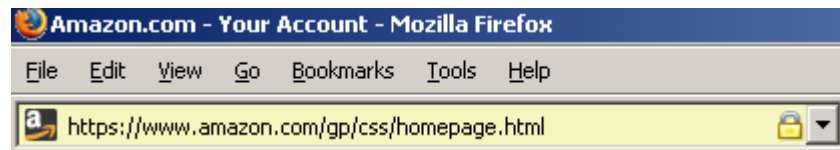


Figure 31: Unambiguous security indicators for SSL

When you do this, you need to take care to avoid the angry-fruit-salad effect in which multiple levels of security indicator overlap to do little more than confuse the user. For example a copy of Firefox with various useful additional security plugins installed might have a yellow URL bar from Firefox telling the user that SSL is in use, a red indicator from the Petnames plugin telling the user that it's an unrecognised site, a green indicator from the Trustbar plugin telling the user that they've been there before, and another yellow indicator from an OCSP responder indicating the that OCSP status isn't available.

⁴ "Redundancy" is a term used to refer to fault-prone systems run in parallel so that if one fails another can take over. "Reduncandy" refers to fault-prone systems run in series so that a fault in any will bring them all down.

When you're using colour or similar highlighting methods in your application, remember that the user has to be aware of the significance of the different colours in order to be able to make a decision based on them, that some users may be colour-blind to particular colour differences, and that colours have different meanings across different cultures. For example the colour red won't automatically be interpreted to indicate danger in all parts of the world, or its meaning as a danger/stop signal may work differently in different countries. In the UK, heavy machinery is started with a green button (go) and stopped with a red button (stop). Across the channel in France, it's started with a red button (a dangerous condition is being created) and stopped with a green button (it's being rendered safe). When it comes to colour-blindness, about 8% of the population will be affected, with the most common type being partial or complete red-green colour-blindness (in case you're wondering how this works with traffic lights, they have a fixed horizontal ordering so that colour-blind people still have some visual indication through the position of the light that's lit). Ensuring that your interface also works without the use of colour, or at least making the colour settings configurable, is one way of avoiding these problems [17]. If you ever get a chance to compare the Paris and London underground/tube/subway maps, see if you can guess which one was designed with colour-blind users in mind.

Here's a simple way of handling visual indications for colour-blind users. Use the configuration dialog shown in Figure 32, which provides a simple, intuitive way of letting colour-blind users choose the colour scheme that provides the best visual indication of a particular condition.

Which of these looks right?

Danger	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Caution	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Safe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 32: Visual cue colour chooser

Another way to handle colour issues, which works if there are only one or two colours in use (for example to indicate safe vs. unsafe) and the colour occurs in a long band like a title bar, is to use a colour gradient fading from a solid colour on one side to a lighter shade on the other. This makes the indicator obvious even to colour-blind users.

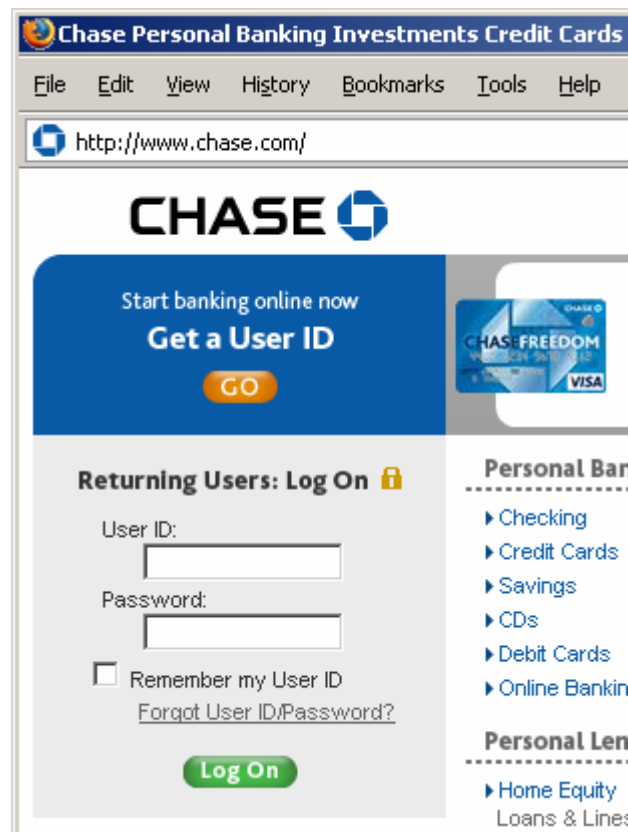


Figure 33: Unprotected login screen

Visual cues can also be used to provide an indication of the absence of security, although how to effectively indicate the absence of a property is in general a hard problem to solve (see the earlier discussion on the Simon Says problem). For example password-entry fields in dialog boxes and web pages always blank out the typed password (typically with asterisks or circles) to give the impression that the password is secret and/or protected in some manner. Even if the password is sent in the clear without any protection (which is the case for many web pages, see Figure 33), it's still blanked out in the user display. Conversely, information such as credit card numbers, which are usually sent encrypted over SSL connections, are displayed to the user. By not blanking the password field when there's no protection being used (see Figure 34), you're providing instant, unmistakable feedback to the user that there's no security active.



Figure 34: Unprotected login screen, with (in)security indicators

The fact that their password is being shown in the clear will no doubt make many users nervous, because they've been conditioned to seeing it masked out. However,

making users nervous is exactly what this measure is meant to do: a password displayed in this manner may now be vulnerable to shoulder surfing, but it's even more vulnerable to network sniffing and similar attacks (this is disregarding the question of why a user would be accessing sensitive information in a password-protected account in an environment that's vulnerable to shoulder-surfing in the first place).

Displaying the password in the clear makes real and present what the user cannot see, that there's no security active to protect either their password or any sensitive information that the password will unlock. To avoid adverse user reactions, you should add a tooltip "Why is my password showing?" to the password-entry box when the password isn't masked (see Figure 34), explaining to users what's going on and the potential consequences of their actions (tooltips have other names in different environments, for example OS X calls them help tags). The tooltips act as a clue box in this type of application.

Although studies of users have shown that they completely ignore tooltips in (equally-ignored) user interface elements like security toolbars [18], it's only acting as an optional explanatory element in this case, so it doesn't matter if users ignore it or not. In any case since the non-masked password has already got their attention and they'll be after an explanation for its presence, the tooltip provides this explanation if they need it.

This combination of measures provides both appropriate warning and enough information for the user to make an informed decision about what to do next.

[anence in Semiconductor Devices](#), specifically remanence issues in static was presented at the [2001 Usenix Security Symposium](#), the [slides for the](#) I read the full paper.


ranging in quality from awful (Kerberos 4, SESAME, and the original at give very broad recommendations on random number generation, none or [Software Generation of Practically Strong Random Numbers](#),  y OS-independant random data a cumulator and generator and an

Figure 35: TargetAlert displaying browser link activation details

Tooltip-style hints are useful in other situations as well. For example you can use them on mouseover of a screen element to provide additional security-relevant information about what'll happen when the user activates that element with the mouse. An example of this type of behaviour is shown in Figure 35, in which the TargetAlert plugin for the Mozilla web browser is indicating that clicking on the link will cause the browser to hang trying to load the Adobe Acrobat plugin. TargetAlert has other indicators to warn the user about links that are executable, pop up new windows, execute Javascript, and so on.

the PSU to a [Zalman Reserator](#) [zalman.co.kr],
and Northbridge. In order to silence my HDD I
nected to the Reserator.

Figure 36: Slashdot displaying the true destination of a link

A variation of this technique is used by the Slashdot web site to prevent link spoofing, in which a link that appears to lead to a particular web site instead leads to a completely different one. This measure, shown in Figure 36, was introduced to counter the widespread practice of having a link to a supposedly informational site lead instead to [goatse.cx](#) (a site that may euphemistically be described as "not work-safe"), the Internet equivalent of a whoopee cushion. A similar such simple measure,

displaying on mouseover the domain name of the site that a link leads to, would help combat the widespread use of disguised links in phishing emails.

```
<form action="http://www.bankofamerica.com">
  <input type="password" name="password">
  <input type="submit" value="submit"
    onclick='this.form.action="http://www.phishing.com"'>
</form>
```

Figure 37: User interface spoofing using Javascript

When you use measures like this, make sure that you display the security state in a manner that can't be spoofed by an attacker. For example web browsers are vulnerable to many levels of user interface spoofing using methods such as HTML to change the appearance of the browser or web page, or Javascript or XUL to modify or over-draw browser UI elements. An example of this type of attack, which uses Javascript to redirect a typed password to a malicious web site, is shown in Figure 37. A better-known example from the web is the use of cross-site scripting (XSS), which allows an attacker to insert Javascript into a target's web page. One such attack, employed against financial institution sites like Barclay's Bank and MasterCard, allowed an attacker to deliver their phishing attack over SSL from the bank's own secure web server [19]. To protect against these types of attack, you should ensure that your security-status display mechanism can't be spoofed or overridden by external means.

Design Example: TLS Password-based Authentication

A useful design exercise for visual cues involves the use of TLS' password-based authentication (TLS-PSK). What's required to effectively apply TLS-PSK are three things:

1. A means of indicating that TLS-PSK security is in effect, namely that both client and server have performed a mutual authentication process.
2. An unmistakable means of obtaining the user password that can't be spoofed by something like a password-entry dialog on a normal web page.
3. An unmistakable link between the TLS-PSK authentication process and the web page that it's protecting.

The obvious way to meet the first requirement is to set the URL bar to a distinctive colour when TLS-PSK is in effect. Users have become accustomed to this as a security indicator because some browsers already do this for standard SSL/TLS, setting the URL bar to light yellow, and browser vendors have indicated that they plan to extend this usage in future browser versions. For TLS-PSK we'll use light blue to differentiate it from the standard SSL/TLS security, producing a non-zero Hamming weight for the security indicators. Using an in-band indicator (for example something present on the web page) is a bad idea, both because as the previous section showed it's quite easily spoofable by an attacker, and because usability tests on such an interface have shown that users just consider it part of the web page and don't pay any attention to it [9].

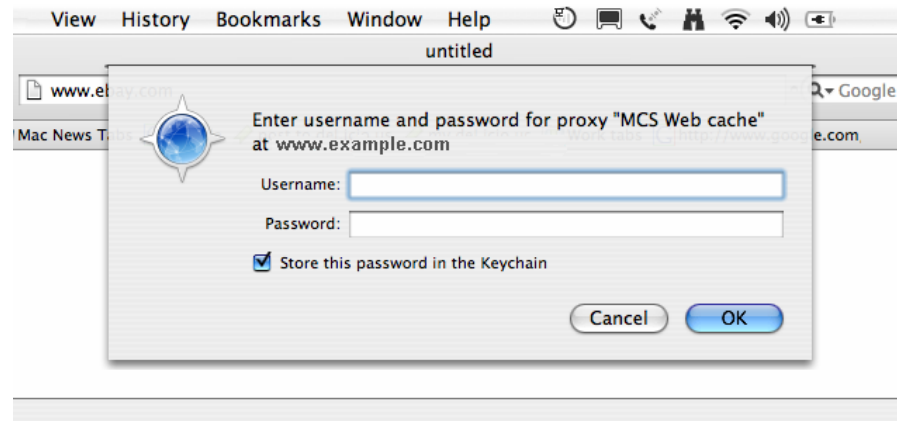


Figure 38: Non-spoofable password-entry dialog

To meet the second and third requirements, instead of popping up a normal password-entry dialog box in front of the web page (which could be coming from hostile code on the web page itself), we make the blue URL bar zoom out into a blue-tinted or blue-bordered password-entry dialog, and then zoom back into the blue URL bar once the TLS-PSK authentication is complete. The Camino browser for OS X already uses a non-spoofable interface of this kind, as shown in Figure 38. When the browser requests a password from the user, the password-entry dialog scrolls out of the browser title bar (outside the normal display area of the browser) in a manner and at a location that no web content can emulate (since this is a complex animation, the single static image of the dialog's final form and location shown above doesn't really do it justice).

This process creates a clear indication even for novice users of a connection between the URL bar indicating that TLS-PSK security is in effect, the TLS-PSK password-entry system, and the final result of the authentication. The user learning task has been simplified to a single bit, "If you don't see the blue indicators and graphical effects, run away".

Finally, this authentication mechanism is an integral part of the critical action sequence. If it's implemented as described above then you can't do TLS-PSK authentication without being exposed to the security interface. Unlike the certificate check in standard SSL/TLS security, you can't choose to avoid it, and as the discussion of users' mental models in a previous section showed, it matches users' expectations of security: When TLS-PSK is in effect, entering your using name and password as a site authenticity check is perfectly valid since only the genuine site will be able to authenticate itself by demonstrating prior knowledge of the name and password. A fake site won't know the password in advance and therefore won't be able to demonstrate its TLS-PSK credentials to the user.

Design Example: Other Password Protection Mechanisms

TLS-PSK is the most powerful password mechanism, but sometimes the need for compatibility with legacy systems means that it's not possible to employ it. There are however a variety of alternatives that you can use that go beyond the current "hand over the user's password to anyone who asks for it" approach. These alternatives work by adding an extra layer of indirection to the password-entry process, sending to the remote system not the actual user password but some unrelated value specific to that particular system. So for example a user password of "mypassword" might translate to a Hotmail password of "5kUqedtM2I", a PayPal password of "Y6WOMZuWLG", and an Amazon password of "xkepKEoVOG". This concept has been around for quite some time, going back more than ten years to the Lucent Personalised Web Assistant, which used a master password supplied to a proxy server (this was before browser plugins) [20][21][22].

The advantage of this extra level of indirection is that it provides password diversification. Every site gets its own unique, random password, so that if one of these derived passwords is ever compromised it won't affect the security of any other

site. Password diversification is an important element in protecting user passwords, since users tend to re-use the same password across multiple sites, with one survey finding that 96% of users reused passwords [23]. This password cross-pollination practice makes it easy for attackers to perform leapfrog attacks in which they obtain the password for a low-value site that users don't take much care to protect and then use it to access a high-value site. The fact that attackers are making use of this has been confirmed by the phishers themselves [24].

An additional benefit to password diversification is that since the derived password is unrelated to the user's actual password, they still get to use strong passwords on every site even if their master password is relatively weak. Finally, this approach provides a good deal of phishing protection. Since passwords are site-specific, a phishing site will be sent a password that's completely unrelated to the one used at the site that it's impersonating.

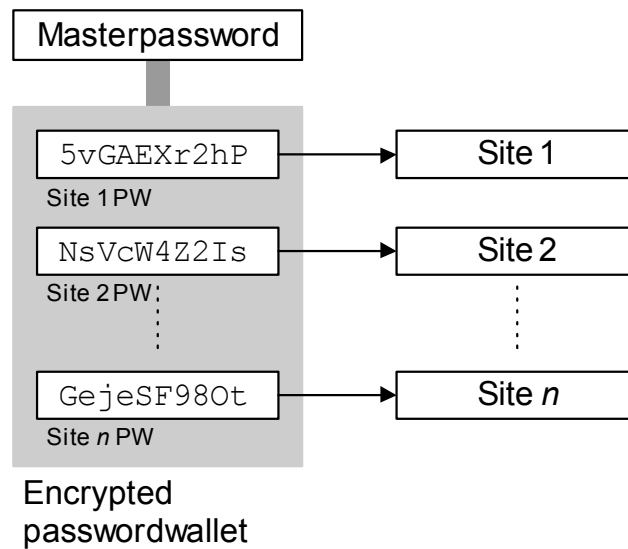


Figure 39: Password diversification using a password wallet

There are two possible approaches to password diversification. The first one is the password wallet technique shown in Figure 39. The user-supplied master password is used to decrypt the wallet, which contains per-site random passwords generated by the application. When the user wants to access a site, the application looks up the appropriate password by server name or URL. If it's a site that the user hasn't visited before, the application can warn the user (in case it's a phishing site) and if they're sure that they want to continue, create a new random password for them.

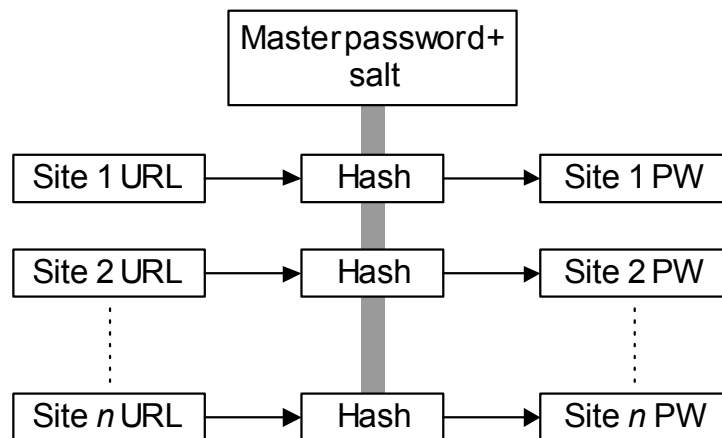


Figure 40: Password diversification using hashing

The second approach to password diversification is shown in Figure 40. The user-supplied master password is hashed with a random salt value and the server name/site

URL to again provide a site-specific password unrelated to the original user password. The random salt makes it impossible for an attacker to guess the user's master password if they acquire one of the site passwords. An additional level of diversification involves hashing the application name into the mix, making the password application-specific as well as site-specific. In this way a compromise of (for example) an SSL/TLS password doesn't compromise the corresponding SSH password.

As with the password wallet approach, this approach guarantees that the spoofed phishing site can never obtain the password for the site that it's impersonating.

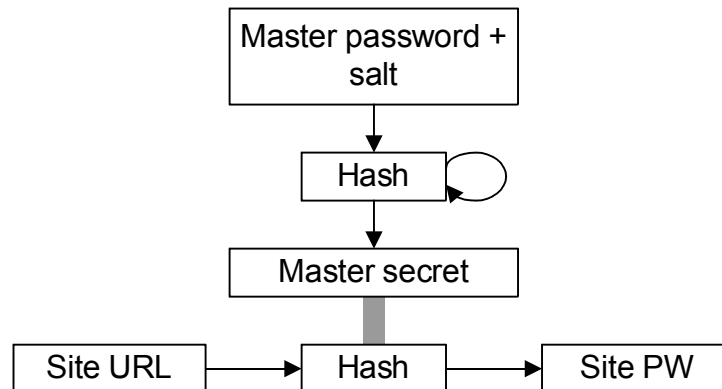


Figure 41: Extra protection for the master password

An enhancement of this technique that provides extra protection against offline password-guessing attacks adds a pre-processing step that converts the master password into a master secret value via a lengthy iterated hashing process and then uses the master secret to generate site passwords instead of applying the master password directly [25]. This additional step is shown in Figure 41, and would typically be done when the software is first installed. By adding this one-off additional step, the time for an attacker to guess each password becomes the sum of the lengthy initial setup time and the quick per-site password generation time. In contrast a legitimate user only experiences the quick per-site password generation time.

Astoundingly, these obvious approaches to password protection, which date back more than ten years, aren't used by any current password-using application. All of them simply connect to anything listening on the appropriate port and hand over the user-entered password (most browsers implement some form of password-storage mechanism, but that just records user-entered passwords rather than managing randomly generated, un-guessable, site-specific ones). The level of interest in this style of password management is demonstrated by the existence of at least half a dozen independently-created Firefox browser plugins that retroactively add this functionality [26], but despite positive third-party evaluations such as "[PwHash] is so seamless that were it installed in every browser since the foundation of the web, users would notice virtually no difference aside from improved security" [25], no browser supports this functionality out of the box. The existence of these various implementations in the form of browser plugins does however provide a nice opportunity for evaluating the usability of various approaches to solving the password-management problem.

The chapter on security usability testing contains further information on requirements for password interfaces.

Design Example: Strengthening Passwords against Dictionary Attacks

One slight drawback to passwords is that they're vulnerable to dictionary attacks, in which an attacker tries every possible word in the dictionary in the hope that one of them is what the user is using as a password. If you're using one of the password-diversification schemes described above, this becomes a great deal more difficult since the passwords are completely random strings and so dictionary attacks don't

work any more. However, there may be cases where you can't do this (for example when the user enters their master password) where it would be good to have some form of protection against a dictionary attack.

Figure 42: Master password entry dialog box

One standard technique for strengthening password protection against dictionary attacks is to iterate the hashing to slow down an attacker. A means of building this into your user interface is shown in Figure 42. This gives the user the choice of a small number of iterations and correspondingly lower dictionary-attack resistance for impatient users, or a larger number of iterations and higher dictionary-attack resistance for more patient users. To determine the correlation between hashing iterations and time, have your application time a small, fixed number of iterations (say 1000) and then use the timing information to mark up the slider controls. This way of doing things has the advantage over hard-coding in a fixed number of iterations that as computers get faster, the attack resistance of the password increases. Conversely, it doesn't penalise users with less powerful machines.

Leave the default processing time at 1 second. Most users won't change this, and it's short enough not to be a noticeable inconvenience. If you want to provide more meaningful feedback on what the delay is buying the user, you can also add an estimate of the resulting protection strength below the slider: "One day to break", "One week to break", and so on.

Legal Considerations

As the earlier section has already pointed out, when you're designing your user interface you need to think about the legal implications of the messages that you present to the user [27]. Aside from the problem of failing to adequately protect users already covered earlier, you also have to worry about over-protecting them in a way that could be seen as detrimental to their or a third party's business. If your security application does something like mistakenly identify an innocent third party's software as malicious, they may be able to sue you for libel, defamation, trade libel/commercial disparagement, or tortious interference, a lesser-known adjunct to libel and defamation in which someone damages the business relationship between two other parties. For example if your application makes a flat-out claim that a program that it's detected is "spyware" (a pejorative term with no widely-accepted meaning) then it had better be *very* sure that it is in fact some form of obviously malicious spyware program. Labelling a grey-area program such as a (beneficial to the user) search toolbar with assorted (not necessarily beneficial to the user)

supplemental functionality as outright spyware might make you the subject of a lawsuit, depending on how affronted the other program's lawyers feel.

This unfortunate requirement for legal protection leads to a direct conflict with the requirement to be as direct with the user as possible in order for the message to sink in. Telling them that program XYZ that your application has detected may possibly be something that, all things considered, they'd prefer not to have on their machine, might be marvellous from a legal point of view but won't do much to discourage a user from allowing it onto their system anyway.

There are two approaches to addressing this inherent conflict of interests. The first (which applies to any security measures, not just the security user interface) is to apply industry best practice as much as possible. For example if there's a particular widely-used and widely-accepted classification mechanism for security issues then using that rather than one that you've developed yourself can be of considerable help in court. Instead of having to explain why your application has arbitrarily declared XYZ to be malicious and prevented it from being installed, you can fall back on the safety net of accepted standards and practices, which makes a libel claim difficult to support since merely following industry practice makes it hard to claim deliberate malicious intent.

A related, somewhat weaker defence if there are no set industry standards is to publicise the criteria under which you classify something as potentially dangerous. In that case it'll be more difficult to sue over a false positive because you were simply following your published policies, and not applying arbitrary and subjective classification mechanisms.

The second defence is to use weasel-words. As was mentioned above, this is rather unfortunate, since it diminishes the impact of your user interface's message on the user. If you're not 100% certain then instead of saying "application XYZ from XYZ Software Corporation is adware", say "an application claiming to be XYZ from XYZ Software Corporation may produce unwanted pop-up messages on your system" (it may be only pretending to be from XYZ Software Corporation, or the pop-up messages could be marginally useful so that not all users would immediately perceive them as unwanted). Since spamware/spyware/adware vendors try as hard as possible to make their applications pseudo-legitimate, you have to choose your wording very carefully to avoid becoming a potential target for a lawsuit. The only thing that saved SpamCop in one spammer-initiated lawsuit was the fact that they merely referred complaints to ISPs (rather than blocking the message) and included a disclaimer that they couldn't verify each and every complaint and that it might in fact be an "innocent bystander" [28], which is great as a legal defence mechanism but less useful as a means of effectively communicating the gravity of the situation to a user.

One simply way of finding the appropriate weasel-words (which was illustrated in the example above) is to describe the properties of a potential security risk rather than applying some subjective tag to it. Although there's no clear definition of the term "adware", everyone will agree that it's a pejorative term. On the other hand no-one can fault you for saying that the application will create possibly unwanted pop-up messages. The more objective and accurate your description of the security issue, the harder it will be for someone to claim in court that it's libellous. This technique saved Lavasoft (the authors of the popular Ad-Aware adware/spyware scanner) in court [29]. The downside to this approach is that it's now up to the user to perform the necessary mental mapping from "potentially unwanted popups" to "adware" (a variant of the `bCanUseTheDamnThing` problem), and not all users will be able to do that.

References

- [1] "The mindlessness of ostensibly thoughtful action: The role of 'placebic' information in interpersonal interaction", Ellen Langer, Arthur Blank, and Benzion Chanowitz, *Journal of Personality and Social Psychology*, **Vol.36**, **No.6** (June 1978), p.635.

- [2] "Gain-Loss Frames and Cooperation in Two-Person Social Dilemmas: A Transformational Analysis", Carsten de Dreu and Christopher McCusker, *Journal of Personality and Social Psychology*, **Vol.72, No.5** (1997), p.1093.
- [3] "The framing of decisions and the psychology of choice", Amos Tversky and Daniel Kahneman, *Science*, **Vol.211, No.4481** (30 January 1981), p.453.
- [4] "Framing of decisions and selection of alternatives in health care", Dawn Wilson, Robert Kaplan, and Lawrence Schneiderman, *Social Behaviour*, **No.2** (1987). p.51.
- [5] "Prospect Theory: An Analysis of Decision under Risk", Daniel Kahneman and Amos Tversky, *Econometrica*, **Vol.47, No.2** (March 1979), p.263.
- [6] "Against the Gods: The Remarkable Story of Risk", Peter Bernstein, John Wiley and Sons, 1998.
- [7] "Taxi Drivers and Beauty Contests", Colin Camerer, *Engineering and Science*, California Institute of Technology, **Vol.60, No.1** (1997), p.11.
- [8] "Age of Propaganda: Everyday Use and Abuse of Persuasion", Anthony Pratkanis and Elliot Aronson, W.H.Freeman and Company, 1992.
- [9] "Design Principles and Patterns for Computer Systems That Are Simultaneously Secure and Usable", Simson Garfinkel, PhD thesis, Massachusetts Institute of Technology, May 2005.
- [10] "In order to demonstrate our superior intellect, we will now ask you a question you cannot answer", Raymond Chen, <http://blogs.msdn.com/oldnewthing/-archive/2004/04/26/120193.aspx>, April 2004.
- [11] "Phishing with Rachna Dhamija", Federico Biancuzzi, 19 June 2006, <http://www.securityfocus.com/columnists/407>.
- [12] "Inoculating SSH Against Address Harvesting", Stuart Schechter, Jaeyon Jung, Will Stockwell, and Cynthia McLain, *Proceedings of the 13th Annual Network and Distributed System Security Symposium (NDSS'06)*, February 2006.
- [13] "VeriSign digital certificates with Firefox", Stuart Fermenick, posting to netscape.public.mozilla.crypto, 24 January 2006, message-ID 11tdkb65dm21r8f@corp.supernews.com.
- [14] "Re: VeriSign digital certificates with Firefox", Nelson Bolyard, posting to netscape.public.mozilla.crypto, 25 January 2006, message-ID ctudnWMSabyYdUreRVn-vQ@mozilla.org.
- [15] "Lightweight OCSP Profile for High Volume Environments", Alex Deacon and Ryan Hurst, draft-ietf-pkix-lightweight-ocsp-profile-03.txt, January 2006.
- [16] Ian Grigg, private communications.
- [17] "How to make figures and presentations that are friendly to color blind people", Masataka Okabe and Kei Ito, http://jfly.iam.u-tokyo.ac.jp/html/-color_blind/.
- [18] "Why Phishing Works", Rachna Dhamija, J.D.Tygar, and Marti Hearst, *Proceedings of the Conference on Human Factors in Computing Systems (CHI'06)*, April 2006, p.581.
- [19] "Bank's own developers a much bigger problem than browsers", 'mhp', 18 July 2004, http://news.netcraft.com/archives/2004/07/18/-banks_own_developers_a_much_bigger_problem_than_browsers.html.
- [20] "How to Make Personalized Web Browsing Simple Secure and Anonymous", Eran Gabber, Phillip Gibbons, Yossi Matias, and Alain Mayer, *Proceedings of Financial Cryptography 1997 (FC'97)*, Springer-Verlag Lecture Notes in Computer Science No.1318, February 1997, p.17.
- [21] "On Secure and Pseudonymous Client-Relationships with Multiple Servers" Eran Gabber, Phillip Gibbons, David Kristol, Yossi Matias, and Alain Mayer, *ACM Transactions on Information and System Security (TISSEC)*, **Vol.2, No.4** (November 1999), p.390.
- [22] "Consistent, Yet Anonymous, Web Access with LPWA", Eran Gabber, Phillip Gibbons, David Kristol, Yossi Matias, and Alain Mayer, *Communications of the ACM*, **Vol.42, No.2** (February 1999), p.42.

- [23] “A Usability Study and Critique of Two Password Managers”, Sonia Chasson, Paul van Oorschot, and Robert Biddle, *Proceedings of the 15th Usenix Security Symposium (Security’06)*, August 2006, p.1.
- [24] “Phishing Social Networking Sites”, “RSnake”, 8 May 2007, <http://ha.ckers.org/blog/20070508/phishing-social-networking-sites/>.
- [25] “A convenient method for securely managing passwords”, J. Alex Halderman, Brent Waters and Ed Felten, *Proceedings of the 14th International World Wide Web Conference (WWW’05)*, May 2005, p.471.
- [26] “Stronger Password Authentication Using Browser Extensions”, Blake Ross, Collin Jackson, Nick Miyake, Dan Boneh, and John Mitchell, *Proceedings of the 14th Usenix Security Symposium (Usenix Security’05)*, August 2005, p.17.
- [27] “Building a Better Filter: How To Create a Safer Internet and Avoid the Litigation Trap”, Erin Egan and Tim Jucovy, *IEEE Security and Privacy*, **Vol.4, No.3** (May/June 2006), p.37.
- [28] OptInRealBig.com, LLC v. IronPort Systems, Inc, US District Court, Northern District of California, Oakland Division, case number 4:04-CV-01687-SBA, 2 September 2004.
- [29] New.net, Inc. v. Lavasoft, U.S. District Court, Central District of California, case number CV 03-3180 GAF.

Security Usability Testing

Designing a usable security interface for an application is inherently difficult (even more so than general user interface design) because of the high level of complexity in the underlying security mechanisms, the nebulous nature of any benefits to the user, and the fact that allowing the user to muddle through (a practice that's sufficient for most interfaces) isn't good enough when they're up against an active and malicious adversary. You therefore need to get the user interface designers in on the process as early as possible, and ensure that the interface drives the security technology and not the other way round. Usability testing is a step that you can't avoid, because even if you choose not to do it explicitly, it'll be done for you implicitly once your application is released. The major difference is that if you perform the testing explicitly, you get to control the testing process and the manner in which results are applied, whereas if you leave it to the market to test, you're liable to get test results like "d00d, your warez SUCKS", or even worse, a CERT advisory.

Usability testing is a two-phase process, pre-implementation testing (trying to figure out what you want to build) and post-implementation testing (verifying that what you eventually built — which given the usual software development process could be quite different from what was planned — is actually the right thing). This section covers both pre- and post-implementation testing of the security usability of an application.

Pre-implementation Testing

Testing at the design stage (before you even begin implementation, for example using a mock-up on paper or a GUI development kit) can be enormously useful in assessing the users' reactions to the interface and as a driver for further design effort [1]. Consider having the designers/developers play the part of the computer when interacting with test users, to allow them to see what their planned interface needs to cope with. Although users aren't professional user interface designers, they are very good at reacting to designs that they don't like, or that won't work in practice. Looking at this from the other side, you could give users the various user interface elements that you'll need to present in your design and ask them to position them on blank page/dialog, and explain how they'd expect each one to work.

One thing to be aware of when you're creating a paper prototype is to make sure that it really is a paper prototype and not a polished-looking mock-up created with a GUI building toolkit or drawing package. If you create a highly-polished design, people will end up nitpicking superficial details and overlook fundamental design issues. For example they might fixate on the colour and style of button placement rather than questioning why the button is there in the first place [2]. If you like doing your UI prototyping in Java, there's a special pluggable napkin look-and-feel for Java that'll give your prototype the required scrawled-on-a-napkin look [3].

A useful design technique to get around the engineering-model lock-in is the "pretend it's magic" trick, in which you try and imagine how the interface would work if it was driven by magic instead of whatever API or programming environment you're working with. Find a user (or users) and ask them how they'd want it to work, without interrupting them every minute or two to tell them that what they're asking for isn't possible and they'll have to start again.

Another useful trick is to sit down and write down each step of the process that you'll be expecting users to perform so that you can see just how painful it (potentially) is in practice. Carrying out this exercise would have quickly helped identify the unworkability of the certificate-enrolment process described in a previous section.

Stereotypical Users

A useful pre-implementation testing technique is to imagine a stereotypical end user (or several types of stereotypical users if this applies) and think about how they'd use the software. What sort of things would they want to do with it? How well would they cope with the security mechanisms? How would they react to security

warnings? The important thing here is that you shouldn't just add a pile of features that you think are cool and then try and figure out how to justify their use by the end user, but that you look at it from the user's point of view and add only those features that they'll actually need and be able to understand. When left to their own devices, developers tend to come up with self-referential designs where the category of "user" doesn't extend any further than people very much like the developer [4].

There's a particular art to the use of stereotypical users, which usability designer Alan Cooper covers in some detail in his book *The Inmates are running the Asylum*. In choosing your user(s), it's important to recreate as much of a real person as you can: Given them names, write a short bio for them, and try and find a representative photo (for example from a magazine or online) that allows you to instantly identify with them. The more specific you can be (at least up to a point), the better.

The reason for this specificity is that a generic cardboard-cut-out user (sometimes referred to as "the elastic user") is far too flexible to provide a very real test of the user interface. Need to choose a key storage location? No problem, the user can handle it. Need to provide an X.500 distinguished name in a web form? Sure, the user can do that. On the other hand 70-year-old Aunty May, whose primary use for her computer is to spam her relatives with emailed jokes, will never go for this. Designing for the elastic user gives you a free hand to do whatever you feel like while still appearing to serve "the user". Creating a user who's as close as possible to a real person (not necessarily an actual person, just something more concrete than a cardboard cut-out) on the other hand lets you directly identify with them and put their reactions to your user interface design into perspective. How would Aunty May handle a request for a public/private key pair file location? By turning off the computer and heading out to do a spot of gardening. Time to rethink your design.

A similar problem occurs with people planning for or deploying security technology. In this case the elastic user becomes a nebulous entity called "the IT department", which takes care of all problems. Take all of the points raised in the previous paragraph and substitute "the IT department can formulate a policy to cover it" for "the user can handle it" and you can see where this type of thinking leads. Only a few large corporations can afford the luxury of having an IT department define policies for every security eventuality, and even then truly effective policies usually only appear after a crisis has occurred. For everyone else, they *are* the IT department, leading to farcical situations such as Aunty May sitting at her home PC in front of a dialog box telling her to contact her system administrator for help.

Note though that you should never employ the technique of stereotypical users as a substitute for studying real users if such access is available. An amazing amount of time is wasted at the design stage of many projects as various contributors argue over what users might in theory do if they were to use the system, rather than simply going to the users and seeing what they actually do.

All too frequently, user interfaces go against the user's natural expectations of how something is supposed to work. For example a survey of a range of users from different backgrounds on how they expected public keys and certificates to be managed produced results that were very, very different from how X.509 says it should be done, suggesting at least one reason for X.509's failure to achieve any real penetration [5].

A final useful function provided by the stereotypical user is that they act as a sanity check for edge cases. The unerring ability of geeks to home in on small problems (and then declare the entire approach un-workable because of the corner case they've thought up) has already been covered. Geeks have major problems distinguishing possibility from probability. To a geek (and especially a security geek) a probability of one in a million is true. To a cryptographer, a probability of 1 in 2^{56} or even 1 in 2^{80} (that's 1 in 1.2 million million million million, a one followed by twenty-four zeroes) is true. To anyone else (except perhaps Terry Pratchett fans), a one-in-a-million chance is false - there's a *possibility* of it being true, but the actual *probability* is miniscule to the point of irrelevance. Personas provide a sanity check for such edge cases. Yes, this is a special case, but would Aunty May ever want to do that?

Input from Users

Asking users how they think that something should work is an extremely useful design technique. Consider the question of storing users' private keys. Should they be stored in one big file on disk? Multiple files? In the registry (if the program is running under Windows)? On a USB token? In their home directory? In a hidden directory underneath their home directory? What happens if users click on one of these files? What if they want to move a particular key to another machine? How about all of their keys? What happens when they stop using the machine or account where the keys are stored? How are the keys protected? How are they backed up? Should they even be backed up?

All of these questions can be debated infinitely, but there's a far simpler (and more effective) way to resolve things. Go and ask the users how they would expect them to be done. Many users won't know, or won't care, but eventually you'll see some sort of common model for key use and handling start to appear. This model will be the one that most clearly matches the user's natural expectations of how things are supposed to work, and therefore the one that they'll find the easiest to use. The problems that can occur when an application doesn't meet users' expectations for key storage was illustrated in one PKI-based tax filing scheme where users weren't able to figure out how key storage worked and solved the problem by requesting a new certificate at each interim filing period (two months). This resulted in an enormous and never-ending certificate churn that completely overloaded the ability of the certificate-issuing process to handle it, and lead to unmanageable large CRLs.

Testing by asking users for input has been used for some years by some companies when developing new user interface features. For example in the early 1980s whenever a new interface feature was implemented for the Apple Lisa, Apple developer Larry Tesler would collar an Apple employee to try it out. If they couldn't figure it out, the feature was redesigned or removed.

Another advantage of asking users what they want is that they frequently come up with issues that the developers haven't even dreamed about. If you do this though, make sure that you occasionally refresh your user pool, because as users spend more and more time with your interface they become less and less representative of the typical user, and therefore less able to pick up potential usability problems.

When you ask users for input, it's important to ask the *right* users. Another problem that the PKI tax filing scheme mentioned above ran into was the difference between the claimed and the actual technology level of the users. When various managers were surveyed during the requirements process, they all replied that their staff had the latest PCs on their desks and were technology-literate. In other words the managers were describing themselves. In actual fact the people doing the tax filing were, as one observer put it, "little old ladies sitting in front of dusty PCs with post-it notes telling them what to do stuck to the monitor". The post-it notes contained paint-by-numbers instructions for the tax filing process, and as soon as one of the post-it's didn't match what was on the screen, the users called the help desk. The result was that most of the electronic filing was being done by proxy by the helpdesk staff, and the system haemorrhaged money at an incredible rate until it was finally upgraded from electronic back to paper-based filing.

The importance of going directly to the end users (rather than relying on testimony from their superiors) can't be over-emphasised. No manager will ever admit that their employees aren't capable of doing something (it would make the manager look bad if they did), so the response to "can your people handle X" is invariably "yes", whether they really can or not. I once went into the paging centre at a large hospital, where messages to and from doctors are dispatched to and from other doctors to talk to the staff about their requirements. After a few minutes there I was somewhat disturbed to discover that this was the first time that anyone had ever asked the users what they actually needed the software to do for them. In the entire lifetime of the hospital, no-one had ever asked the users what they needed! Needless to say, using the software was a considerable struggle (it was an extreme example of task-directed

design), and even a preliminary set of minor changes to the interface improved the users' satisfaction considerably.

Post-implementation Testing

Once you've finished your application, take a few non-technical people, sit them in a room with a copy of the software running, and see how they handle it. Which parts take them the longest? At what points do they have to refer to the manual, or even ask for help? Did they manage to get the task done in a secure manner, meaning that *their* expectations of security (not just yours) were met? Can a section that caused them problems be redesigned or even eliminated by using a safe default setting? Real testing before deployment (rather than shipping a version provisionally tagged as a beta release and waiting for user complaints) is an important part of the security usability evaluation process.

Logging of users' actions during this process can help show up problem areas, either because users take a long time to do something or because their actions generate many error messages. Logging also has the major advantage that (except for privacy concerns) it's totally non-invasive, so that users can ignore the logging and just get on with what they're doing. Microsoft used logging extensively in designing the new interface for Office 2007/Office 12, analysing 1.3 billion Office 2003 sessions in order to determine what users were and weren't using [6].

There's another useful litmus test that you can use for your post-implementation testing to find potential security weaknesses. Imagine that your application has been deployed for awhile and there's been a report of a catastrophic security failure in it. Yes, we know that your application is perfect in every way, but somehow some part of it has failed and the only error information that you have to work with is the report that it failed. Where do you think the failure was? How would you fix it?

This type of analysis is an interesting psychological technique called a premortem strategy [7]. The US Navy gave it the name "crystal-ball technique" in its review of decision-making under stress that occurred after the erroneous shootdown of a civilian airliner by the USS Vincennes [8]. In the Navy version, people are told to assume that they have a crystal ball that's told them that their favoured hypothesis is incorrect, so that they have to come up with an alternative explanation for an event.

No matter what you call it, what premortem analysis does is compensate for the overconfidence in a work that anyone who's intimately involved in its creation develops over extended exposure to it. If you ask a designer or programmer to review their application, their review will be rather half-hearted, since they want to believe that what they've created is pretty good. The premortem strategy helps them break their emotional attachment to the project's success and objectively identify likely points of failure. Real-world testing has shown that it takes less than ten minutes for failures and their likely causes to be discovered [9].

User Testing

User interface design is usually a highly iterative process, so that the standard { design, implement, test } cycle described above probably won't be enough to shake out all potential problems, particularly in the case of something as complex and hard to predict as security user interface design. Instead of a single cycle, you may need to use multiple cycles of user testing, starting with a relatively generic design (sometimes known as low-fi prototyping) and then refining it based on user feedback and experience.

This testing process needn't be complex or expensive. Usability expert Jakob Nielsen has shown that once you go beyond about five users tested, you're not getting much more information in terms of usability results [10]. This phenomenon occurs because as you add more and more users, there's increasing overlap in what they do, so that you learn less and less from each new user that you add. So if you have (say) 20 test users, it's better to use them in four different sets of tests on different versions or iterations of the interface than to commit all 20 to a single test. A variation of this situation occurs when a single group contains highly distinct subgroups of users, such

as one where half the users are technical and the other half are non-technical. In this case you should treat each subgroup as a separate unit for 5-user test purposes, since they're likely to produce very different test results.

A useful tool to employ during this iterative design process is to encourage users to think out loud as they're using the software. This verbalisation of users' thoughts helps track not just *what* users are doing but *why* they're doing it, allowing you to locate potential stumbling blocks and areas that cause confusion. Make sure though that you actually analyse a user's comments about potential problems. If a user misses an item in a dialog or misreads a message, they may end up in trouble at some point further down the road, and come up with complex rationalisations about why the application is broken at the point where they realise that they're in trouble, rather than at the point where they originally made the error.

Note also that the very act of verbalising (and having to provide an explanation for) their actions can make a user think much more about what they're doing, and as a result change their behaviour. Tests with users have shown that they're much better at performing a user interface task when they're required to think out loud about what they're doing.

To get around this, you can allow the user to perform less thinking out loud, and instead prompt them at various points for thoughts on what they're doing. Asking questions like "What do you expect will happen if you do this?" or "Is that what you expected would happen?" are excellent ways of turning up flawed assumptions in your design.

A variation of thinking out loud is constructive interaction, in which two users use a system together and comment on each other's actions (imagine your parents sitting in front of their PC trying to figure out how to send a photo attachment via their Hotmail account). This type of feedback-gathering is somewhat more natural than thinking out loud, so there's less chance of experimental bias being introduced.

Another trick that you can use during user interface testing is to insert copier's traps into the interface to see if users really are paying attention. Copier's traps are little anomalies inserted into maps by mapmakers that allow them to detect if a competitor has copied one of their maps, since a map prepared from original mapping data won't contain the fictitious feature shown in the trap.

You can use the same technique in your user interface to see if users really have understood the task that they're performing or whether they're just muddling through. In a standard application, muddling through a task like removing red-eye from a photo is fine as long as the end result looks OK, but in a security context with an active and malicious adversary it can be downright dangerous even if the result does appear to be OK. Adding a few copier's traps during the testing phase will tell you whether the interface really is working as intended, or whether the user has simply managed to bluff their way through.

Usability Testing Examples

This section presents a number of case studies of security usability problems that were turned up by user testing. Unfortunately almost all of the testing was reactive rather than proactive and has resulted in few changes to products either because it's too late to fix things now or because the affected organisations aren't interested in making changes. As well as providing for interesting usability case studies, these examples could be seen as a strong argument for pre-release testing.

Encrypted Email

An example of the conflict between user expectations and security design was turned up when security usability studies showed that email users typically weren't aware that (a) messages can be modified as they move across the Internet, (b) encrypting a message doesn't provide any protection against such modification, and (c) signing a message does protect it. The users had assumed that encrypting a message provided integrity protection but signing it simply appended the equivalent of a pen-and-paper

signature to the end of it [11]. Real-world testing and user feedback is required to identify these issues so that they can be addressed, for example by explaining signing as protecting the message from tampering rather than the easily-misunderstood “signing”. Similarly, the fact that encryption doesn’t provide integrity protection can be addressed either at the user interface level by warning the user that the encrypted message isn’t protected from modification (trying to “fix” the user), or at the technical level by adding a MDC (modification detection code) inside the encryption layer or a MAC (message authentication code) outside it (actually fixing the problem). Of these two, the latter is the better option since it “fixes” the encryption to do what users expect without additionally burdening the user. This is the approach taken by OpenPGP, which added a SHA-1 hash to the encrypted data (S/MIME doesn’t appear to be interested in fixing this). Modifying the application to do what the user wants is always preferable to trying to modify the user to do what the application wants.

Browser Cookies

Another example of a problem that would have been turned up by post-implementation testing occurs with the handling of cookies in browsers. This has slowly (and painfully) improved over the years from no user control over what a remote web site could do to rather poor control over what it could do. The reason for this was that cookies are a mechanism designed purely for the convenience of the remote site to make the stateless HTTP protocol (slightly) stateful. No-one ever considered the consequences for users, and as a result it’s now extremely hard to fix the problem and make the cookie mechanism safe [12]. For example once a browser connects to a remote site, it automatically sends any cookies it has for the site to the remote server instead of requiring that the server explicitly request them. While more recent browsers allow users to prevent some types of cookies from being stored, it’s not the storage that’s the problem but their usage by the remote system, and the user has no control over that since changing current browsers’ behaviour would require the redesign of vast numbers of web sites. Similarly, while in recent browsers users have been given the ability to selectively enable storage of cookies from particular sites, clearing them afterwards is still an all-or-nothing affair. There’s no way to say “clear all cookies except for the ones from the sites I’ve chosen to keep”.

With more testing of the user side of the cookie mechanism, it should have been obvious that having the user’s software volunteering information to a remote system in this manner was a poor design decision. Now that usability researchers have looked at it and pointed out the problems, it’s unfortunately too late to change the design.

(Note that fixing cookies wouldn’t have solved the overall problem of site control over data stored on the user’s machine, because there are cookie-equivalent mechanisms that sites can use in place of cookies, and these can’t be made safe (or at least safer) in the way that cookies can without significantly curtailing browser operations. For example the browser cache operates in somewhat the same way as cookies, allowing site-controlled data to be temporarily stored on the user’s machine. By setting the Last-Modified field in the header (which is required in order for caching to work) and reading it back when the browser sends its If-Modified-Since in future requests, a server can achieve the same effect as storing a cookie on the client’s machine. There are other tricks available to servers if the client tries to sidestep this cache-cookie mechanism [13]. So even with a better user interface and a fixed design that makes the cookie client-controlled, malicious servers will always have a cookie-like mechanism available to them).

Key Storage

Post-implementation testing can often turn up highly surprising results arising from issues that would never have occurred to implementers. A representative example from outside the security world occurred in the evolution of what we’re now familiar with as the ‘OK’ button, which in its early days was labelled quite differently since it was felt that ‘OK’ was a bit too colloquial for serious computer use. In 1981 when Apple was performing early user testing on the nascent Macintosh user interface, the

button was labelled 'Do It'. However, the testing revealed that 'Do It' was a bit too close visually to 'Dolt', and some users were becoming upset that a computer touted for its user-friendliness was calling them dolts [14]. The designers, who knew that the text said Do It because they were the ones who had written it, would never have been able to see this problem because they knew a priori what the text was meant to say. The alternative interpretation was only revealed through testing with users uncontaminated by involvement in the Macintosh design effort.

Getting back to the security world, the developers of the Tor anonymity system found that Tor users were mailing out their private keys to other Tor users, despite the fact that they were supposed to know not to do this. Changing the key filename to include a `secret_` prefix at the front solved the problem by making it explicit to users that this was something that shouldn't be shared [15]. PGP solves the problem in a similar manner by only allowing the public key components to be exported from a PGP keyring, even if the user specifies that the PGP private keyring be used as the source for the export.

Conversely, Windows/PKCS #12 takes exactly the opposite approach, blurring any distinction between the two in the form of a single "digital identity" or PKCS #12/PFX file, so that users are unaware that they're handing over their private keys as part of their digital identity (one paper likens this practice to "pouring weed killer into a fruit juice bottle and storing it on an easily accessible shelf in the kitchen cupboard") [16]. The term "digital identity" is in fact so meaningless to users that in one usability test they weren't able to usefully explain what it was *after they'd used it for more than half an hour* [17]. Think about this yourself for a second: Excluding the stock response of "It's an X.509 certificate", how would you define the term "digital identity"?

Another issue with private keys held in crypto tokens like USB keys or smart cards involves how users perceive these devices. In theory, USB tokens are superior to smart cards in every way: They're a more convenient form factor, less physically fragile, easier to secure (because they're not limited to the very constrained smart card form factor), more flexible through the ability to add additional circuitry, don't require a separate reader, and so on. Smart cards only have one single advantage over user USB tokens: the USB tokens are (conceptually) very close to standard keys, which get shared among members of the family, lent to relatives or friends who may be visiting, or left with the neighbours so that they can feed the cat and water the plants when the owners are away.

Smart cards, when the correct measures are used, don't have this problem. If you take a smart card and personalise it for the user with a large photo of the owner, their name and date of birth, a digitised copy of their signature, and various extras like a fancy hologram and other flashy bits, they'll be strongly inclined to guard it closely and highly reluctant to lend it out to others. The (somewhat unfortunate) measure of making the card an identity-theft target ensures that it'll get looked after better than an anonymous USB token.

Banking Passwords

The threat model for passwords on the Internet is quite different from the historic threat model, which dates back to the 1960s with users logging onto centralised mainframes via dedicated terminals. In this mainframe environment, the attacker keeps trying passwords against a user account until they guess the right one. What this means is that the user name stays constant and the password varies. The defence against this type of attack is the traditional "three strikes and you're out" one in which three incorrect password attempts set off alarms, cause a delay of several minutes before you can try again, or in the most paranoid cases lock the account, a marvellous self-inflicted denial-of-service attack.

That was the threat model (and corresponding defence) from forty years ago. Today, the threat is quite different. In response to the three-strikes-and-you're-out defence, attackers are keeping the password constant and varying the user name instead of the other way round. With a large enough number of users, they'll eventually find a user

that's using the password that they're trying against each account, and since they only try one password per account (or more generally a value less than the lockout threshold), they never trigger the defence mechanisms. This is a 21st-century Internet attack applied against an anachronistic threat model that hasn't really existed for some decades.

Consider the following example of this attack, based on research carried out on the Norwegian banking system in 2003-4 [18]. The Norwegian banks used a standard four-digit PIN and locked the account after the traditional three attempts. Using the fixed-PIN/varying-userID approach, an attacker would be able to access one account out of every 220,000 tried (a botnet would be ideal for this kind of attack). On the next scan with a different PIN, they'd get another account. Although this sounds like an awfully low yield, the only human effort required is pointing a botnet at the target and then sitting back and waiting for the results to start rolling in. The banking password authentication mechanisms were never designed to withstand this type of attack, since they used as the basis for their defence the 1960s threat model that works just fine when the user is at an ATM.

There are many variants of this attack. Some European banks use dynamic PIN calculators, which generate a new time-based or pseudorandom-sequence based PIN for each logon. In order to accommodate clock drift or a value in the sequence being lost (for example due to a browser crash or network error), the servers allow a window of a few values in either direction of the currently expected value. As with the static-password model, this works really well against an attacker that tries to guess the PIN for a single account, but really badly against an attacker that tries a fixed PIN across all accounts, because as soon as their botnet has hit enough accounts they'll come up a winner.

For all of these attacks (and further variations not covered here), a basic level of post-release analysis would have uncovered the flaw in the threat model. Unfortunately the testing was only performed some years later by academic researchers, and the affected organisations mostly ignored their findings [18].

Password Managers

The previous chapter looked at the use of strengthened password mechanisms to protect users' passwords, and mentioned that facilities of this type are already available in some cases, typically as plugins for the Firefox web browser. How do these plugins stand up in practice? A usability study of two popular password managers, PwdHash and Password-Multiplier, found that they fall far short of their authors' expectations due to a variety of user interface problems [19].

The biggest problem with these browser plugins is that they are exactly that, browser plugins. The lack of integration into the browser created almost all of the usability problems that users experienced. For example if the plugin wasn't installed or was bypassed by a malicious web page using Javascript or a similar technique, the user would end up entering their master password on a remote login page instead of having the plugin provide a site-specific random password.

This reiterates an important point that's already been made elsewhere: In order to be effective, a security measure has to be a native part of the underlying application. It has to be present and active at all times. It can't be an optional add-on component that may or may not be currently active, or for which users have to expend conscious effort to notice its presence, because they simply won't notice its absence (see the earlier discussion on the psychological aspects of the security user interface for more on this problem).

A second problem with the lack of direct integration is that the add-on nature of the browser plugins lead to complex and awkward interaction mechanisms because of the lack of direct access to browser-internal mechanisms. A direct consequence of this awkwardness was that only one single task of the five that users were asked to complete in the study had a success rate over 50%, with failure rates being as high as 84%. Alarming, one of the failure modes that was revealed was that users tried

entering every password they could think of when they couldn't access the site using the plugin.

For the plugin tested in the usability study, users were required to use special attention-key sequences like '@@' or Alt-P or F2 to activate the security mechanisms, and these were only effective if the cursor was already present in the password text fields. Users either forgot to use the attention sequence, or got them wrong, or used them at the wrong time. They therefore found it very hard to tell whether they'd successfully activated and applied the plugin security mechanisms, and several said that if they hadn't been participating in a study they'd have long since signed up for a new account with a standard password rather than struggle further with the password-manager plugins.

These problems came about entirely because of the need to implement the security features as a plugin. If they'd been built directly into the browser, none of this would have occurred.

Another interesting feature that was turned up by the user testing was that people were profoundly uneasy about the fact that they no longer knew the passwords that they were using, leading to complaints like "I wish it would show me my password when it first generates it. I won't lose it or share it!" [19]. This loss of control negatively affected users' perceptions of the password manager. One way of mitigating this problem, already provided by the rudimentary password-saving features built into existing browsers, is to display the password when the user requests it. This helps fight the users' perception that they've lost control of their passwords when they let the password manager handle them.

File Sharing

A similar problem to the Tor one was turned up by post-implementation testing of the Kazaa file-sharing application ("post-implementation testing" in this case means that after the software had been in use for awhile, some researchers went out and had a look at how it was being used) [20]. They found that Kazaa exhibited a considerable number of user interface problems, with only two of twelve users tested being able to determine which files they were sharing with the rest of the world. Both design factors and the Kazaa developers' lack of knowledge of user behaviour through pre- or post-implementation testing contributed to these problems. For example Kazaa manages shared files through two independent locations, via the "Shared Folders" dialog box and the "My Media" downloads folder. Items that were shared through one weren't reflected in the other, so if a user chose to download files to their Windows C: drive, they inadvertently shared the entire drive with other Kazaa users (!!!) without the "Shared Folders" dialog indicating this. The number of users caught out by this was indicated by over four hundred sample searches carried out in a period of twelve hours, with 61% of the searches returning hits for Kazaa users' Outlook Express mail files, a representative file that would never (knowingly) be shared with the rest of the world. Another study, which looked only for banking files, found large numbers of files containing sensitive banking information being inadvertently shared by bank employees [21]. Kazaa's poor default settings have even lead one lawyer to comment that it offers "no reasonable expectation of privacy" [22].

An aspect of user behaviour that was unanticipated by the Kazaa developers was the fact that users were in general unaware that sharing a folder (directory) would share the contents of all of the subdirectories beneath it, and were also unaware that sharing a folder shared all of the files in it rather than just a particular file type such as music files. Part of this problem was again due to the user interface design, where clicking on a parent folder such as "My Documents" (which is automatically recommended for sharing by Kazaa when it's set up) gave no indication that all files and subfolders beneath it would also be shared.

As with the mismatch of user expectations over message encryption that were covered earlier, there are two ways to address this problem. The first is to attempt to "fix" the user by warning them that they're sharing all files and subdirectories, an action that the previous sections have shown is likely to have little effect on security

(users will satisfy their way past it — they want to trade files, not read warnings). A much better approach uses activity-based planning to avoid ever putting the user in a situation where such a warning is necessary. With this style of interface, the user is given the option to share music in the current folder, share pictures in the current folder, share movies in the current folder (let's face it, Kazaa isn't used to exchange knitting patterns), or go to an advanced sharing mode interface. This advanced/expert mode interface allows the specification of additional file types to share and an option to share such file types in subdirectories, disabled by default.


Some P2P applications in fact do the exact opposite of this, searching users' hard drives for any folders containing music or videos and then sharing *the entire folder* that contains the music file(s) [21]. This is compounded by the fact that many users don't really understand the concept of folders and tend to save documents wherever the 'Save' dialog happens to be pointing to when it pops up (one sysadmin describes the resulting collection of data as "not so much filed as sprayed at random across the filesystem"). As a result, the letter to the bank is stored next to the holiday photos, the Quicken account data, and the video of the dancing bunnies, all shared with anyone else with an Internet connection. Compounding the problem, the set-and-forget nature of P2P applications and the lack of interaction with the user once they've been started leaves users with no indication that saving or copying any new files into shared folders is publishing that information for the entire Internet to see.

An additional safety feature would be to provide the user with a capsule summary of the types and number of files being shared ("21 video files, 142 sound files, 92 images, 26 documents, 4 spreadsheets, 17 programs, 218 other files") as an additional warning about what it is they're doing ("Why does it say that I'm sharing documents and spreadsheets when I thought I was only sharing sounds and images?"). Strangely enough, the My Media folder (ostensibly meant for incoming files) provides exactly this summary information, while the Shared Folders interface doesn't, merely showing a directory tree view. This simple change to the user interface now makes the application behave in the way that the user expects it to, with no loss of functionality but a significant gain in security. Even a quick change to the current user interface, having it auto-expand the first one or two levels of directories in the tree view to show that all of the sub-folders are selected, would at least go some way towards fixing the interface's security problems.

Site Images

In 2005 the US Federal Financial Institutions Examination Council (FFIEC) issued guidance requiring that US financial institutions use two-factor authentication (strictly speaking they said that single-factor authentication was inadequate and required that "financial institutions offering Internet-based products and services to their customers should use effective methods to authenticate the identity of customers") [23]. The poor security practices of US financial institutions have already been covered in previous chapters; in this case they redefined "two-factor authentication" so that it no longer required the use of a security token like a SecurID or a challenge/response calculator of the type used by European banks (which would have cost money to deploy), but merely required them to display a personalised image on the user's logon page [24]. In other words their definition of "two-factor authentication" was "twice as much one-factor authentication".

They then compounded the error by training users to ignore the standard HTTPS indicators in favour of the site images. Figure 43 and Figure 44 provide two examples of this problem.



Online Banking

Confirm that your SiteKey is correct

If you recognize your SiteKey, you'll know for sure that you are at the valid Bank of America site. Confirming your SiteKey is also how you'll know that it's safe to enter your Passcode and click the **Sign In** button.

An asterisk (*) indicates a required field.

Your SiteKey:



If you don't recognize your personalized SiteKey, don't enter your Passcode.

* **Passcode:**

(4 - 20 Characters, case sensitive)


[Sign In](#)

Figure 43: Training users to ignore HTTP indicators

When security researchers looked at the effectiveness of these security indicators, the results were alarming, but predictable: Users were ignoring the existing HTTPS indicators (in the study not one user was stopped by the absence of HTTPS indicators), but also not paying much attention to the absence of the site image either. Simply replacing the image with a message telling users that “*bank-name* is currently upgrading our award-winning *site-image brand-name* feature. Please contact customer service if your *site-image brand-name* does not reappear within the next 24 hours” was enough to convince 92% of the participants in the study that it was safe to use the site [25]. Although it wouldn’t have been too hard to simply copy the site image from the genuine site (it took about a minute to defeat the purported additional challenge-question security measures to obtain the sample image shown in Figure 43), an attacker doesn’t even have to go to this minimal level of effort to defeat it — a maintenance message is all that’s required, and thanks to the banks’ conditioning of users the SSL indicators are bypassed to boot.

Personal Investors

Search


Open an account | Log on | Forms | Contact us | Site help

Home
My Portfolio
Research Funds & Stocks
Account Types & Services
Planning &

Security Center » Enhanced Logon FAQs

How do I know I'm on the authentic Vanguard website?

When you enter your Vanguard user name from a recognized computer, we'll show you your security image and caption. At that point, you'll know you're on the authentic Vanguard.com website and can safely enter your Vanguard password.

Figure 44: More user insecurity training

The effectiveness so trivial a measure as removing the site images through a bogus “under construction” message is a follow-on effect of the “all the ads all the time” nature of today’s web sites. Just as users expect ASP and Javascript problems, transient network outages, broken links and 404 errors, and similar issues whenever they go online, they’re also quite used to constantly-mutating web sites where almost

anything can change between visits. As with the SSL indicators mentioned in an earlier section, trying to detect security problems using a mechanism with a close to 100% false positive rate isn't notably useful.

Other attacks on site images include a standard man-in-the-middle attack (which is quite simple to perform, despite claims from the marketing manager of the service that it's impossible) [26], or just displaying a random image from the selection provided by the bank. Although the effectiveness of the latter approach hasn't been experimentally evaluated, the results of other studies on users' attention to security indicators of this type suggests that a significant number of users won't notice that anything is amiss.

In any case the redefinition of "two-factor authentication" to mean "twice as much one-factor authentication" presented the merest speed-bump to malware authors, who bypassed it with little effort. For example the Gozi Trojan, among its many other capabilities, has a "grabs" module that hooks into the browser's Javascript engine to obtain any extra credentials communicated via AJAX mechanisms rather than a standard password-entry dialog [27]. There's no indication from the malware community that the twice-as-much-one-factor approach is presenting any difficulty to attackers.

Signed Email

Today virtually all use of signed messaging occurs in automated protocols and processes like EDI buried deep down in the IT infrastructure. The vision that flourished during the crypto wars of the 1990s that everyone would eventually be using signed and/or encrypted email has pretty much evaporated. So why is no-one signing their messages?

Part of the blame can be laid at the feet of the un-usability of the PKI or PKI-like mechanisms that are required to support signing, but another part of the problem is the fact that while geeks will do something with a computer just because it's geeky, the rest of the world needs a reason to do things with their computers. Why would the average user care about signed email? If it's from someone that they know then they'll verify the message's authenticity based on the message contents, so-called semantic integrity, and not a digital signature [28]. On the other hand if it's from someone that they don't know then it doesn't matter whether the message is signed or not. In neither case is the large amount of effort required in order to work with digital signatures justified in the eyes of the typical user.

This doesn't apply only to everyday users. When the S/MIME standards group debated whether they should switch to using S/MIME signed email for their discussions, they came to the same conclusion. In other words one of the groups that sets the standards for digitally signed messages decided that there wasn't much point to actually using them. Although it can be argued (endlessly) that everyone should be using mechanisms like signed email to prevent things like phishing attacks, a user base that has problems with something as basic as a padlock icon will never be able to cope with the massive complexity that comes with digitally signed email. So despite the best efforts of the protocol designers and programmers and the ensuing result that the majority of the world's desktops have digital-signature-enabled email clients built into them, the market has decided that, by and large, digitally-signed email just isn't worth the effort. As with several of the other examples presented here, real-world user testing would have saved considerable misspent effort (both at the IT and the government/legislative level, consider all of the moribund digital signature legislation that half the world's governments were busy passing in the late 1990s) and helped focus efforts elsewhere.

(Another concern, which because of the lack of digital signature usage comes up mostly among privacy advocates, is the problem of incrimination. There is already concern among some users about the size of the digital footprint or data shadow that they create in their everyday use of computers and the Internet. Digitally signing everything, the equivalent of creating a notarised document under some digital signature regimes, doesn't help allay these concerns).

Post-delivery Reviews

A final stage of testing is the post-delivery review, sometimes referred to as a retrospective. Most advocates of this process suggest that 3-12 months after release is the best time to carry out this type of review, this being the point at which users have become sufficiently familiar with the software to locate problem areas, and at which point the software has had sufficient exposure to the real world to reveal any flaws in the design or its underlying assumptions.

This final stage of the design process is extremely important when deploying a security system. The reason why the Walker spy ring was able to compromise the NSA-designed security of the US Navy so effectively was that the NSA and Navy in combination had ended up creating an overall system that was (as the post-mortem report mentioned earlier puts it) “inherently insecure and unusable”, despite the fact that it had been built on (theoretically) secure components [29]. The report goes on to say that “time and again, individuals made decisions based on assumptions that proved to be woefully incorrect. In many cases, these assumptions were based on nothing more than wishful thinking, or on the fact that it would be very convenient if certain things were true [...] Just as good design involves finding out how the encryptor behaves as the battery loses its charge or the device gets splashed with water, so also good system design should take into account what happens when the operators do not behave as they ought to — whether through malice, carelessness, or simple inability to carry out the requirements with the resources available. The latter two cases can be minimized or even eliminated through better design: that is, the designer must make it as easy as possible to do the right thing and as hard as possible to do the wrong thing. This needs to be an iterative process, based on close observation of what ordinary sailors actually do during fleet deployments, and incorporating improvements and innovations as they become available”.

The rest of the report constitutes a fascinating insight into just how badly a theoretically secure system that ignores real-world considerations can fail in practice, with almost every aspect of the system compromised in one way or the other once it came into contact with the real world. This shows just how important both studying real users (during the pre-implementation phase) and observing how it’s used once it’s deployed (during the post-implementation phase) can be in ensuring that the system actually has the properties that it’s supposed to have.

Post-delivery reviews are important for shaking out emergent properties unanticipated by the designers that even post-implementation testing with users can’t locate. For example when the folks who wrote RFC 1738 provided for URLs of the form `user@hostname`, they never considered that a malicious party could use this to construct URLs like `http://www.bankofamerica.com@1234567/`, which points to a server whose numeric IP address is 1234567 while appearing to users to be a legitimate bank server’s address. Testing in a hostile environment (the real world) provides additional feedback on secure user interface design. Although it’s unlikely that attackers will cooperate in performing this type of testing for you, over the years a large body of knowledge has been established that you can use to ensure that your application doesn’t suffer from the same weaknesses. Books on secure programming like *Building Secure Software* by John Viega and Gary McGraw and *Writing Secure Software* by Michael Howard and David LeBlanc contain in-depth discussions of “features” to avoid when you create an application that needs to process or display security-relevant information.

References

- [1] “Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces”, Carolyn Snyder, Morgan Kaufmann, 2003.
- [2] “Matching design sketches to the desired level of design feedback”, Jan Mikovsky, 26 October 2006, http://mikovsky.blogs.com/-flowstate/2006/10/using_crude_ske.html.
- [3] “Napkin Look and Feel”, <http://napkinlaf.sourceforge.net/>.

- [4] "About Face 2.0: The Essentials of Interaction Design", Alan Cooper and Robert Reimann, John Wiley and Sons, 2003.
- [5] "PKI Technology Survey and Blueprint", Peter Gutmann, *Proceedings of the 2006 New Security Paradigms Workshop (NSPW'06)*, October 2006.
- [6] "Inside Deep Thought (Why the UI, Part 6)", Jensen Harris, 31 October 2005, <http://blogs.msdn.com/jensenh/archive/2005/10/31/-487247.aspx>.
- [7] "Critical Thinking Skills in Tactical Decision Making: A Model and A Training Method", Marvin Cohen, Jared Freeman, and Bryan Thompson, in "Making Decisions Under Stress: Implications for Individual and Team Training", American Psychological Association (APA), 1998, p.155.
- [8] "Critical Thinking Skills in Tactical Decision Making: A Model and A Training Strategy", Marvin Cohen, Jared Freeman, and Bryan Thompson, "Making Decisions Under Stress: Implications for Individual and Team Training", American Psychological Association (APA), 1998, p.155.
- [9] "Sources of Power: How People Make Decisions", Gary Klein, MIT Press, 1998.
- [10] "Why You Only Need to Test With 5 Users", Jakob Nielsen, <http://www.useit.com/alertbox/20000319.html>, March 2000.
- [11] "Johnny 2: A User Test of Key Continuity Management with S/MIME and Outlook Express", Simson Garfinkel and Robert Miller, *Proceedings of the 2005 Symposium on Usable Privacy and Security (SOUPS'05)*, July 2005, p.13.
- [12] "Cookies and Web Browser Design: Toward Realizing Informed Consent Online", Lynette Millett, Batya Friedman, and Edward Felten, *Proceedings of the 2001 Conference on Human Factors in Computing Systems (CHI'01)*, April 2001, p.46.
- [13] "Silence on the Wire", Michal Zalewski, No Starch Press, 2004.
- [14] "Revolution in the Valley", Andy Hertzfeld, O'Reilly Media Inc, 2005.
- [15] Nick Mathewson, private communications.
- [16] "Lessons Learned in Implementing and Deploying Crypto Software", Peter Gutmann, *Proceedings of the 11th Usenix Security Symposium*, August 2002, p.315.
- [17] "Design Principles and Patterns for Computer Systems That Are Simultaneously Secure and Usable", Simson Garfinkel, PhD thesis, Massachusetts Institute of Technology, May 2005.
- [18] "Case Study: Online Banking Security", Kjell Hole, Vebjørn Moen, and Thomas Tjøstheim, *IEEE Security and Privacy*, **Vol.4, No.2** (March/April 2006), p.14.
- [19] "A Usability Study and Critique of Two Password Managers", Sonia Chasson, Paul van Oorschot, and Robert Biddle, *Proceedings of the 15th Usenix Security Symposium (Security'06)*, August 2006, p.1.
- [20] "Usability and privacy: a study of Kazaa P2P file-sharing", Nathaniel Good and Aaron Krekelberg, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 2003, p.137.
- [21] "Inadvertent Disclosure: Information Leaks in the Extended Enterprise", M.Eric Johnson and Scott Dynes, *Proceedings of the 6th Workshop on the Economics of Information Security (WEIS'07)*, June 2007.
- [22] "RIAA 'extortion': why the only RICO they fear is Suave", Eric Bangeman, 6 May 2007, <http://arstechnica.com/news.ars/post/20070506-riaa-extortion-why-the-only-rico-they-fear-is-suave.html>.
- [23] "Authentication in an Internet Banking Environment", Federal Financial Institutions Examination Council, October 2005, http://www.ffiec.gov/pdf/authentication_guidance.pdf.
- [24] "Fraud Vulnerabilities in SiteKey Security at Bank of America", Jim Youll, Challenge/Response LLC, 18 July 2006, <http://cr-labs.com/publications/SiteKey-20060718.pdf>.
- [25] "The Emperor's New Security Indicators", Stuart Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer, *IEEE Symposium on Security and Privacy*, May 2007, to appear.

- [26] "A Deceit-Augmented Man In The Middle Attack Against Bank of America's SiteKey® Service", Christopher Soghoian and Markus Jakobsson, 10 April 2007, <http://paranoia.dubfire.net/2007/04/deceit-augmented-man-in-middle-attack.html>.
- [27] "Gozi Trojan", Don Jackson, 20 March 2007, <http://www.secureworks.com/research/threats/gozi>.
- [28] "A Case (Study) For Usability in Secure Email Communication", Apu Kapadia, *IEEE Security and Privacy*, **Vol.5, No.2** (March/April 2007), p.80.
- [29] "An Analysis of the System Security Weaknesses of the US Navy Fleet Broadcasting System, 1967-1974, as exploited by CWO John Walker", Laura Heath, Master of Military Art and Science thesis, US Army Command and General Staff College, Ft.Leavenworth, Kansas, 2005.

Data Enveloping

Encryption envelopes are the easiest way to use cryptlib. An envelope is a container object whose behaviour is modified by the data and resources that you push into it. To use an envelope, you add to it other container and action objects and resources such as passwords that control the actions performed by the envelope, and then push in and pop out data that's processed according to the resources that you've pushed in. cryptlib takes care of the rest. For example to encrypt the message "This is a secret" with the password "Secret password" you would do the following:

```
create the envelope;
add the password attribute "Secret password" to the envelope;
push data "This is a secret" into the envelope;
pop encrypted data from the envelope;
destroy the envelope;
```

That's all that's necessary. Since you've added a password attribute, cryptlib knows that you want to encrypt the data in the envelope with the password, so it encrypts the data and returns it to you. This process is referred to as enveloping the data.

The opposite, de-enveloping process consists of:

```
create the envelope;
push encrypted data into the envelope;
add the password attribute "Secret password" to the envelope;
pop decrypted data from the envelope;
destroy the envelope;
```

cryptlib knows the type of encrypted data that it's working with (it can inform you that you need to push in a password if you don't know that in advance), decrypts it with the provided password, and returns the result to you.

This example illustrates a feature of the de-enveloping process that may at first seem slightly unusual: You have to push in some encrypted data before you can add the password attribute needed to decrypt it. This is because cryptlib will automatically determine what to do with the data you give it, so if you added a password before you pushed in the encrypted data cryptlib wouldn't know what to do with the password.

Signing data is almost identical, except that you add a signature key attribute instead of a password. You can also add a number of other encryption attributes depending on the type of functionality you want. Since all of these require further knowledge of cryptlib's capabilities, only basic data, compressed-data, and password-based enveloping will be covered in this section.

Due to constraints in the underlying data formats that cryptlib supports, you can't perform more than one step of compression, encryption, or signing using a single envelope (the resulting data stream can't be encoded using most of the common data formats supported by cryptlib). If you want to perform more than one of these operations, you need to use multiple envelopes, one for each of the processing steps you want to perform. If you try and add an encryption attribute to an envelope which is set up for signing, or a signing attribute to an envelope which is set up for encryption, or some other conflicting combination, cryptlib will return a parameter error to indicate that the attribute type is invalid for this envelope since it's already being used for a different purpose.

Creating/Destroying Envelopes

Envelopes are accessed through envelope objects that work in the same general manner as the other container objects used by cryptlib. Before you can envelope or de-envelope data you need to create the appropriate type of envelope for the job. If you want to envelope some data, you would create the envelope with **cryptCreateEnvelope**, specifying the user who is to own the device object or CRYPT_UNUSED for the default, normal user, and the format for the enveloped data (for now you should use CRYPT_FORMAT_CRYPTLIB, the default format):

```
CRYPT_ENVELOPE cryptEnvelope;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

/* Perform enveloping */

cryptDestroyEnvelope( cryptEnvelope );
```

The Visual Basic version is:

```
Dim cryptEnvelope As Long

cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_CRYPTLIB

' Perform enveloping

cryptDestroyEnvelope cryptEnvelope
```

The C#, Java, and Python versions (here as elsewhere) migrate the output value to the return value, and return errors by throwing exceptions. The Python version is:

```
cryptEnvelope = cryptCreateEnvelope( cryptUser,
    CRYPT_FORMAT_CRYPTLIB )
```

The C# and Java version is:

```
int cryptEnvelope;

cryptEnvelope = crypt.CreateEnvelope( cryptUser,
    crypt.FORMAT_CRYPTLIB );
```

If you want to de-envelope the result of the previous enveloping process, you would create the envelope with format `CRYPT_FORMAT_AUTO`, which tells cryptlib to automatically detect and use the appropriate format to process the data:

```
CRYPT_ENVELOPE cryptEnvelope;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Perform de-enveloping */

cryptDestroyEnvelope( cryptEnvelope );
```

Note that the `CRYPT_ENVELOPE` is passed to the **cryptCreateEnvelope** by reference as the function modifies it when it creates the envelope. In all other routines in cryptlib, `CRYPT_ENVELOPE` is passed by value.

Sometimes when you're processing data in an envelope, you may not be able to add all of the data in an envelope, for example when you're trying to de-envelope a message that's been truncated due to a transmission error, or when you don't retrieve all of the processed data in the envelope before destroying it. When you destroy the envelope cryptlib will return `CRYPT_ERROR_INCOMPLETE` as a warning that not all of the data has been processed. The envelope will be destroyed as usual, but the warning is returned to indicate that you should have added further data or retrieved processed data before destroying the envelope.

The Data Enveloping Process

Although this section only covers basic data and password-based enveloping, the concepts that it covers apply to all the other types of enveloping as well, so you should familiarise yourself with this section even if you're only planning to use the more advanced types of enveloping such as digitally signed data enveloping. The general model for enveloping data is:

```
add any attributes such as passwords or keys
push in data
pop out processed data
```

To de-envelope data:

```

push in data
(cryptlib will inform you what resource(s) it needs to process the
data)
add the required attribute such as a password or key
pop out processed data

```

The enveloping/de-enveloping functions perform a lot of work in the background. For example when you add a password attribute to an envelope and follow it with some data, the function hashes the variable-length password down to create a fixed-length key for the appropriate encryption algorithm, generates a temporary session key to use to encrypt the data that you'll be pushing into the envelope, uses the fixed-length key to encrypt the session key, encrypts the data (taking into account the fact that most encryption modes can't encrypt individual bytes but require data to be present in fixed-length blocks), and then cleans up by erasing any keys and other sensitive information still in memory. This is why it's recommended that you use the envelope interface rather than trying to do the same thing yourself.

The **cryptPushData** and **cryptPopData** functions are used to push data into and pop data out of an envelope. For example to push the message "Hello world" into an envelope, you would use:

```
cryptPushData( envelope, "Hello world", 11, &bytesCopied );
```

The same operation in C# and Java is:

```
int bytesCopied = crypt.PushData( envelope, "Hello world" );
```

In Python this is:

```
bytesCopied = cryptPushData( envelope, "Hello world" )
```

The function will return an indication of how many bytes were copied into the envelope in `bytesCopied`. Usually this is the same as the number of bytes you pushed in, but if the envelope is almost full or you're trying to push in a very large amount of data, only some of the data may be copied in. This is useful when you want to process a large quantity of data in multiple sections, which is explained further on.

When you push in data, cryptlib may return an advisory `CRYPT_ENVELOPE_RESOURCE` status, which indicates that additional information such as a password or decryption key is required in order to continue. Until you supply the necessary resource, cryptlib can't process the data that you've given it, and any further attempts to push or pop data will fail with a `CRYPT_ENVELOPE_RESOURCE`. The handling of de-encryption resources is covered in more detail in the following sections.

Popping data works similarly to pushing data:

```
cryptPopData( envelope, buffer, bufferSize, &bytesCopied );
```

In this case you supply a buffer to copy the data to, and an indication of how many bytes you want to accept, and the function will return the number of bytes actually copied in `bytesCopied`. This could be anything from zero up to the full buffer size, depending on how much data is present in the envelope.

Once you've pushed the entire quantity of data that you want to process into an envelope, you need to use **cryptFlushData** to tell the envelope object to wrap up the data processing. If you try to push in any more data after this point, cryptlib will return a `CRYPT_ERROR_COMPLETE` error to indicate that processing of the data in the envelope has been completed and no more data can be added. Since the enveloped data contains all the information necessary to de-envelope it, it isn't necessary to perform the final flush during de-enveloping.

You can add enveloping and de-enveloping attributes to an envelope in the usual manner with **cryptSetAttribute** and **cryptSetAttributeString**. For example to add the password "password" to an envelope, you would set the `CRYPT_ENVINFO_PASSWORD` attribute:

```
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
"password", 8 );
```

The same operation in Visual Basic is:

```
cryptSetAttributeString cryptEnvelope, CRYPT_ENVINFO_PASSWORD, _  
    "password", 8
```

The various types of attributes that you can add are explained in more detail further on.

Data Size Considerations

When you add data to an envelope, cryptlib processes and encodes it in a manner that allows arbitrary amounts of data to be added. If cryptlib knows in advance how much data will be pushed into the envelope, it can use a more efficient encoding method since it doesn't have to take into account an indefinitely long data stream. You can notify cryptlib of the overall data size by setting the CRYPT_ENVINFO_DATASIZE attribute:

```
cryptSetAttribute( envelope, CRYPT_ENVINFO_DATASIZE, dataSize );
```

This tells cryptlib how much data will be added, and allows it to use the more efficient encoding format. If you push in more data than this before you wrap up the processing with **cryptFlushData**, cryptlib will return CRYPT_ERROR_OVERFLOW; if you push in less, it will return CRYPT_ERROR_UNDERFLOW.

There is one exception to this rule, which occurs when you're using the PGP/OpenPGP data format. PGP requires that the length be indicated at the start of every message, so you always have to set the CRYPT_ENVINFO_DATASIZE attribute when you perform PGP enveloping. If you try and push data into a PGP envelope without setting the data size, cryptlib will return CRYPT_ERROR_NOTINITED to tell you that it can't envelope the data without knowing its overall size in advance. PGP/OpenPGP enveloping is explained in more detail in "PGP" on page 180.

The amount of data popped out of an envelope never matches the amount pushed in, because the enveloping process adds encryption headers, digital signature information, and assorted other paraphernalia which is required to process a message. In many cases the overhead involved in wrapping up a block of data in an envelope can be noticeable, so you should always push and pop as much data at once into and out of an envelope as you can. For example if you have a 100-byte message and push it in as 10 lots of 10 bytes, this is much slower than pushing a single lot of 100 bytes. This behaviour is identical to the behaviour in applications like disk or network I/O, where writing a single big file to disk is a lot more efficient than writing 10 smaller files, and writing a single big network data packet is more efficient than writing 10 smaller data packets.

Push and popping unnecessarily small blocks of data when the total data size is unknown can also affect the overall enveloped data size. If you haven't told cryptlib how much data you plan to process with CRYPT_ENVINFO_DATASIZE then each time you pop a block of data from an envelope, cryptlib has to wrap up the current block and add header information to it to allow it to be de-enveloped later on.

Because this encoding overhead consumes extra space, you should again try to push and pop a single large data block rather than many small ones (to prevent worst-case behaviour, cryptlib will coalesce adjacent small blocks into a minimum block size of 10 bytes, so it won't return an individual block containing less than 10 bytes unless it's the last block in the envelope). This is again like disk data storage or network I/O, where many small files or data packets lead to greater fragmentation and wasted storage space or network overhead than a single large file or packet.

By default the envelope object which is created will have a 16K data buffer on DOS and 16-bit Windows systems, and a 32K buffer elsewhere. The size of the internal buffer affects the amount of extra processing that cryptlib needs to perform; a large buffer will reduce the amount of copying to and from the buffer, but will consume more memory (the ideal situation to aim for is one in which the data fits completely within the buffer, which means that it can be processed in a single operation). Since the process of encrypting and/or signing the data can increase its overall size, you

should make the buffer 1-2K larger than the total data size if you want to process the data in one go. The minimum buffer size is 4K, and on 16-bit systems the maximum buffer size is 32K-1.

If want to use a buffer which is smaller or larger than the default size, you can specify its size using the CRYPT_ATTRIBUTE_BUFFERSIZE attribute after the envelope has been created. For example if you knew you were going to be processing a single 80K message on a 32-bit system (you can't process more than 32K-1 bytes at once on a 16-bit system) you would use:

```
CRYPT_ENVELOPE cryptEnvelope;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );
cryptSetAttribute( cryptEnvelope, CRYPT_ATTRIBUTE_BUFFERSIZE,
    90000L );

/* Perform enveloping */

cryptDestroyEnvelope( cryptEnvelope );
```

(the extra 10K provides a generous safety margin for message expansion due to the enveloping process). When you specify the size of the buffer, you should try and make it as large as possible, unless you're pretty certain you'll only be seeing messages up to a certain size. Remember, the larger the buffer, the less processing overhead is involved in handling data. However, if you make the buffer excessively large it increases the probability that the data in it will be swapped out to disk, so it's a good idea not to go overboard on buffer size. You don't have to process the entire message at once, cryptlib provides the ability to envelope or de-envelope data in multiple sections to allow processing of arbitrary amounts of data even on systems with only small amounts of memory available.

Basic Data Enveloping

In the simplest case the entire message you want to process will fit into the envelope's internal buffer. The simplest type of enveloping does nothing to the data at all, but just wraps it and unwraps it:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

/* Create the envelope */
cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

/* Destroy the envelope */
cryptDestroyEnvelope( cryptEnvelope );
```

The Visual Basic equivalent is:

```
Dim cryptEnvelope As Long
Dim bytesCopied As Long

' Create the envelope
cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_CRYPTLIB

' Add the data size information and data, wrap up the processing, and
' pop out the processed data
cryptSetAttribute cryptEnvelope, CRYPT_ENVINFO_DATASIZE, messageLength
cryptPushData cryptEnvelope, message, messageLength, bytesCopied
cryptFlushData cryptEnvelope
cryptPopData cryptEnvelope, envelopedData, envelopedDataBufferSize, _
    bytesCopied
```

```
' Destroy the envelope
cryptDestroyEnvelope cryptEnvelope
```

The Python version is:

```
# Create the envelope
cryptEnvelope = cryptCreateEnvelope( cryptUser,
    CRYPT_FORMAT_CRYPTLIB )

# Add the data size information and data, wrap up the processing, and
# pop out the processed data
cryptEnvelope.ENVINFO_DATASIZE = len( message )
bytesCopied = cryptPushData( cryptEnvelope, message )
cryptFlushData( cryptEnvelope )
bytesCopied = cryptPopData( cryptEnvelope, envelopedData,
    envelopedDataBufferSize )

# Destroy the envelope
cryptDestroyEnvelope( cryptEnvelope )
```

The C# or Java version is:

```
int bytesCopied;

// Create the envelope
cryptEnvelope = crypt.CreateEnvelope( cryptUser,
    crypt.FORMAT_CRYPTLIB );

// Add the data size information and data, wrap up the processing, and
// pop out the processed data
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_DATASIZE,
    message.Length );
bytesCopied = crypt.PushData( cryptEnvelope, message );
crypt.FlushData( cryptEnvelope );
bytesCopied = crypt.PopData( cryptEnvelope, envelopedData,
    envelopedDataBufferSize );

// Destroy the envelope
crypt.DestroyEnvelope( cryptEnvelope );
```

(the above code is for C#, the Java version is virtually identical except that the `message.Length` of a C# byte array is `message.length` in Java).

To de-envelope the resulting data you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

/* Create the envelope */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Push in the enveloped data and pop out the recovered message */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataSize,
    &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize,
    &bytesCopied );

/* Destroy the envelope */
cryptDestroyEnvelope( cryptEnvelope );
```

The Visual Basic form is:

```
Dim cryptEnvelope As Long
Dim bytesCopied As Long

' Create the envelope
cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO

' Push in the enveloped data and pop out the recovered message
cryptPushData cryptEnvelope, envelopedData, envelopedDataSize, _
    bytesCopied
cryptFlushData cryptEnvelope
cryptPopData cryptEnvelope, message, messageBufferSize, bytesCopied

' Destroy the envelope
cryptDestroyEnvelope cryptEnvelope
```

This type of enveloping isn't terribly useful, but it does demonstrate how the enveloping process works.

Compressed Data Enveloping

A variation of basic data enveloping is compressed data enveloping which compresses or decompresses data during the enveloping process. Compressing data before signing or encryption improves the overall enveloping throughput (compressing data and encrypting the compressed data is faster than just encrypting the larger, uncompressed data), increases security by removing known patterns in the data, and saves storage space and network bandwidth.

To tell cryptlib to compress data that you add to an envelope, you should set the `CRYPT_ENVINFO_COMPRESSION` attribute before you add the data. This attribute doesn't take a value, so you should set it to `CRYPT_UNUSED`. The code to compress a message is then:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

/* Tell cryptlib to compress the data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_COMPRESSION,
    CRYPT_UNUSED );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

De-enveloping compressed data works exactly like decompressing normal data, cryptlib will transparently decompress the data for you and return the decompressed result when you call **cryptPopData**.

The compression/decompression process can cause a large change in data size between what you push and what you pop back out, so you typically end up pushing much more than you pop or popping much more than you push. In particular, you may end up pushing multiple lots of data before you can pop any compressed data out, or pushing a single lot of compressed data and having to pop multiple lots of decompressed data. This applies particularly to the final stages of enveloping when you flush through any remaining data, which signals the compressor to wrap up processing and move any remaining data into the envelope. This means that the flush can return `CRYPT_ERROR_OVERFLOW` to indicate that there's more data to be flushed, requiring multiple iterations of flushing and copying out data:

```
/* ... */

/* Flush out any remaining data */
do
{
    cryptFlushData( cryptEnvelope );
    cryptPopData( cryptEnvelope, outBuffer, BUFFER_SIZE, &bytesCopied );
}
while( bytesCopied > 0 );
```

To handle this in a more general manner, you should use the processing techniques described in "Enveloping Large Data Quantities" on page 155.

Password-based Encryption Enveloping

To do something useful (security-wise) to the data, you need to add a container or action object or other type of attribute to tell the envelope to secure the data in some

way. For example if you wanted to encrypt a message with a password you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

/* Add the password */
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password, passwordLength );

/* Add the data size information and data, wrap up the processing, and
    pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

The same operation in Java (for C# replace the `.length` with `.Length`) is:

```
int cryptEnvelope = crypt.CreateEnvelope( cryptUser,
    crypt.FORMAT_CRYPTLIB );

/* Add the password */
crypt.SetAttributeString( cryptEnvelope, crypt.ENVINFO_PASSWORD,
    password );

/* Add the data size information and data, wrap up the processing, and
    pop out the processed data */
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_DATASIZE,
    message.length );
int bytesCopied = crypt.PushData( cryptEnvelope, message );
crypt.FlushData( cryptEnvelope );
bytesCopied = crypt.PopData( cryptEnvelope, envelopedData,
    envelopedData.length );

crypt.DestroyEnvelope( cryptEnvelope );
```

To de-envelope the resulting data you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Push in the enveloped data and the password required to de-envelope
    it, and pop out the recovered message */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password, passwordLength );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

The de-enveloping process in Java is:

```
int cryptEnvelope = crypt.CreateEnvelope( cryptUser,
    crypt.FORMAT_AUTO );
int bytesCopied;

// Push in the enveloped data and the password required to
// de-envelope it, and pop out the recovered message
try {
    bytesCopied = crypt.PushData( cryptEnvelope, envelopedData );
}
catch ( CryptException e ) {
    if ( e.getStatus() != crypt.ENVELOPE_RESOURCE )
        throw e;
```

```

    }
    crypt.SetAttributeString( cryptEnvelope, crypt.ENVINFO_PASSWORD,
        password );
    crypt.FlushData( cryptEnvelope );
    crypt.PopData( cryptEnvelope, messageBuffer, messageBuffer.length );

    // Destroy the envelope
    crypt.DestroyEnvelope( cryptEnvelope );

```

The Visual Basic equivalent is:

```

Dim cryptEnvelope As Long
Dim bytesCopied As Long

cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO

' Push in the enveloped data and the password required to
' de-envelope it, and pop out the recovered message
cryptPushData cryptEnvelope, envelopedData, envelopedDataSize, _
    bytesCopied
cryptSetAttributeString cryptEnvelope, CRYPT_ENVINFO_PASSWORD, _
    password, Len( password )
cryptFlushData cryptEnvelope
cryptPopData cryptEnvelope, message, messageBufferSize, bytesCopied

' Destroy the envelope
cryptDestroyEnvelope cryptEnvelope

```

When you push in the password-protected data, **cryptPushData** will return **CRYPT_ENVELOPE_RESOURCE** to indicate that an additional resource (in this case the password) is required in order to continue. This is an advisory status that isn't needed in this case but can be useful for advanced de-envelope processing as described in "De-enveloping Mixed Data" on page 153.

If you add the wrong password, cryptlib will return a **CRYPT_ERROR_-WRONGKEY** error. You can use this to request a new password from the user and try again. For example to give the user the traditional three attempts at getting the password right you would replace the code to add the password with:

```

for( i = 0; i < 3; i++ )
{
    password = ...;
    if( cryptSetAttributeString( envelope, CRYPT_ENVINFO_PASSWORD,
        password, passwordLength ) == CRYPT_OK )
        break;
}

```

Conventional Encryption Enveloping

In addition to encrypting enveloped data with a password, it's possible to bypass the password step and encrypt the data directly using an encryption context. This context can either be used to encrypt the data directly (**CRYPT_ENVINFO_SESSIONKEY**) or indirectly by wrapping up a session key (**CRYPT_ENVINFO_KEY**). For example to encrypt data directly using IDEA with a raw session key you would do the following:

```

CRYPT_ENVELOPE cryptEnvelope;
CRYPT_CONTEXT cryptContext;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

/* Create the session key context and add it */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_IDEA );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEY,
    "0123456789ABCDEF", 16 );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SESSIONKEY,
    cryptContext );
cryptDestroyContext( cryptContext );

```

```
/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

To de-envelope the resulting data you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
CRYPT_CONTEXT cryptContext;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Push in the enveloped data and the session key context required to
   de-envelope it, and pop out the recovered message */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_IDEA );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEY,
    "0123456789ABCDEF", 16 );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SESSIONKEY,
    cryptContext );
cryptDestroyContext( cryptContext );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

Encrypting the data directly by using the context to wrap up the session key and then encrypting the data with that functions identically, except that the context is added as `CRYPT_ENVINFO_KEY` rather than `CRYPT_ENVINFO_SESSIONKEY`. The only real difference between the two is the underlying data format that cryptlib generates. As with the password-based de-enveloping, cryptlib will return an advisory `CRYPT_ENVELOPE_RESOURCE` status when you push in the data to let you know that you need to provide a decryption key in order to continue.

Raw session-key based enveloping isn't recommended since it bypasses much of the automated key management which is performed by the enveloping code, and requires the direct use of low-level encryption contexts. If all you want to do is change the underlying encryption algorithm used from the default triple DES, it's easier to do this by setting the `CRYPT_OPTION_ENCR_ALGO` attribute for the envelope as described in "Working with Configuration Options" on page 359.

Authenticated Enveloping

Encryption protects the confidentiality of the data in an envelope, but it doesn't provide any integrity protection - an attacker can modify the encrypted form of the data and obtain a corresponding modification of the decrypted form, and the simple use of encryption can't provide any protection against this. To provide integrity protection for the contents of an envelope, you need to use an authenticated envelope. You can tell cryptlib to add authentication to an envelope by setting the `CRYPT_ENVINFO_INTEGRITY` attribute before you push data into the envelope. By default this has a setting of `CRYPT_INTEGRITY_NONE`, which means that the contents are protected by encryption only. If you want to provide authentication (without encryption) for the data, you can set the `CRYPT_ENVINFO_INTEGRITY` to `CRYPT_INTEGRITY_MACONLY` (a MAC is a cryptographic message authentication code used to protect the contents of the envelope). The data processing works just like a standard encrypted envelope:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );
```

```

/* Tell cryptlib that we want integrity-protection instead of
   encryption and add the password */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_INTEGRITY,
    CRYPT_INTEGRITY_MACONLY );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password, passwordLength );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );

```

Note that you have to set the `CRYPT_ENVINFO_INTEGRITY` attribute before you add an encryption key or password, otherwise cryptlib will assume that you want to use the key to encrypt rather than authenticate the envelope contents.

When you de-envelope the data, cryptlib will use the MAC to check the integrity of the envelope contents. If the data has been modified, cryptlib will return `CRYPT_ERROR_SIGNATURE` once you've pushed the final portion of the enveloped data into the envelope:

```

CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Push in the enveloped data and the password required to de-envelope
   it, checking whether the data has been altered */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password, passwordLength );
status = cryptFlushData( cryptEnvelope );
if( status == CRYPT_ERROR_SIGNATURE )
    /* Data has been altered */;

/* The data is un-altered, pop out the recovered message */
cryptPopData( cryptEnvelope, message, messageBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );

```

You can also read the result of the integrity check by reading the `CRYPT_ENVINFO_SIGNATURE_RESULT` attribute, which works in much the same way as it does for signed envelopes, which are discussed in “Digitally Signed Enveloping” on page 164.

De-enveloping Mixed Data

Sometimes you won't know exactly what type of processing has been applied to the data that you're trying to de-envelope, so you can let cryptlib tell you what to do. When cryptlib needs some sort of resource (such as a password or an encryption key) to process the data that you've pushed into an envelope, it will return a `CRYPT_ENVELOPE_RESOURCE` status if you try and push in any more data or pop out the processed data. This status code is returned as soon as cryptlib knows enough about the data that you're pushing into the envelope to be able to process it properly. Typically, as soon as you start pushing in encrypted, signed, or otherwise processed data, **cryptPushData** will return `CRYPT_ENVELOPE_RESOURCE` to tell you that it needs some sort of resource in order to continue.

If you knew that the data that you were processing was either plain, unencrypted data, compressed data, or password-encrypted data created using the code shown earlier, then you could de-envelope it with:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied, status;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Push in the enveloped data and pop out the recovered message */
status = cryptPushData( cryptEnvelope, envelopedData,
    envelopedDataLength, &bytesCopied );
if( status == CRYPT_ENVELOPE_RESOURCE )
    cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
        password, passwordLength );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

If the data is enveloped without any processing or is compressed data, cryptlib will de-envelope it without requiring any extra input. If the data is enveloped using password-based encryption, cryptlib will return CRYPT_ENVELOPE_RESOURCE to indicate that it needs a password before it can continue.

This illustrates the manner in which the enveloped data contains enough information to allow cryptlib to process it automatically. If the data had been enveloped using some other form of processing (for example public-key encryption or digital signatures), cryptlib would ask you for the private decryption key or the signature check key at this time (it's actually slightly more complex than this, the details are explained in "Enveloping with Multiple Attributes" on page 166).

De-enveloping with a Large Envelope Buffer

If you've increased the envelope buffer size to allow the processing of large data quantities, the de-enveloping process may be slightly different. When de-enveloping data, cryptlib only reads an initial fixed amount of data before stopping and asking for user input such as the password or private key which is required to process the data. This is to avoid the situation where an envelope absorbs megabytes or even gigabytes of data only to report that it can't even begin to process it for lack of a decryption key. In this case the envelope will return CRYPT_ERROR_RESOURCE to indicate that it requires further information in order to continue. Once you've added the necessary de-enveloping attribute(s), you can either pop what's already been processed and continue as normal (see "Enveloping Large Data Quantities" on page 155) or, for a sufficiently large envelope buffer, push in the remaining data before popping it all at once:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied, status;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Push in the enveloped data and see if we need any special handling */
status = cryptPushData( cryptEnvelope, envelopedData,
    envelopedDataLength, &bytesCopied );
if( status == CRYPT_ENVELOPE_RESOURCE )
{
    /* Add the necessary de-enveloping attributes */
    /* ... */

    /* If only some of the data was accepted because the envelope
       stopped to request further instructions, push in the rest now */
    if( bytesCopied < envelopedDataLength )
    {
        int remainingBytesCopied;

        status = cryptPushData( cryptEnvelope, envelopedData + bytesIn,
            envelopedDataLength - bytesIn, &remainingBytesCopied );
        bytesIn += remainingBytesCopied;
    }
}
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize, &bytesCopied );
```



```
cryptDestroyEnvelope( cryptEnvelope );
```

This code checks whether the envelope has absorbed all of the enveloped data and, if not, pushes the remainder after adding the attribute(s) necessary for processing it. Once all of the data has been pushed, it pops the result as usual.

Obtaining Envelope Security Parameters

If you want to know the details of the encryption mechanism that's being used to protect the enveloped data, you can read various CRYPT_CTXINFO_XXX attributes from the envelope object which will return information from the encryption context(s) that are being used to secure the data. For example if you're encrypting or decrypting data you can get the encryption algorithm and mode and the key size being used with:

```
CRYPT_ALGO_TYPE cryptAlgo;
CRYPT_MODE_TYPE cryptMode;
int keySize;

cryptGetAttribute( cryptEnvelope, CRYPT_CTXINFO_ALGO, &cryptAlgo );
cryptGetAttribute( cryptEnvelope, CRYPT_CTXINFO_MODE, &cryptMode );
cryptGetAttribute( cryptEnvelope, CRYPT_CTXINFO_KEYSIZE, &keySize );
```

Enveloping Large Data Quantities

Sometimes, a message may be too big to process in one go or may not be available in its entirety, an example being data which is being sent or received over a network interface where only the currently transmitted or received portion is available. Although it's much easier to process a message in one go, it's also possible to envelope and de-envelope it a piece at a time (bearing in mind the earlier comment that the enveloping is most efficient when you push and pop data a single large block at a time rather than in many small blocks). With unknown amounts of data to be processed it generally isn't possible to use CRYPT_ENVINFO_DATASIZE, so in the sample code below this is omitted.

There are several strategies for processing data in multiple parts. The simplest one simply pushes and pops a fixed amount of data each time:

```
loop
  push data
  pop data
```

Since there's a little overhead added by the enveloping process, you should always push in slightly less data than the envelope buffer size. Alternatively, you can use the CRYPT_ATTRIBUTE_BUFFERSIZE to specify an envelope buffer which is slightly larger than the data block size that you want to use. The following code uses the first technique to password-encrypt a file in blocks of BUFFER_SIZE – 4K bytes:

```
CRYPT_ENVELOPE cryptEnvelope;
void *buffer;
int bufferCount;

/* Create the envelope with a buffer of size BUFFER_SIZE and add the
   password attribute */
cryptCreateEnvelope( &cryptEnvelope, cryptUser,
  CRYPT_FORMAT_CRYPTLIB );
cryptSetAttribute( cryptEnvelope, CRYPT_ATTRIBUTE_BUFFERSIZE,
  BUFFER_SIZE );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
  password, passwordLength );

/* Allocate a buffer for file I/O */
buffer = malloc( BUFFER_SIZE );
```

```
/* Process the entire file */
while( !endOfFile( inputFile ) )
{
    int bytesCopied;

    /* Read a (BUFFER_SIZE - 4K) block from the input file, envelope
       it, and write the result to the output file */
    bufferCount = readFile( inputFile, buffer, BUFFER_SIZE - 4096 );
    cryptPushData( cryptEnvelope, buffer, bufferCount, &bytesCopied );
    cryptPopData( cryptEnvelope, buffer, BUFFER_SIZE, &bytesCopied );
    writeFile( outputFile, buffer, bytesCopied );
}

/* Flush the last lot of data out of the envelope */
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, buffer, BUFFER_SIZE, &bytesCopied );
if( bytesCopied > 0 )
    writeFile( outputFile, buffer, bytesCopied );
free( buffer );

cryptDestroyEnvelope( cryptEnvelope );
```

The Visual Basic version is:

```
Dim cryptEnvelope As Long
Dim buffer() As Byte
Dim bufferCount As Integer
Dim bytesCopied As Long

' Create the envelope with a buffer of size BUFFER_SIZE and add the
' password attribute
cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_CRYPTLIB
cryptSetAttribute cryptEnvelope, CRYPT_ATTRIBUTE_BUFFERSIZE, _
    BUFFER_SIZE
cryptSetAttributeString cryptEnvelope, CRYPT_ENVINFO_PASSWORD, _
    password, Len( password )

' Allocate a buffer for file I/O
buffer = String( BUFFER_SIZE, vbNullChar )

Do While Not EndOfFile( inputFile )
    ' Read a (BUFFER_SIZE - 4K) block from the input file, envelope
    ' it, and write the result to an output file
    bufferCount = ReadFile inputFile, buffer, BUFFERSIZE - 4096
    cryptPushData cryptEnvelope, buffer, bufferCount, bytesCopied
    cryptPopData cryptEnvelope, buffer, BUFFER_SIZE, bytesCopied
    WriteFile outputFile, buffer, bytesCopied
Loop
cryptFlushData cryptEnvelope, buffer, BUFFER_SIZE, bytesCopied
If bytesCopied > 0 Then WriteFile outputFile, buffer, bytesCopied

cryptDestroyEnvelope cryptEnvelope
```

The code allocates a `BUFFER_SIZE` byte I/O buffer, reads up to `BUFFER_SIZE - 4K` bytes from the input file, and pushes it into the envelope. It then tells cryptlib to pop up to `BUFFER_SIZE` bytes of enveloped data back out into the buffer, takes whatever is popped out, and writes it to the output file. When it has processed the entire file, it pushes in the usual zero-length data block to flush any remaining data out of the buffer.

Note that the upper limit on `BUFFER_SIZE` depends on the system that you're running the code on. If you need to run it on a 16-bit system, `BUFFER_SIZE` is limited to 32K-1 bytes because of the length limit imposed by 16-bit integers, and the default envelope buffer size is 16K bytes unless you specify a larger default size using the `CRYPT_ATTRIBUTE_BUFFERSIZE` attribute.

Going to a lot of effort to exactly match a certain data size such as a power of two when pushing and popping data isn't really worthwhile, since the overhead added by the envelope encoding will always change the final encoded data length.

When you're performing compressed data enveloping or de-enveloping, the processing usually results in a large change in data size, in which case you may need

to use the technique described below that can handle arbitrarily-sized input and output quantities.

Alternative Processing Techniques

A slightly more complex technique is to always stuff the envelope as full as possible before trying to pop anything out of it:

```
loop
do
    push data
    while push status != CRYPT_ERROR_OVERFLOW
    pop data
```

This results in the most efficient use of the envelope's internal buffer, but is probably overkill for the amount of code complexity required:

```
CRYPT_ENVELOPE cryptEnvelope;
void *inBuffer, *outBuffer;
int bytesCopiedIn, bytesCopiedOut, bufferCount;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password, passwordLength );

/* Allocate input and output buffers */
inBuffer = malloc( BUFFER_SIZE );
outBuffer = malloc( BUFFER_SIZE );

/* Process the entire file */
while( !endOfFile( inputFile ) )
{
    int offset = 0;

    /* Read a buffer full of data from the file and push and pop it
       to/from the envelope */
    bufferCount = readFile( inputFile, inBuffer, BUFFER_SIZE );
    while( bufferCount > 0 )
    {
        /* Push as much as we can into the envelope */
        cryptPushData( cryptEnvelope, inBuffer + offset, bufferCount,
            &bytesCopiedIn );
        offset += bytesCopiedIn;
        bufferCount -= bytesCopiedIn;

        /* If we couldn't push everything in, the envelope is full, so
           we empty a buffers worth out */
        if( bufferCount > 0 )
        {
            cryptPopData( cryptEnvelope, outBuffer, BUFFER_SIZE,
                &bytesCopiedOut );
            writeFile( outputFile, outBuffer, bytesCopiedOut );
        }
    }
}

/* Flush out any remaining data */
do
{
    cryptFlushData( cryptEnvelope );
    cryptPopData( cryptEnvelope, outBuffer, BUFFER_SIZE,
        &bytesCopiedOut );
    if( bytesCopiedOut > 0 )
        writeFile( outputFile, outBuffer bytesCopiedOut );
}
while( bytesCopiedOut > 0 );
free( inBuffer );
free( outBuffer );

cryptDestroyEnvelope( cryptEnvelope );
```

Running the code to fill/empty the envelope in a loop is useful when you're applying a transformation such as data compression, which dramatically changes the length of the enveloped/de-enveloped data. In this case it's not possible to tell how much data

you can push into or pop out of the envelope because the length is transformed by the compression operation. It's also generally good practice to not write code that makes assumptions about the amount of internal buffer space available in the envelope, the above code will make optimal use of the envelope buffer no matter what its size.

Enveloping with Many Enveloping Attributes

There may be a special-case condition when you begin the enveloping that occurs if you've added a large number of password, encryption, or keying attributes to the envelope so that the header prepended to the enveloped data is particularly large. For example if you encrypt a message with different keys or passwords for several dozen recipients, the header information for all the keys could become large enough that it occupies a noticeable portion of the envelope's buffer. In this case you can push in a small amount of data to flush out the header information, and then push and pop data as usual:

```
add many password/encryption/keying attributes;
push a small amount of data;
pop data;
loop
    push data;
    pop data;
```

If you use this strategy then you can trim the difference between the envelope buffer size and the amount of data you push in at once down to about 1K; the 4K difference shown earlier took into account the fact that a little extra data would be generated the first time data was pushed due to the overhead of adding the envelope header:

```
CRYPT_ENVELOPE cryptEnvelope;
void *buffer;
int bufferCount;

/* Create the envelope and add many passwords */
cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password1, password1Length );
/* ... */
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password100, password100Length );

buffer = malloc( BUFFER_SIZE );

/* Read up to 100 bytes from the input file, push it into the envelope
   to flush out the header data, and write all the data in the
   envelope to the output file */
bufferCount = readFile( inputFile, buffer, 100 );
cryptPushData( cryptEnvelope, buffer, bufferCount, &bytesCopied );
cryptPopData( cryptEnvelope, buffer, BUFFER_SIZE, &bytesCopied );
writeFile( outputFile, buffer, bytesCopied );

/* Process the entire file */
while( !feof( inputFile ) )
{
    int bytesCopied;

    /* Read a BUFFER_SIZE block from the input file, envelope it, and
       write the result to the output file */
    bufferCount = readFile( inputFile, buffer, BUFFER_SIZE );
    cryptPushData( cryptEnvelope, buffer, bufferCount, &bytesCopied );
    cryptPopData( cryptEnvelope, buffer, BUFFER_SIZE, &bytesCopied );
    writeFile( outputFile, buffer, bytesCopied );
}

/* Flush the last lot of data out of the envelope */
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, buffer, BUFFER_SIZE, &bytesCopied );
if( bytesCopied > 0 )
    writeFile( outputFile, buffer, bytesCopied );
free( buffer );

cryptDestroyEnvelope( cryptEnvelope );
```

In the most extreme case (hundreds or thousands of passwords, encryption, or keying attributes added to an envelope), the header could fill the entire envelope buffer, and you would need to pop the initial data in multiple sections before you could process any more data using the usual push/pop loop. If you plan to use this many resources, it's better to specify the use of a larger envelope buffer using `CRYPT_ATTRIBUTE_BUFFERSIZE` in order to eliminate the need for such special-case processing for the header.

De-enveloping data that has been enveloped with multiple keying resources also has special requirements and is covered in the next section.

Advanced Enveloping

The previous chapter covered basic enveloping concepts and simple password-based enveloping. Extending beyond these basic forms of enveloping, you can also envelope data using public-key encryption or digitally sign the contents of the envelope. These types of enveloping require the use of public and private keys that are explained in various other chapters that cover key generation, key databases, and certificates.

cryptlib automatically manages objects such as public and private keys and keysets, so you can destroy them as soon as you've pushed them into the envelope. Although the object will appear to have been destroyed, the envelope maintains its own reference to it which it can continue to use for encryption or signing. This means that instead of the obvious:

```
create the key object;
create the envelope;
add the key object to the envelope;
push data into the envelope;
pop encrypted data from the envelope;
destroy the envelope;
destroy the key object;
```

it's also quite safe to use something like:

```
create the envelope;
create the key object;
add the key object to the envelope;
destroy the key object;
push data into the envelope;
pop encrypted data from the envelope;
destroy the envelope;
```

Keeping an object active for the shortest possible time makes it much easier to track, it's a lot easier to let cryptlib manage these things for you by handing them off to the envelope.

Public-Key Encrypted Enveloping

Public-key based enveloping works just like password-based enveloping except that instead of adding a password attribute you add a public key or certificate (when encrypting) or a private decryption key (when decrypting). For example if you wanted to encrypt data using a public key contained in `pubKeyContext`, you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

/* Add the public key */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_PUBLICKEY,
    pubKeyContext );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

You can also use a certificate in place of the public key, the envelope will handle both in the same way. The certificate is typically obtained by reading it from a keyset, either directly using `cryptGetPublicKey` as described in “Reading a Key from a Keyset” on page 226, or by setting the `CRYPT_ENVINFO_RECIPIENT` attribute as

described in “S/MIME Enveloping” on page 171. Using the CRYPT_ENVINFO_RECIPIENT attribute is the preferred option since it lets cryptlib handle a number of the complications that arise from reading keys for you.

When cryptlib encrypts the data in the envelope, it will use the algorithm specified with the CRYPT_OPTION_ENCR_ALGO option. If you want to change the encryption algorithm which is used, you can set the CRYPT_OPTION_ENCR_ALGO attribute for the envelope (or as a global configuration option) to the algorithm type you want, as described in “Working with Configuration Options” on page 359. Alternatively, you can push a raw session-key context into the envelope before you push in a public key, in which case cryptlib will use the context to encrypt the data rather than generating one itself.

The same operation in Java (for C# replace the `.length` with `.Length`) is:

```
int cryptEnvelope = crypt.CreateEnvelope( cryptUser,
    crypt.FORMAT_CRYPTLIB );

/* Add the public key */
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_PUBLICKEY,
    pubKeyContext );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_DATASIZE,
    message.length );
int bytesCopied = crypt.PushData( cryptEnvelope, message );
crypt.FlushData( cryptEnvelope );
bytesCopied = crypt.PopData( cryptEnvelope, envelopedData,
    envelopedData.length );

crypt.DestroyEnvelope( cryptEnvelope );
```

De-enveloping is slightly more complex since, unlike password-based enveloping, there are different keys used for enveloping and de-enveloping. In the simplest case if you know in advance which private decryption key is required to decrypt the data, you can add it to the envelope in the same way as with password-based enveloping:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Push in the enveloped data and the private decryption key required
   to de-envelope it, and pop out the recovered message */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_PRIVATEKEY,
    privKeyContext );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize, &bytesCopied
    );

cryptDestroyEnvelope( cryptEnvelope );
```

Although this leads to very simple code, it’s somewhat awkward since you may not know in advance which private key is required to decrypt a message. To make the private key handling process easier, cryptlib provides the ability to automatically fetch decryption keys from a private key keyset for you, so that instead of adding a private key, you add a private key keyset object and cryptlib takes care of obtaining the key for you. Alternatively, you can use a crypto device such as a smart card or Fortezza card to perform the decryption.

Using a private key from a keyset is slightly more complex than pushing in the private key directly since the private key stored in the keyset is usually encrypted or PIN-protected and will require a password or PIN supplied by the user to access it. This means that you have to supply a password to the envelope before the private key can be used to decrypt the data in it. This works as follows:

```
create the envelope;
add the decryption keyset;
push encrypted data into the envelope;
if( required resource = private key )
    add password to decrypt the private key;
pop decrypted data from the envelope;
destroy the envelope;
```

When you add the password, cryptlib will use it to try to recover the private key stored in the keyset you added previously. If the password is incorrect, cryptlib will return `CRYPT_ERROR_WRONGKEY`, otherwise it will recover the private key and then use that to decrypt the data. The full code to decrypt public-key enveloped data is therefore:

```
CRYPT_ENVELOPE cryptEnvelope;
CRYPT_ATTRIBUTE_TYPE requiredAttribute;
int bytesCopied, status;

/* Create the envelope and add the private key keyset and data */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_DECRYPT,
    privKeyKeyset );
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );

/* Find out what we need to continue and, if it's a private key, add
the password to recover it from the keyset */
cryptGetAttribute( cryptEnvelope, CRYPT_ATTRIBUTE_CURRENT,
    &requiredAttribute );
if( requiredAttribute != CRYPT_ENVINFO_PRIVATEKEY )
    /* Error */;
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password, passwordLength );
cryptFlushData( cryptEnvelope );

/* Pop the data and clean up */
cryptPopData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptDestroyEnvelope( cryptEnvelope );
```

The Visual Basic equivalent is:

```
Dim cryptEnvelope As Long
Dim requiredAttribute As CRYPT_ATTRIBUTE_TYPE
Dim bytesCopied As Long
Dim status As Long

' Create the envelope and add the private key and data
cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO
cryptSetAttribute cryptEnvelope, CRYPT_ENVINFO_KEYSET_DECRYPT, _
    privateKeyset
cryptPushData cryptEnvelope, envelopedData, envelopedDataLength, _
    bytesCopied

' Find out what we need to continue, and if it's a private key,
' add the password to recover it from the keyset
cryptGetAttribute cryptEnvelope, CRYPT_ATTRIBUTE_CURRENT, _
    requiredAttribute
If ( requiredAttribute <> CRYPT_ENVINFO_PRIVATEKEY ) Then
    ' Error
End If
cryptSetAttributeString cryptEnvelope, CRYPT_ENVINFO_PASSWORD, _
    password, len( password )
cryptFlushData cryptEnvelope

' Pop the data and clean up
cryptPopData cryptEnvelope, message, messageLength, bytesCopied
cryptDestroyEnvelope cryptEnvelope
```

In the unusual case where the private key isn't protected by a password or PIN, there's no need to add the password since cryptlib will use the private key as soon as you access the attribute information by reading it using **cryptGetAttribute**.

In order to ask the user for a password, it can be useful to know the name or label attached to the private key so you can display it as part of the password request

message. You can obtain the label for the required private key by reading the envelope's `CRYPT_ENVINFO_PRIVATEKEY_LABEL` attribute:

```
char label[ CRYPT_MAX_TEXTSIZE + 1 ];
int labelLength;

cryptGetAttributeString( cryptEnvelope,
    CRYPT_ENVINFO_PRIVATEKEY_LABEL, label, &labelLength );
label[ labelLength ] = '\0';
```

You can then use the key label when you ask the user for the password for the key.

Using a crypto device to perform the decryption is somewhat simpler since the PIN will already have been entered after **cryptDeviceOpen** was called, so there's no need to supply it as `CRYPT_ENVINFO_PASSWORD`. To use a crypto device, you add the device in place of the private key keyset:

```
CRYPT_ENVELOPE cryptEnvelope;
CRYPT_ATTRIBUTE_TYPE requiredAttribute;
int bytesCopied, status;

/* Create the envelope and add the crypto device and data */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_DECRYPT,
    cryptDevice );
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );

/* Find out what we need to continue. Since we've told the envelope
   to use a crypto device, it'll perform the decryption as soon as we
   ask it to using the device, so we shouldn't have to supply anything
   else */
cryptGetAttribute( cryptEnvelope, CRYPT_ATTRIBUTE_CURRENT,
    &requiredAttribute );
if( requiredAttribute != CRYPT_ATTRIBUTE_NONE )
    /* Error */;
cryptFlushData( cryptEnvelope );

/* Pop the data and clean up */
cryptPopData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptDestroyEnvelope( cryptEnvelope );
```

Note how **cryptGetAttribute** now reports that there's nothing further required (since the envelope has used the private key in the crypto device to performed the decryption), and you can continue with the de-enveloping process.

Code that can handle the use of either a private key keyset or a crypto device for the decryption is a straightforward extension of the above:

```
CRYPT_ENVELOPE cryptEnvelope;
CRYPT_ATTRIBUTE_TYPE requiredAttribute;
int bytesCopied, status;

/* Create the envelope and add the keyset or crypto device and data */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_DECRYPT,
    cryptKeysetOrDevice );
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );

/* Find out what we need to continue. If what we added was a crypto
   device, the decryption will occur once we query the envelope. If
   what we added was a keyset, we need to supply a password for the
   decryption to happen */
cryptGetAttribute( cryptEnvelope, CRYPT_ATTRIBUTE_CURRENT,
    &requiredAttribute );
if( requiredAttribute != CRYPT_ATTRIBUTE_NONE )
{
    char label[ CRYPT_MAX_TEXTSIZE + 1 ];
    int labelLength;

    if( requiredAttribute != CRYPT_ENVINFO_PASSWORD )
        /* Error */;
```

```
/* Get the label for the private key and obtain the required
password from the user */
cryptGetAttributeString( cryptEnvelope,
    CRYPT_ENVINFO_PRIVATEKEY_LABEL, label, &labelLength );
label[ labelLength ] = '\0';
getPassword( label, password, &passwordLength );

/* Add the password required to decrypt the private key */
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password, passwordLength );
}
cryptFlushData( cryptEnvelope );

/* Pop the data and clean up */
cryptPopData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptDestroyEnvelope( cryptEnvelope );
```

Digitally Signed Enveloping

Digitally signed enveloping works much like the other enveloping types except that instead of adding an encryption or decryption attribute you supply a private signature key (when enveloping) or a public key or certificate (when de-enveloping). For example if you wanted to sign data using a private signature key contained in sigKeyContext, you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

/* Add the signing key */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
    sigKeyContext );

/* Add the data size information and data, wrap up the processing, and
pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

The signature key could be a native cryptlib key, but it could also be a key from a crypto device such as a smart card or Fortezza card. They both work in the same way for signing data.

The Java version of the signed enveloping process (for C# replace the .length with .Length) is:

```
int cryptEnvelope = crypt.CreateEnvelope( cryptUser,
    crypt.FORMAT_CRYPTLIB );

/* Add the public key */
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_SIGNATURE,
    sigKeyContext );

/* Add the data size information and data, wrap up the processing, and
pop out the processed data */
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_DATASIZE,
    message.length );
int bytesCopied = crypt.PushData( cryptEnvelope, message );
crypt.FlushData( cryptEnvelope );
bytesCopied = crypt.PopData( cryptEnvelope, envelopedData,
    envelopedData.length );

crypt.DestroyEnvelope( cryptEnvelope );
```

The Visual Basic equivalent is:

```
cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_CRYPTLIB
```

```

' Add the signing key
cryptSetAttribute cryptEnvelope, CRYPT_ENVINFO_SIGNATURE, _
    sigKeyContext

' Add the data size information and data, wrap up the processing,
' and pop out the processed data
cryptSetAttribute cryptEnvelope, CRYPT_ENVINFO_DATASIZE, messageLength
cryptPushData cryptEnvelope, message, messageLength, bytesCopied
cryptFlushData cryptEnvelope
cryptPopData cryptEnvelope, envelopedData, envelopedDataBufferSize, _
    bytesCopied

cryptDestroyEnvelope cryptEnvelope

```

When cryptlib signs the data in the envelope, it will hash it with the algorithm specified with the `CRYPT_OPTION_ENCR_HASH` option. If you want to change the hashing algorithm which is used, you can set the `CRYPT_OPTION_ENCR_HASH` attribute for the envelope (or as a global configuration option) to the algorithm type you want, as described in “Working with Configuration Options” on page 359. Alternatively, you can push a hash context into the envelope before you push in a signature key, in which case cryptlib will associate the signature key with the last hash context you pushed in.

If you’re worried about some obscure (and rather unlikely) attacks on private keys, you can enable the `CRYPT_OPTION_MISC_SIDECHANNELPROTECTION` option as explained in “Working with Configuration Options” on page 359.

As with public-key based enveloping, verifying the signed data requires a different key for this part of the operation, in this case a public key or key certificate. In the simplest case if you know in advance which public key is required to verify the signature, you can add it to the envelope in the same way as with the other envelope types:

```

CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Add the enveloped data and the signature check key required to
   verify the signature, and pop out the recovered message */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
    sigCheckKeyContext );
cryptPopData( cryptEnvelope, message, messageBufferSize, &bytesCopied
    );

cryptDestroyEnvelope( cryptEnvelope );

```

Although this leads to very simple code, it’s somewhat awkward since you may not know in advance which public key or key certificate is required to verify the signature on the message. To make the signature verification process easier, cryptlib provides the ability to automatically fetch signature verification keys from a public-key keyset for you, so that instead of supplying a public key or key certificate, you add a public-key keyset object before you start de-enveloping and cryptlib will take care of obtaining the key for you. This option works as follows:

```

create the envelope;
add the signature check keyset;
push signed data into the envelope;
pop plain data from the envelope;
if( required resource = signature check key )
    read signature verification result;

```

The full code to verify signed data is therefore:

```

CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied, signatureResult, status;

/* Create the envelope and add the signature check keyset */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_SIGCHECK,
    sigCheckKeyset );

/* Push in the signed data and pop out the recovered message */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize,
    &bytesCopied );

/* Determine the result of the signature check */
cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_RESULT,
    &signatureResult );

```

The same process in Java (for C# replace the .length with .Length) is:

```

/* Create the envelope and add the signature check keyset */
int cryptEnvelope = crypt.CreateEnvelope(cryptUser,
    crypt.FORMAT_AUTO );
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_KEYSET_SIGCHECK,
    sigCheckKeyset );

/* Push in the signed data and pop out the recovered message */
int bytesCopied = crypt.PushData( cryptEnvelope, envelopedData );
crypt.FlushData( cryptEnvelope );
bytesCopied = crypt.PopData( cryptEnvelope, message, message.length );

/* Determine the result of the signature check */
int signatureResult = crypt.GetAttribute( cryptEnvelope,
    crypt.ENVINFO_SIGNATURE_RESULT );

```

The Visual Basic version is:

```

Dim signatureResult As Long

' Create the envelope and add the signature check keyset
cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO
cryptSetAttribute cryptEnvelope, CRYPT_ENVINFO_KEYSET_SIGCHECK, _
    sigCheckKeyset

' Push in the signed data and pop out the recovered message
cryptPushData cryptEnvelope, envelopedData, envelopedDataLength, _
    bytesCopied
cryptPopData cryptEnvelope, message, messageBufferSize, bytesCopied

' Determine the result of the signature check
cryptGetAttribute cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_RESULT, _
    signatureResult

```

The signature result will typically be CRYPT_OK (the signature verified), CRYPT_ERROR_SIGNATURE (the signature did not verify), or CRYPT_ERROR_NOTFOUND (the key needed to check the signature wasn't found in the keyset).

Most signed data in use today uses a format popularised in S/MIME that includes the signature verification key with the data being signed as a certificate chain. For this type of data you don't need to provide a signature verification key, since it's already included with the signed data. Details on creating and processing data in this format is given in "S/MIME Enveloping" on page 171.

Enveloping with Multiple Attributes

Sometimes enveloped data can have multiple sets of attributes applied to it, for example encrypted data might be encrypted with two different passwords to allow it to be decrypted by two different people:

```

CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

```

```

/* Add two different passwords to the envelope */
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password1, password1Length );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password2, password2Length );

/* Add the data size information and data, wrap up the processing, and
    pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );

```

In this case either of the two passwords can be used to decrypt the data. This can be extended indefinitely, so that 5, 10, 50, or 100 passwords could be used (of course with 100 different passwords able to decrypt the data, it's questionable whether it's worth the effort of encrypting it at all, however this sort of multi-user encryption could be useful for public-key encrypting messages sent to collections of people such as mailing lists). The same applies for public-key enveloping, in fact the various encryption types can be mixed if required so that (for example) either a private decryption key or a password could be used to decrypt data.

Similarly, an envelope can have multiple signatures applied to it:

```

CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied

cryptCreateEnvelope( &cryptEnvelope, cryptUser,
    CRYPT_FORMAT_CRYPTLIB );

/* Add two different signing keys to the envelope */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
    cryptSigKey1 );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
    cryptSigKey2 );

/* Add the data size information and data, wrap up the processing, and
    pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );

```

In this case the envelope will be signed by both keys. As with password-based enveloping, this can also be extended indefinitely to allow additional signatures on the data, although it would be somewhat unusual to place more than one or two signatures on a piece of data.

When de-enveloping data that has been enveloped with a choice of multiple attributes, cryptlib builds a list of the attributes required to decrypt or verify the signature on the data, and allows you to query the required attribute information and choose the one you want to work with.

Processing Multiple De-enveloping Attributes

The attributes required for de-enveloping are managed through the use of an attribute cursor as described in “Attribute Lists and Attribute Groups” on page 38. You can use the attribute cursor to determine which attribute is required for the de-enveloping process. Once you're iterating through the attributes, all that's left to do is to plug in the appropriate handler routines to manage each attribute requirement that could be encountered. As soon as one of the attributes required to continue is added to the envelope, cryptlib will delete the required-attribute list and continue, so the attempt to move the cursor to the next entry in the list will fail and the program will drop out of

the processing loop. For example to try a password against all of the possible passwords that might decrypt the message that was enveloped above, you would use:

```
int status

/* Get the decryption password from the user */
password = ...;

if( cryptSetAttribute( envelope, CRYPT_ATTRIBUTE_CURRENT_GROUP,
CRYPT_CURSOR_FIRST ) == CRYPT_OK )
do
{
CRYPT_ATTRIBUTE_TYPE requiredAttribute;

/* Get the type of the required attribute at the cursor position
*/
cryptGetAttribute( envelope, CRYPT_ATTRIBUTE_CURRENT,
&requiredAttribute );

/* Make sure we really do require a password resource */
if( requiredAttribute != CRYPT_ENVINFO_PASSWORD )
/* Error */;

/* Try the password. If everything is OK, we'll drop out of the
loop */
status = cryptSetAttributeString( envelope,
CRYPT_ENVINFO_PASSWORD, password, passwordLength );
}
while( status == CRYPT_WRONGKEY && \
cryptSetAttribute( envelope, CRYPT_ATTRIBUTE_CURRENT_GROUP,
CRYPT_CURSOR_NEXT ) == CRYPT_OK );
```

This steps through each required attribute in turn and tries the supplied password to see if it matches. As soon as the password matches, the data can be decrypted, and we drop out of the loop and continue the de-enveloping process.

To extend this a bit further, let's assume that the data could be enveloped using a password or a public key (requiring a private decryption key to decrypt it, either one from a keyset or a crypto device such as a smart card or Fortezza card). The code inside the loop above then becomes:

```
CRYPT_ATTRIBUTE_TYPE requiredAttribute;

/* Get the type of the required resource at the cursor position */
cryptGetAttribute( envelope, CRYPT_ATTRIBUTE_CURRENT,
&requiredAttribute );

/* If the decryption is being handled via a crypto device, we don't
need to take any further action, the data has already been
decrypted */
if( requiredAttribute != CRYPT_ATTRIBUTE_NONE )
{
/* Make sure we really do require a password attribute */
if( requiredAttribute != CRYPT_ENVINFO_PASSWORD && \
requiredAttribute != CRYPT_ENVINFO_PRIVATEKEY )
/* Error */;

/* Try the password. If everything is OK, we'll drop out of the
loop */
status = cryptSetAttributeString( envelope, CRYPT_ENVINFO_PASSWORD,
password, passwordLength );
}
```

If what's required is a CRYPT_ENVINFO_PASSWORD, cryptlib will apply it directly to decrypt the data. If what's required is a CRYPT_ENVINFO_PRIVATEKEY, cryptlib will either use the crypto device to decrypt the data if it's available, or otherwise use the password to try to recover the private key from the keyset and then use that to decrypt the data.

Iterating through each required signature attribute when de-enveloping signed data is similar, but instead of trying to provide the necessary decryption information you would provide the necessary signature check information (if requested, many envelopes carry their own signature verification keys with them) and display the

resulting signature information. Unlike encryption de-enveloping attributes, cryptlib won't delete the signature information once it has been processed, so you can re-read the information multiple times:

```
int status

if( cryptSetAttribute( envelope, CRYPT_ATTRIBUTE_CURRENT_GROUP,
CRYPT_CURSOR_FIRST ) == CRYPT_OK )
do
{
    CRYPT_ATTRIBUTE_TYPE requiredAttribute;
    int sigResult;

    /* Get the type of the required attribute at the cursor position
    */
    cryptGetAttribute( envelope, CRYPT_ATTRIBUTE_CURRENT,
&requiredAttribute );

    /* Make sure we really do have signature */
    if( requiredAttribute != CRYPT_ENVINFO_SIGNATURE )
        /* Error */;

    /* Get the signature result */
    status = cryptSetAttribute( envelope,
CRYPT_ENVINFO_SIGNATURE_RESULT, &sigResult );
}
while( cryptStatusOK( status ) && \
    cryptSetAttribute( envelope, CRYPT_ATTRIBUTE_CURRENT_GROUP,
CRYPT_CURSOR_NEXT ) == CRYPT_OK );
```

This steps through each signature in turn and reads the result of the signature verification for that signature, stopping when an invalid signature is found or when all signatures are processed.

Nested Envelopes

Sometimes it may be necessary to apply multiple levels of processing to data, for example you may want to both sign and encrypt data. cryptlib allows enveloped data to be arbitrarily nested, with each nested content type being either further enveloped data or (finally) the raw data payload. For example to sign and encrypt data you would do the following:

```
create the envelope;
add the signature key;
push in the raw data;
pop out the signed data;
destroy the envelope;

create the envelope;
add the encryption key;
push in the previously signed data;
pop out the signed, encrypted data;
destroy the envelope;
```

This nesting process can be extended arbitrarily with any of the cryptlib content types.

Since cryptlib's enveloping isn't sensitive to the content type (that is, you can push in any type of data and it'll be enveloped in the same way), you need to notify cryptlib of the actual content type being enveloped if you're using nested envelopes. You can set the content type being enveloped using the `CRYPT_ENVINFO_CONTENTTYPE` attribute, giving as value the appropriate `CRYPT_CONTENT_type`. For example to specify that the data being enveloped is signed data, you would use:

```
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_CONTENTTYPE,
CRYPT_CONTENT_SIGNEDDATA );
```

The default content type is plain data, so if you don't explicitly set a content type cryptlib will assume it's just raw data. The other content types are described in "Other Certificate Object Extensions" on page 338.

Using the nested enveloping example shown above, the full enveloping procedure would be:

```
create the envelope;
add the signature key;
(cryptlib sets the content type to the default 'plain data')
push in the raw data;
pop out the signed data;
destroy the envelope;

create the envelope;
set the content type to 'signed data';
add the encryption key;
push in the previously signed data;
pop out the signed, encrypted data;
destroy the envelope;
```

This will mark the innermost content as plain data (the default), the next level as signed data, and the outermost level as encrypted data.

Unwrapping nested enveloped data is the opposite of the enveloping process. For each level of enveloped data, you can obtain its type (once you've pushed enough of it into the envelope to allow cryptlib to decode it) by reading the `CRYPT_ENVINFO_CONTENTTYPE` attribute:

```
CRYPT_ATTRIBUTE_TYPE contentType;

cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_CONTENTTYPE,
                  &contentType );
```

Processing nested enveloped data therefore involves unwrapping successive layers of data until you finally reach the raw data content type.

S/MIME

S/MIME is a standard format for transferring signed, encrypted, or otherwise processed data as a MIME-encoded message (for example as email or embedded in a web page). The MIME-encoding is only used to make the result palatable to mailers, it's also possible to process the data without the MIME encoding.

The exact data formatting and terminology used requires a bit of further explanation. In the beginning there was PKCS #7, a standard format for signed, encrypted, or otherwise processed data. When the earlier PEM secure mail standard failed to take off, PKCS #7 was wrapped up in MIME encoding and christened S/MIME version 2. Eventually PKCS #7 was extended to become the Cryptographic Message Syntax (CMS), and when that's wrapped in MIME it's called S/MIME version 3.

In practice it's somewhat more complicated than this since there's significant blurring between S/MIME version 2 and 3 (and PKCS #7 and CMS). The main effective difference between the two is that PKCS #7/SMIME version 2 is completely tied to X.509 certificates, certification authorities, certificate chains, and other paraphernalia, CMS can be used without requiring all these extras if necessary, and S/MIME version 3 restricts CMS back to requiring X.509 for S/MIME version 2 compatibility.

The cryptlib native format is CMS used in the configuration that doesn't tie it to the use of certificates (so it'll work with PGP/OpenPGP keys, raw public/private keys, and other keying information as well as with X.509 certificates). In addition to this format, cryptlib also supports the S/MIME format which is tied to X.509 — this is just the cryptlib native format restricted so that the full range of key management options aren't available. If you want to interoperate with other implementations, you should use this format since many implementations can't work with the newer key management options that were added in CMS.

You can specify the use of the restricted CMS/SMIME format when you create an envelope with the formatting specifier `CRYPT_FORMAT_CMS` or `CRYPT_FORMAT_SMIME` (they're almost identical, the few minor differences are explained in “Extra Signature Information” on page 177), which tells cryptlib to use the restricted CMS/SMIME rather than the (default) unrestricted CMS format. You can also use the format specifiers with `cryptExportKeyEx` and `cryptCreateSignatureEx` (which take as their third argument the format specifier) as explained in “Exchanging Keys” on page 278, and “Signing Data” on page 284.

S/MIME Enveloping

Although it's possible to use the S/MIME format directly with the mid-level signature and encryption functions, S/MIME requires a considerable amount of extra processing above and beyond that required by cryptlib's default format, so it's easiest to let cryptlib take care of this extra work for you by using the enveloping functions to process S/MIME data.

To create an envelope that uses the S/MIME format, call `cryptCreateEnvelope` as usual but specify a format type of `CRYPT_FORMAT_SMIME` instead of the usual `CRYPT_FORMAT_CRYPTLIB`:

```
CRYPT_ENVELOPE cryptEnvelope;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_SMIME );

/* Perform enveloping */

cryptDestroyEnvelope( cryptEnvelope );
```

Creating the envelope in this way restricts cryptlib to using the standard X.509-based S/MIME data format instead of the more flexible data format which is used for envelopes by default.

Encrypted Enveloping

S/MIME supports password-based enveloping in the same way as ordinary cryptlib envelopes (in fact the two formats are identical). Public-key encrypted enveloping is supported only when the public key is held in an X.509 certificate. Because of this restriction the private decryption key must also have a certificate attached to it. Apart from these restrictions, public-key based S/MIME enveloping works the same way as standard cryptlib enveloping. For example to encrypt data using the key contained in an X.509 certificate you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_SMIME );

/* Add the certificate */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_PUBLICKEY,
    certificate );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

Since the certificate will originally come from a keyset, a simpler alternative to reading the certificate yourself and explicitly adding it to the envelope is to let cryptlib do it for you by first adding the keyset to the envelope and then specifying the email address of the recipient or recipients of the message with the CRYPT_ENVINFO_RECIPIENT attribute:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_SMIME );

/* Add the encryption keyset and recipient email address */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_ENCRYPT,
    cryptKeyset );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_RECIPIENT,
    "person@company.com", 18 );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

The same thing in Java (for C# replace the `.length` with `.Length`) is:

```
int cryptEnvelope = crypt.CreateEnvelope( cryptUser,
    crypt.FORMAT_SMIME );

/* Add the encryption keyset and recipient email address */
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_KEYSET_ENCRYPT,
    cryptKeyset );
crypt.SetAttributeString( cryptEnvelope, crypt.ENVINFO_RECIPIENT,
    "person@company.com" );
```

```

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
crypt.SetAttribute( cryptEnvelope, crypt.ENVINFO_DATASIZE,
    message.length );
int bytesCopied = crypt.PushData( cryptEnvelope, message );
crypt.FlushData( cryptEnvelope );
bytesCopied = crypt.PopData( cryptEnvelope, envelopedData,
    envelopedData.length );

crypt.DestroyEnvelope( cryptEnvelope );

```

The Visual Basic equivalent is:

```

cryptCreateEnvelope cryptEnvelope, cryptUser, CRYPT_FORMAT_SMIME

' Add the encryption keyset and recipient email address
cryptSetAttribute cryptEnvelope, CRYPT_ENVINFO_KEYSET_ENCRYPT, _
    cryptKeyset
cryptSetAttributeString cryptEnvelope, CRYPT_ENVINFOR_RECIPIENT, _
    "person@company.com", 18

' Add the data size information and data, wrap up the processing,
' and pop out the processed data
cryptSetAttribute cryptEnvelope, CRYPT_ENVINFO_DATASIZE, messageLength
cryptPushData cryptEnvelope, message, messageLength, bytesCopied
cryptFlushData cryptEnvelope
cryptPopData cryptEnvelope, envelopedData, envelopedDataBufferSize, _
    bytesCopied

cryptDestroyEnvelope cryptEnvelope

```

For each message recipient that you add, cryptlib will look up the key in the encryption keyset and add the appropriate information to the envelope to encrypt the message to that person. This is the recommended way of handling public-key encrypted enveloping, since it lets cryptlib handle the certificate details for you and makes it possible to manage problem areas such as cases where the same email address is present in multiple certificates of which only one is valid for message encryption. If you want to handle this case yourself, you have to use a keyset query to search the duplicate certificates and select the appropriate one as described in “Handling Multiple Certificates with the Same Name” on page 230.

The encryption keyset doesn’t have to be local. If you use an HTTP keyset as described in “HTTP Keysets” on page 221, cryptlib will fetch the required certificate directly from the remote CA, saving you the effort of having to maintain and update a local set of certificates. This use of HTTP keysets makes it very easy to distribute certificates over the Internet.

De-enveloping works as for standard enveloping:

```

CRYPT_ENVELOPE cryptEnvelope;
CRYPT_ATTRIBUTE_TYPE requiredAttribute;
int bytesCopied, status;

/* Create the envelope and add the private key keyset and data */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_DECRYPT,
    privKeyKeyset );
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );

/* Find out what we need to continue and, if it's a private key, add
   the password to recover it */
cryptGetAttribute( cryptEnvelope, CRYPT_ATTRIBUTE_CURRENT,
    &requiredAttribute );
if( requiredAttribute != CRYPT_ENVINFO_PRIVATEKEY )
    /* Error */;
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
    password, passwordLength );
cryptFlushData( cryptEnvelope );

/* Pop the data and clean up */
cryptPopData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptDestroyEnvelope( cryptEnvelope );

```

More information on public-key encrypted enveloping, including its use with crypto devices such as smart cards and Fortezza cards, is given in “Public-Key Encrypted Enveloping” on page 160.

Digitally Signed Enveloping

S/MIME digitally signed enveloping works just like standard enveloping except that the signing key is restricted to one that has a full chain of X.509 certificates (or at least a single certificate) attached to it. For example if you wanted to sign data using a private key contained in `sigKeyContext`, you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_SMIME );

/* Add the signing key */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
    sigKeyContext );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

When you sign data in this manner, cryptlib includes any certificates attached to the signing key alongside the message. Although you can sign a message using a key with a single certificate attached to it, it's safer to use one that has a full certificate chain associated with it because including only the key certificate with the message requires that the recipient locate any other certificates that are required to verify the signature. Since there's no easy way to do this, signing a message using only a standalone certificate can cause problems when the recipient tries to verify the signature.

Verifying the signature on the data works slightly differently from the normal signature verification process since the signed data already carries with it the complete certificate chain required for verification. This means that you don't have to push a signature verification keyset or key into the envelope because the required certificate is already included with the data:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied, sigCheckStatus;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );

/* Push in the enveloped data and pop out the recovered message */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize, &bytesCopied
    );

/* Determine the result of the signature check */
cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_RESULT,
    &sigCheckStatus );

cryptDestroyEnvelope( cryptEnvelope );
```

Since the certificate is included with the data, anyone could alter the data, re-sign it with their own certificate, and then attach their certificate to the data. To avoid this problem, cryptlib provides the ability to verify the chain of certificates, which works in combination with cryptlib's certificate trust manager. You can obtain the certificate object containing the signing certificate chain with:

```
CRYPT_CERTIFICATE cryptCertChain;

cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
    &cryptCertChain );
```

You can work with this certificate chain as usual, for example you may want to display the certificates and any related information to the user. At the least, you should verify the chain using **cryptCheckCert**. You may also want to perform a validity check using RTCS, revocation checking using CRLs or OCSP, and any other certificate checks that you consider necessary. More details on working with certificate chains are given in “Certificate Chains” on page 310, details on basic signed enveloping (including its use with crypto devices like smart cards and Fortezza cards) are given in “Digitally Signed Enveloping” on page 164, details on validity checking with RTCS are given in “Certificate Status Checking using RTCS” on page 247, and details on revocation checking with OCSP are given in “Certificate Revocation Checking using OCSP” on page 252.

Detached Signatures

So far, the signature for the signed data has always been included with the data itself, allowing it to be processed as a single blob. cryptlib also provides the ability to create detached signatures in which the signature is held separate from the data. This leaves the data being signed unchanged and produces a standalone signature as the result of the encoding process.

To specify that an envelope should produce a detached signature rather than standard signed data, you should set the envelope’s CRYPT_ENVINFO_DETACHED-SIGNATURE attribute to ‘true’ (any nonzero value) before you push in any data

```
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DETACHEDSIGNATURE,
    1 );
```

Apart from that, the creation of detached signatures works just like the creation of standard signed data, with the result of the enveloping process being the standalone signature (without the data attached):

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_SMIME );

/* Add the signing key and specify that we're using a detached
signature */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
    sigKeyContext );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DETACHEDSIGNATURE,
    1 );

/* Add the data size information and data, wrap up the processing, and
pop out the detached signature */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, detachedSignature,
    detachedSignatureBufferSize, &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

Verifying a detached signature requires an extra processing step since the signature is no longer bundled with the data. First, you need to push in the detached signature (to tell cryptlib what to do with any following data). After you’ve pushed in the signature and followed it up with the usual **cryptFlushData** to wrap up the processing, you need to push in the data that was signed by the detached signature as the second processing step:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied, sigCheckStatus;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
```

```
/* Push in the detached signature */
cryptPushData( cryptEnvelope, detachedSignature, detachedSigLength,
               &bytesCopied );
cryptFlushData( cryptEnvelope );

/* Push in the data */
cryptPushData( cryptEnvelope, data, dataLength, NULL );
cryptFlushData( cryptEnvelope );

/* Determine the result of the signature check */
cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_RESULT,
                  &sigCheckStatus );

cryptDestroyEnvelope( cryptEnvelope );
```

Since the data wasn't enveloped to begin with, there's nothing to de-envelope, which means there's nothing to pop out of the envelope apart from the signing certificate chain that you can obtain as before by reading the CRYPT_ENVINFO_SIGNATURE attribute.

In case you're not sure whether a signature includes data or not, you can query its status by checking the value of the CRYPT_ENVINFO_DETACHEDSIGNATURE attribute after you've pushed in the signature:

```
int isDetachedSignature;

/* Push in the signed enveloped data */
cryptPushData( cryptEnvelope, signedData, signedDataLength,
               &bytesCopied );

/* Check the signed data type */
cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_DETACHEDSIGNATURE,
                  &isDetachedSignature );
if( isDetachedSignature )
    /* Detached signature */;
else
    /* Signed data + signature */;
```

Alternative Detached Signature Processing

Besides the method described above there is a second way to verify a detached signature which involves hashing the data yourself and then adding the hash to the envelope rather than pushing the data into the envelope and having it hashed for you. This is useful in situations where the signed data is present separate from the signature, or is in a non-standard format (for example an AuthentiCode signed file) that can't be recognised by the enveloping code.

Verifying a detached signature in this manner is a slight variation of the standard detached signature verification process in which you first add to the envelope the hash value for the signed data and then push in the detached signature:

```
CRYPT_CONTEXT hashContext;
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied, sigCheckStatus;

/* Create the hash context and hash the signed data */
cryptCreateContext( &hashContext, cryptUser, CRYPT_ALGO_SHA );
cryptEncrypt( hashContext, signedData, dataLength );
cryptEncrypt( hashContext, signedData, 0 );

/* Create the envelope and add the hash */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_HASH, hashContext );
cryptDestroyContext( hashContext );

/* Add the detached signature */
cryptPushData( cryptEnvelope, signatureData, signatureDataLength,
               &bytesCopied );
cryptFlushData( cryptEnvelope );

/* Determine the result of the signature check */
cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_RESULT,
                  &sigCheckStatus );
```

```
cryptDestroyEnvelope( cryptEnvelope );
```

When you push in the detached signature cryptlib will verify that the hash information in the signature matches the hash that you've supplied. If the two don't match, cryptlib will return `CRYPT_ERROR_SIGNATURE` to indicate that the signature can't be verified using the given values. Because of this check, you must add the hash before you push in the detached signature.

Extra Signature Information

S/MIME signatures can include with them extra information such as the time at which the message was signed. Normally cryptlib will add and verify this information for you automatically, with the details of what's added based on the setting of the `CRYPT_OPTION_CMS_DEFAULTATTRIBUTES` option as described in "Working with Configuration Options" on page 359. If this option is set to false (zero), cryptlib won't add any additional signature information, which minimises the size of the resulting signature. If this option is set to true (any nonzero value), cryptlib will add default signing attributes such as the signing time for you.

You can also handle the extra signing information yourself if you require extra control over what's included with the signature. The extra information is specified as a `CRYPT_CERTTYPE_CMS_ATTRIBUTES` certificate object. To include this information with the signature you should add it to the envelope alongside the signing key as `CRYPT_ENVINFO_SIGNATURE_EXTRADATA`:

```
CRYPT_ENVELOPE cryptEnvelope;
CRYPT_CERTIFICATE cmsAttributes;

/* Create the CMS attribute object */
cryptCreateCert( &cmsAttributes, cryptUser,
    CRYPT_CERTTYPE_CMS_ATTRIBUTES );
/* ... */

/* Create the envelope and add the signing key and signature
   information */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_CMS );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
    sigKeyContext );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_EXTRADATA,
    cmsAttributes );
cryptDestroyCert( cmsAttributes );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

You can also use this facility to extend or overwrite the attributes added by cryptlib. For example if you wanted to add a security label to the data being signed, you would add it to the CMS attribute object and add that to the envelope. cryptlib will then add any additional required information (for example the signing time) and finally generate the signature using the combined collection of attributes. This means that you can fill in whatever attributes you want, and cryptlib will take care of the rest for you.

Verifying a signature that includes this extra information works just like standard signature verification since cryptlib handles it all for you. Just as you can obtain a certificate chain from a signature, you can also obtain the extra signature information from the envelope:

```
CRYPT_CERTIFICATE cmsAttributes;

cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_EXTRADATA,
    &cmsAttributes );
```

You can now work with the signing attributes as in the same manner as standard certificate attributes, for example you may want to display any relevant information to the user. More details on working with these attributes are given in “Certificate Extensions” on page 320, and the attributes themselves are covered in “Other Certificate Object Extensions” on page 338.

The example above created a `CRYPT_FORMAT_CMS` envelope, which means that cryptlib will add certain default signing attributes to the signature when it creates it. If the envelope is created with `CRYPT_FORMAT_SMIME` instead of `CRYPT_FORMAT_CMS`, cryptlib will add an extra set of S/MIME-specific attributes that indicate the preferred encryption algorithms for use when an S/MIME enabled mailer is used to send mail to the signer. This information is used for backwards-compatibility reasons because many S/MIME mailers will quietly default to using very weak 40-bit keys if they’re not explicitly told to use proper encryption such as triple DES or AES (cryptlib will never use weakened encryption since it doesn’t even provide this capability).

Because of this default-to-insecure encryption problem, cryptlib includes with a `CRYPT_FORMAT_SMIME` signature additional information to indicate that the sender should use a non-weakened algorithm such as triple DES, AES, CAST-128, or IDEA. With a `CRYPT_FORMAT_CMS` signature this additional S/MIME-specific information isn’t needed so cryptlib doesn’t include it.

Timestamping

In addition to the standard signature information which is provided by the signer, cryptlib also supports the use of a message timestamp which is provided by an external timestamp authority (TSA). Timestamping signed data in an envelope is very simple and requires only the addition of a `CRYPT_ENVINFO_TIMESTAMP` attribute to tell cryptlib which TSA to obtain the timestamp from. The TSA is specified as a TSP session object as described in “Secure Sessions” on page 190. For example to specify a TSA located at `http://www.timestamp.com/-tsa/request.cgi`, you would create the TSP session with:

```
CRYPT_SESSION cryptSession;

/* Create the TSP session and add the server name */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_TSP );
cryptSetAttributeString( cryptSession, CRYPT_SESSION_SERVER_NAME,
    "http://www.timestamp.com/-tsa/request.cgi", 40 );
```

You can also specify additional session information in the usual manner for cryptlib sessions, after which you add the session to the envelope. Once you’ve added it, you can destroy it since it’s now managed by the envelope:

```
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_TIMESTAMP,
    cryptSession );
cryptDestroySession( cryptSession );
```

When cryptlib signs the data in the envelope, it will communicate with the TSA to obtain a timestamp on the signature, which is then included with the other signed data. This timestamp can be verified at a later date to prove that the envelope was indeed signed at the indicated time.

Since communicating with a TSA over a network can be a slow process, the signature generation may take somewhat longer than usual. When the timestamp is created cryptlib doesn’t communicate any part of the message or any indication of its contents to the TSA, it merely sends it the message signature information which is then countersigned by the TSA. In this way no confidential or sensitive information is leaked to the outside world through the timestamping process.

A time-stamped message appears the same as a standard signed message, with the exception that the timestamp data is present as additional signature information of type `CRYPT_ENVINFO_TIMESTAMP`. You can read the timestamp data in the same way that you read other extra signature information:


```
CRYPT_ENVELOPE timeStamp;  
  
cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_TIMESTAMP,  
                  &timeStamp );
```

The returned timestamp is a standard signed envelope object that you can check in the usual manner, for example by verifying the signature on the timestamp data and checking the certificates used for the timestamp signature.

PGP

PGP is a standard format for encrypting, signing, and compressing data. The original format, PGP 2.x or PGP classic, has since been superseded by OpenPGP, partially implemented in PGP 5.0 and later fully in NAI PGP, GPG, and various variations such as the ckt builds. cryptlib can read both the PGP 2.x and OpenPGP formats, including handling for assorted variations and peculiarities of different implementations. As output cryptlib produces data in the OpenPGP format, which can be read by any recent PGP implementation. Note that PGP 2.x used the patented IDEA encryption algorithm (see “Algorithms” on page 387 for details), if you’re using the code for commercial purposes you need to either obtain a license for IDEA or use only the OpenPGP format (which cryptlib does by default anyway, so this usually isn’t a concern).

You can specify the use of the PGP format when you create an envelope with the formatting specifier `CRYPT_FORMAT_PGP`, which tells cryptlib to use the PGP format rather than the (default) CMS format. cryptlib doesn’t restrict the use of PGP envelopes to PGP keys. Any type of keys, including standard cryptlib keys and X.509 certificates, can be used with PGP envelopes. By extension it’s also possible to use smart cards, crypto accelerators, and Fortezza cards with PGP envelopes (as an extreme example, it’s possible to use a Fortezza card to create a PGP envelope).

PGP Enveloping

To create an envelope that uses the PGP format, call **cryptCreateEnvelope** as usual but specify a format type of `CRYPT_FORMAT_PGP` instead of the usual `CRYPT_FORMAT_CRYPTLIB`:

```
CRYPT_ENVELOPE cryptEnvelope;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_PGP );

/* Perform enveloping */

cryptDestroyEnvelope( cryptEnvelope );
```

Creating the envelope in this way restricts cryptlib to using the PGP data format instead of the more flexible data format which is used for envelopes by default. This imposes a number of restrictions on the use of envelopes that are described in more detail in the sections that cover individual PGP enveloping types. One restriction that applies to all enveloping types is that PGP requires the presence of the `CRYPT_ENVINFO_DATASIZE` attribute before data can be enveloped. This attribute is described in more detail in “Data Size Considerations” on page 146. If you try to push data into an envelope without setting the `CRYPT_ENVINFO_DATASIZE` attribute, cryptlib will return `CRYPT_ERROR_NOTINITED` to indicate that you haven’t provided the information which is needed for the enveloping to proceed.

Encrypted Enveloping

PGP supports password-based enveloping in the same general way as ordinary cryptlib envelopes. However, due to constraints imposed by the PGP format, it’s not possible to mix password- and public-key-based key exchange actions in the same envelope. In addition it’s not possible to specify more than one password for an envelope. If you try to add more than one password, or try to add a password when you’ve already added a public key or vice versa, cryptlib will return `CRYPT_ERROR_INITED` to indicate that the key exchange action has already been set.

Public-key based PGP enveloping works the same way as standard cryptlib enveloping. For example to encrypt data using the a public key you would use:

```

CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_PGP );

/* Add the public key */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_PUBLICKEY,
    publicKey );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );

```

Since the key will originally have come from a keyset, a simpler alternative to reading the key yourself and explicitly adding it to the envelope is to let cryptlib do it for you by first adding the keyset to the envelope and then specifying the email address of the recipient or recipients of the message with the CRYPT_ENVINFO_RECIPIENT attribute:

```

CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_PGP );

/* Add the encryption keyset and recipient email address */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_ENCRYPT,
    cryptKeyset );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_RECIPIENT,
    "person@company.com", 18 );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );

```

For each message recipient that you add, cryptlib will look up the key in the encryption keyset and add the appropriate information to the envelope to encrypt the message to that person. This is the recommended way of handling public-key encrypted enveloping, since it lets cryptlib handle the key details for you and makes it possible to manage problem areas such as cases where the same email address is present for multiple keys of which only one is valid for message encryption.

De-enveloping works as for standard enveloping:

```

CRYPT_ENVELOPE cryptEnvelope;
CRYPT_ATTRIBUTE_TYPE requiredAttribute;
int bytesCopied, status;

/* Create the envelope and add the private key keyset and data */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_DECRYPT,
    privKeyKeyset );
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
    &bytesCopied );

```

```
/* Find out what we need to continue and, if it's a private key, add
the password to recover it */
cryptGetAttribute( cryptEnvelope, CRYPT_ATTRIBUTE_CURRENT,
&requiredAttribute );
if( requiredAttribute != CRYPT_ENVINFO_PRIVATEKEY )
/* Error */;
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_PASSWORD,
password, passwordLength );
cryptFlushData( cryptEnvelope );

/* Pop the data and clean up */
cryptPopData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptDestroyEnvelope( cryptEnvelope );
```

More information on public-key encrypted enveloping, including its use with crypto devices such as smart cards, is given in “Public-Key Encrypted Enveloping” on page 160.

Digitally Signed Enveloping

PGP digitally signed enveloping works just like standard enveloping. For example if you wanted to sign data using a private key contained in `sigKeyContext`, you would use:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_PGP );

/* Add the signing key */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
sigKeyContext );

/* Add the data size information and data, wrap up the processing, and
pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
&bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

Verifying the signature works in the usual way:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied, signatureResult, status;

/* Create the envelope and add the signature check keyset */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_SIGCHECK,
sigCheckKeyset );

/* Push in the signed data and pop out the recovered message */
cryptPushData( cryptEnvelope, envelopedData, envelopedDataLength,
&bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, message, messageBufferSize,
&bytesCopied );

/* Determine the result of the signature check */
cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_RESULT,
&signatureResult );
```

The signature result will typically be `CRYPT_OK` (the signature verified), `CRYPT_ERROR_SIGNATURE` (the signature did not verify), or `CRYPT_ERROR_NOTFOUND` (the key needed to check the signature wasn’t found in the keyset).

When you sign data in the PGP format, the nested content type is always set to plain data. This is a limitation of the PGP format that always signs data as the innermost step, so that what’s signed is always plain data. In addition to this restriction, it’s not possible to have more than one signer per envelope. Multiple signers requires the use of nested envelopes, however it’s necessary to intersperse a layer of encryption or

compression between each signature pass since PGP can't easily distinguish which signature belongs to which signature pass. In general it's best not to try to apply multiple signatures to a piece of data.

Detached Signatures

So far, the signature for the signed data has always been included with the data itself, allowing it to be processed as a single blob. cryptlib also provides the ability to create detached signatures in which the signature is held separate from the data. This leaves the data being signed unchanged and produces a standalone signature as the result of the encoding process.

To specify that an envelope should produce a detached signature rather than standard signed data, you should set the envelope's CRYPT_ENVINFO_DETACHEDSIGNATURE attribute to 'true' (any nonzero value) before you push in any data

```
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DETACHEDSIGNATURE, 1
);
```

Apart from that, the creation of detached signatures works just like the creation of standard signed data, with the result of the enveloping process being the standalone signature (without the data attached):

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_PGP );

/* Add the signing key and specify that we're using a detached
signature */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE,
sigKeyContext );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DETACHEDSIGNATURE, 1
);

/* Add the data size information and data, wrap up the processing, and
pop out the detached signature */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, detachedSignature,
detachedSignatureBufferSize, &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

Verifying a detached signature works somewhat differently from standard cryptlib detached signature processing since the PGP format doesn't differentiate between standard and detached signatures. Because of this lack of differentiation, it's not possible for cryptlib to automatically determine whether a signature should have data associated with it or not. Normally, cryptlib assumes that a signature is associated with the data being signed, which is the most common case. When verifying a detached signature, you need to use the alternative signature processing technique that involves hashing the data yourself and then adding the hash to the envelope rather than pushing the data into the envelope and having it hashed for you. Since PGP hashes further information after hashing the data to be signed, you shouldn't complete the hashing before you push the hash context into the envelope. This is in contrast to standard cryptlib detached signature processing which requires that you complete the hashing before pushing the context into the envelope:

```
CRYPT_CONTEXT hashContext;
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied, sigCheckStatus;

/* Create the hash context and hash the signed data without completing
the hashing */
cryptCreateContext( &hashContext, cryptUser, CRYPT_ALGO_SHA );
cryptEncrypt( hashContext, data, dataLength );
```

```
/* Create the envelope and add the signature check keyset */
cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_AUTO );
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_SIGCHECK,
    sigCheckKeyset );

/* Add the hash and follow it with the detached signature */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_HASH, hashContext );
cryptPushData( cryptEnvelope, data, dataLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptDestroyContext( hashContext );

/* Determine the result of the signature check */
cryptGetAttribute( cryptEnvelope, CRYPT_ENVINFO_SIGNATURE_RESULT,
    &sigCheckStatus );

cryptDestroyEnvelope( cryptEnvelope );
```

When you push in the detached signature cryptlib will verify that the hash information in the signature matches the hash that you've supplied. If the two don't match, cryptlib will return `CRYPT_ERROR_SIGNATURE` to indicate that the signature can't be verified using the given values. Because of this check, you must add the hash before you push in the detached signature.

From Envelopes to email

The enveloping process produces binary data as output that then needs to be wrapped up in the appropriate MIME headers and formatting before it can really be called S/MIME or PGP mail. The exact mechanisms used depend on the mailer code or software interface to the mail system you're using. General guidelines for the different enveloped data types are given below.

Note that cryptlib is a security toolkit and not a mail client or server. Although cryptlib provides all the crypto functionality needed to implement S/MIME and PGP, it cannot send or receive email, process MIME message parts or base64 or PGP ASCII encoding, or otherwise act as a mail agent. These functions are performed by mail-handling software. For mail-processing operations you need to combine it with mail-handling software of the kind described further on.

S/MIME email

MIME is the Internet standard for communicating complex data types via email, and provides for tagging of message contents and safe encoding of data to allow it to pass over data paths that would otherwise damage or alter the message contents. Each MIME message has a top-level type, subtype, and optional parameters. The top-level types are application, audio, image, message, multipart, text, and video.

Most of the S/MIME secured types have a content type of `application/pkcs7-mime`, except for detached signatures that have a content type of `application/pkcs7-signature`. The content type usually also includes an additional `smime-type` parameter whose value depends on the S/MIME type and is described in further detail below. In addition it's usual to include a content-disposition field whose value is also explained below.

Since MIME messages are commonly transferred via email and this doesn't handle the binary data produced by cryptlib's enveloping, MIME also defines a means of encoding binary data as text. This is known as content-transfer-encoding.

Data

The innermost, plain data content should be converted to canonical MIME format and have a standard MIME header which is appropriate to the data content, with optional encoding as required. For the most common type of content (plain text), the header would have a content-type of `text/plain`, and possibly optional extra information such as a content transfer encoding (in this case `quoted-printable`), content disposition, and whatever other MIME headers are appropriate. This formatting is normally handled for you by the mailer code or software interface to the mail system you're using.

Signed Data

For signed data the MIME type is `application/pkcs7-mime`, the `smime-type` parameter is `signed-data`, and the extensions for filenames specified as parameters is `.p7m`. A typical MIME header for signed data is therefore:

```
Content-Type: application/pkcs7-mime; smime-type=signed-data;
             name=smime.p7m
Content-Transfer-Encoding: base64
Content-Disposition: attachment; filename=smime.p7m
```

encoded signed data

Detached Signature

Detached signatures represent a special instance of signed data in which the data to be signed is carried as one MIME body part and the signature is carried as another body part. The message is encoded as a multipart MIME message with the overall message

having a content type of `multipart/signed` and a protocol parameter of `application/pkcs7-signature`, and the signature part having a content type of `application/pkcs7-signature`.

Since the data precedes the signature, it's useful to include the hash algorithm used for the data as a parameter with the content type (cryptlib processes the signature before the data so it doesn't require it, but other implementations may not be able to do this). The hash algorithm parameter is given by `micalg=sha1` or `micalg=md5` as appropriate. When receiving S/MIME messages you can ignore this value since cryptlib will automatically use the correct type based on the signature.

A typical MIME header for a detached signature is therefore:

```
Content-Type: multipart/signed; protocol=application/pkcs7-signature;
    micalg=sha1; boundary=boundary

--boundary
Content-Type: text/plain Content-Transfer-Encoding: quoted-printable

signed text

--boundary
Content-Type: application/pkcs7-signature; name=smime.p7s
Content-Transfer-Encoding: base64
Content-Disposition: attachment; filename=smime.p7s

encoded signature

--boundary--
```

Encrypted Data

For encrypted data the MIME type is `application/pkcs7-mime`, the `smime-type` parameter is `enveloped-data`, and the extension for filenames specified as parameters is `.p7m`. A typical MIME header for encrypted data is therefore:

```
Content-Type: application/pkcs7-mime; smime-type=enveloped-data;
    name=smime.p7m
Content-Transfer-Encoding: base64
Content-Disposition: attachment; filename=smime.p7m

encoded encrypted data
```

Nested Content

Unlike straight CMS nested content, S/MIME nested content requires a new level of MIME encoding for each nesting level. For the minimum level of nesting (straight signed or encrypted data) you need to first MIME-encode the plain data, then envelope it to create CMS signed or encrypted data, and then MIME-encode it again. For the typical case of signed, encrypted data you need to MIME-encode, sign, MIME-encode again, encrypt, and then MIME-encode yet again (rumours that S/MIME was designed by a consortium of network bandwidth vendors and disk drive manufacturers are probably unfounded).

Since the nesting information is contained in the MIME headers, you don't have to specify the nested content type using `CRYPT_ENVINFO_CONTENTTYPE` as you do with straight CMS enveloped data (this is one of the few actual differences between `CRYPT_FORMAT_CMS` and `CRYPT_FORMAT_SMIME`), cryptlib will automatically set the correct content type for you. Conversely, you need to use the MIME header information rather than `CRYPT_ENVINFO_CONTENTTYPE` when de-enveloping data (this will normally be handled for you by the mailer code or software interface to the mail system you're using).

PGP email

Traditionally, PGP has employed its own email encapsulation format that predates MIME and isn't directly compatible with it. A PGP message is delimited with the string `-----BEGIN PGP MESSAGE-----` and `-----END PGP MESSAGE--`

---, with the (binary) message body present in base64-encoded format between the delimiters. The body is followed by a base64-encoded CRC24 checksum calculated on the message body before base64-encoding. In addition the body may be preceded by one or more lines of type-and-value pairs containing additional information such as software version information, and separated from the body by a blank line. More details on the format are given in the PGP standards documents.

An example of a PGP email message is:

```
-----BEGIN PGP MESSAGE-----
Version: cryptlib 3.1

base64-encoded message body
base64-encoded CRC24 checksum
-----END PGP MESSAGE-----
```

Signed data with a detached signature is delimited with -----BEGIN PGP SIGNED MESSAGE----- at the start of the message, followed by -----BEGIN PGP SIGNATURE----- and -----END PGP SIGNATURE----- around the signature that follows. The signature follows the standard PGP message-encoding rules given above:

```
-----BEGIN PGP SIGNED MESSAGE-----
Hash: SHA1

message body
-----BEGIN PGP SIGNATURE-----
Version: cryptlib 3.1

base64-encoded signature
base64-encoded CRC24 checksum
-----END PGP SIGNATURE-----
```

The example above shows another use for the type-and-value lines, in this case to indicate the hashing algorithm used in the signature to allow one-pass processing of the message.

In addition to the traditional PGP format, there exists a mechanism for encapsulating the traditional PGP format in an additional layer of MIME wrapping. This isn't true MIME message handling since it merely wraps MIME headers around the existing PGP email encapsulation rather than using the full MIME capabilities directly as does S/MIME. This format is almost never used, with software expected to use the traditional PGP format instead. If you need more information about PGP/MIME, you can find it in the PGP standards documentation.

Implementing S/MIME and PGP email using cryptlib

Most of the MIME processing and encoding issues described above will be handled for you by the mail software that cryptlib is used with. To use cryptlib to handle S/MIME and PGP email messages, you would typically register the various MIME types with the mail software and, when they are encountered, the mailer will hand the message content (the data that remains after the MIME wrapper has been removed) to cryptlib. cryptlib can then process the data and hand the processed result back to the mailer. The same applies for generating S/MIME and PGP email messages.

Note that cryptlib is a security toolkit and not a mail client or server. Although cryptlib provides all the crypto functionality needed to implement S/MIME and PGP, it cannot send or receive email, process MIME message parts, or otherwise act as a mail agent. For mail-processing operations you need to combine it with mail-handling software of the kind described below.

c-client/IMAP4

c-client is a portable Swiss army chainsaw interface to a wide variety of mail and news handling systems. One of the services it provides is full handling of MIME message parts which involves breaking a message down into a sequence of BODY structures each of which contains one MIME body part. The `type` member contains the content type (typically `TYPEMULTIPART` or `TYPEAPPLICATION` for the

types used in S/MIME or PGP), the `subtype` member contains the MIME subtype, the `parameter` list contains any required parameters, and the `contents.binary` member contains outgoing binary data straight from the cryptlib envelope (c-client will perform any necessary encoding such as base64 if required). All of this information is converted into an appropriately-formatted MIME message by c-client before transmission.

Since IMAP supports the fetching of individual MIME body parts from a server, `contents.binary` can't be used to access incoming message data since only the header information may have been fetched, with the actual content still residing on the server. To fetch a particular body part, you need to use `mail_fetchbody`. If the body part is base64-encoded (denoted by the `encoding` member of the `BODY` having the value `ENCBASE64`) then you also need to call `rfc822_base64` to decode the data so cryptlib can process it. In the unlikely event that the binary data is encoded as quoted-printable (denoted by `ENCQUOTEDPRINTABLE`, at least one broken mailer occasionally does this) you need to call `rfc822_qpprint` instead. In either case the output can be pushed straight into a cryptlib envelope.

Eudora

Eudora handles MIME content types through plug-in translators that are called through two functions, `ems_can_translate` and `ems_translate_file`. Eudora calls `ems_can_translate` with an `emsMIMETYPE` parameter that contains information on the MIME type contained in the message. If this is an S/MIME or PGP type (for example `application/pkcs7-mime`) the function should return `EMSR_NOW` to indicate that it can process this MIME type, otherwise it returns `EMSR_CANT_TRANSLATE`.

Once the translator has indicated that it can process a message, Eudora calls `ems_translate_file` with input and output files to read the data from and write the processed result to. The translation is just the standard cryptlib enveloping or de-enveloping process depending on whether the translator is an on-arrival or on-display one (used for de-enveloping incoming messages) or a Q4-transmission or Q4-completion one (used for enveloping outgoing messages).

MAPI

MAPI (Microsoft's mail API) defines two types of mailer extensions that allow cryptlib-based S/MIME and PGP functionality to be added to Windows mail applications. The first type is a spooler hook or hook provider, which can be called on delivery of incoming messages and on transmission of outgoing messages. The second type is a preprocessor, which is less useful and operates on outgoing messages only. The major difference between the two in terms of implementation complexity is that hook providers are full (although simple) MAPI service providers while preprocessors are extensions to transport providers (that is, if you've already written a transport provider you can add the preprocessor without too much effort; if you don't have a transport provider available, it's quite a bit more work). In general it's probably easiest to use a single spooler hook to handle inbound and outbound messages. You can do this by setting both the `HOOK_INBOUND` and `HOOK_OUTBOUND` flags in the hook's `PR_RESOURCE_FLAGS` value.

Messages are passed to hooks via `ISpoolerHook::OutboundMsgHook` (for outgoing messages) and `ISpoolerHook::InboundMsgHook` (for incoming messages). The hook implementation itself is contained in a DLL that contains the `HPPProviderInit` entry point and optional further entry points used to configure it, for example a message service entry point for program-based configuration and a `WIZARDENTRY` for user-based configuration.

Windows 95/98/ME and NT/2000/XP/Vista Shell

Windows allows a given MIME content type to be associated with an application to process it. You can set up this association by calling `MIMEAssociationDialog` and setting the `MIMEASSOCDLG_FL_REGISTER_ASSOC` flag in the

`dwInFlags` parameter, which will (provided the user approves it) create an association between the content type you specify in the `pcszMIMEContentType` parameter and the application chosen by the user. This provides a somewhat crude but easy to set up mechanism for processing S/MIME and PGP data using a cryptlib-based application.

Secure Sessions

cryptlib's secure session interface provides a session-oriented equivalent to envelope objects that can be used to secure a communications link with a host or server or otherwise communicate with another system over a network. Secure sessions can include SSH, SSL, and TLS sessions, general request/response-style communications sessions can include protocols such as the certificate management protocol (CMP), simple certificate enrolment protocol (SCEP), real-time certificate status protocol (RTCS), online certificate status protocol (OCSP), and timestamping protocol (TSP). As with envelopes, cryptlib takes care of all of the session details for you so that all you need to do is provide basic communications information such as the name of the server or host to connect to and any other information required for the session such as a password or certificate. cryptlib takes care of establishing the session and managing the details of the communications channel and its security parameters.

Secure sessions are very similar to envelopes, with the main difference being that while an envelope is a pure data object into which you can push data and pop the processed form of the same data, a session is a communications object into which you push data and then pop data that constitutes a response from a remote server or client. This means that a session object can be viewed as a bottomless envelope through which you can push or pop as much data as the other side can accept or provide.

As with an envelope, you use a session object by adding to it action objects and resources such as user names and passwords that control the interaction with the remote server or client and then push in data intended for the remote system and pop out data coming from the remote system. For example to connect to a server using SSH and obtain a directory of files using the `ls` command you would do the following:

```
create the session;
add the server name, user name, and password;
activate the session;
push data "ls";
pop the result of the ls command;
destroy the session
```

That's all that's necessary. Since you've added a user name and password, cryptlib knows that it should establish an encrypted session with the remote server and log on using the given user name and password. From then on all data which is exchanged with the server is encrypted and authenticated using the SSH protocol.

Creating an SSH server session is equally simple. In this case all you need is the server key:

```
create the session;
add the server key;
activate the session;
pop client data;
push server response;
destroy the session
```

When you activate the session, cryptlib will listen for an incoming connection from a client and return once a secure connection has been negotiated, at which point communication proceeds as before.

Creating/Destroying Session Objects

Secure sessions are accessed as session objects that work in the same general manner as other cryptlib objects. You create a session using **cryptCreateSession**, specifying the user who is to own the session object or `CRYPT_UNUSED` for the default, normal user, and the type of session that you want to create. This creates a session object ready for use in securing a communications link or otherwise communicating with a remote server or client. Once you've finished with the session, you use **cryptDestroySession** to end the session and destroy the session object:

```

CRYPT_SESSION cryptSession;

cryptCreateSession( &cryptSession, cryptUser, sessionType );

/* Communicate with the remote server or client */

cryptDestroySession( cryptSession );

```

The available session types are:

Session	Description
CRYPT_SESSION_CMP	Certificate management protocol (CMP).
CRYPT_SESSION_OCSP	Online certificate status protocol (OCSP).
CRYPT_SESSION_RTCS	Real-time certificate status protocol (RTCS).
CRYPT_SESSION_SCEP	Simple certificate enrolment protocol (SCEP).
CRYPT_SESSION_SSH	Secure shell (SSH).
CRYPT_SESSION_SSL	Secure sockets layer (SSL and TLS).
CRYPT_SESSION_TSP	Timestamping protocol (TSP).

This section will mainly cover the secure communications session types such as SSH, SSL, and TLS. CMP, SCEP, RTCS, and OCSP client sessions are certificate management services that are covered in “Obtaining Certificates using CMP”, “Obtaining Certificates using SCEP”, “Certificate Status Checking using RTCS”, and “Certificate Revocation Checking using OCSP” on pages 252, 247, 247, and 253, and a TSP client session is an S/MIME service which is covered in “Timestamping” on page 178. RTCS, OCSP and TSP server sessions are standard session types and are also covered here. CMP and SCEP server sessions are somewhat more complex and are covered in “Managing a CA using CMP or SCEP” on page 261. The general principles covering sessions apply to all of these session types, so you should familiarise yourself with the operation of session objects and associated issues such as network proxies and timeouts before trying to work with these other session types.

By default the secure communications session object which is created will have an internal buffer whose size is appropriate for the type of security protocol which is being employed. The size of the buffer may affect the amount of extra processing that cryptlib needs to perform, so that a large buffer can reduce the amount of copying to and from the buffer, but will consume more memory. If want to use a buffer for a secure communications session which is larger than the default size, you can specify its size using the CRYPT_ATTRIBUTE_BUFFERSIZE attribute after you’ve created the session. For example if you wanted to set the buffer for an SSH session to 64 kB you would use:

```

CRYPT_SESSION cryptSession;

cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSH );
cryptSetAttribute( cryptSession, CRYPT_ATTRIBUTE_BUFFERSIZE, 65536L );

/* Communicate with the remote server or client */

cryptDestroySession( cryptSession );

```

Since cryptlib streams data through the session object, the internal buffer size doesn’t limit how much data you can push and pop (for example you could push 1 MB of data into a session object with a 32 kB internal buffer), the only reason you’d want to change the size is to provide tighter control over memory usage by session objects. Unless you’re absolutely certain that the other side will only send very small data quantities, you shouldn’t shrink the buffer below the default size set by cryptlib since the protocols that cryptlib implements have certain fixed bounds on packet sizes that need to be met, making the buffer too small would make it impossible to process data being sent by the other side.

Note that the CRYPT_SESSION is passed to **cryptCreateSession** by reference as the function modifies it when it creates the session. In all other routines in cryptlib, CRYPT_SESSION is passed by value.

Client vs. Server Sessions

cryptlib distinguishes between two types of session objects, client sessions and server sessions. Client sessions establish a connection to a remote server while server sessions wait for incoming communications from a remote client. To distinguish between client and server objects, you use a session type ending in _SERVER when you create the session object. For example to create an SSL/TLS server object instead of an SSL/TLS client you would specify its type on creation as CRYPT_SESSION_SSL_SERVER instead of CRYPT_SESSION_SSL.

Because server sessions wait for an incoming connection request to arrive, you need to run each one in its own thread if you want to handle multiple connections simultaneously (cryptlib is fully thread-safe so there's no problem with having multiple threads processing incoming connections). For example to handle up to 10 connections at once you would do the following:

```
for i = 1 to 10 do
    start_thread( server_thread );
```

where the server_thread is:

```
loop
    create the session;
    add required information to the session;
    activate the session;
    process client request(s);
    destroy the session;
```

More information on using cryptlib with multiple threads is given in "Multi-threaded cryptlib Operation" on page 46.

Binding to the default ports used by the various session protocols may require special privileges on some systems that don't allow normal users to bind to ports below 1024. If you need to bind to a reserved port you should consult your operating system's documentation for details on any restrictions that may apply, and may need to take special precautions if binding to one of these ports requires the use of elevated security privileges. Alternatively, you can bind to a non-default port outside the reserved range by specifying the port using the CRYPT_SESSINFO_SERVER_PORT attribute. You can also specify which interface you want to bind to if the system has more than one by using the CRYPT_SESSINFO_SERVER_NAME attribute. If you're testing code before deploying it, it's a good idea to specify that you want to bind to **localhost** to avoid listening on arbitrary externally-visible interfaces. For example to listen on local port 2000 you would use:

```
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    "localhost", 9 );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_SERVER_PORT, 2000 );
```

Server Names/URLs

Server names can be given using IP addresses (in dotted-decimal form for IPv4 or colon-delimited form for IPv6), DNS names, or full URLs, with optional ports and other information provided in the usual manner. You can specify the server name or URL using the CRYPT_SESSINFO_SERVER_NAME attribute and the port (if you're not using the default port from the protocol and it isn't already specified in the URL) using the CRYPT_SESSINFO_SERVER_PORT attribute. For example to specify a connection to the server **www.server.com** on port 80 you would use:

```
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    "www.server.com", 14 );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_SERVER_PORT, 80 );
```

Alternatively, you could specify both in the same name:

```
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    "www.server.com:80", 17 );
```

Since this is a web server for which port 80 is the default port, you could also use the more common:

```
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    "http://www.server.com", 20 );
```

SSL and TLS use a predefined port and are often used in conjunction with HTTP, so you can specify these URLs with or without the `http://` or `https://` schema prefixes. SSH similarly uses a predefined port and can be used with or without the `ssh://`, `scp://`, or `sftp://` schema prefixes. All of these protocols allow you to specify user information before the host name, separated with an '@' sign. For example to connect as "user" to the SSH server `ssh.server.com` you could use:

```
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    "ssh://user@ssh.server.com", 25 );
```

which saves having to explicitly specify the user name with the `CRYPT_SESSINFO_USERNAME` attribute.

All of the PKI protocols use HTTP as their transport mechanism, so cryptlib will automatically default to using HTTP transport whether you include the `http://` schema specifier or not. The CMP and TSP protocols also have alternative, deprecated transport mechanisms identified by `cmp://...` (for CMP) and `tcp://...` (for TSP) instead of `http://...`. These are occasionally used by CAs or timestamp servers, you may need to use these instead of the HTTP default.

Server Private Keys

Most server sessions require the use of a private key in one form or another to decrypt data from the client or sign responses returned to the client. The server key is typically stored in a private key file, but for extra security may be held in a crypto device such as a crypto coprocessor or accelerator. In addition, for most session types the server key needs to be associated with a certificate or certificate chain leading up to a trusted root certificate, so that you can't use just a raw private key as the server key. You can obtain the required certificate or certificate chain by creating it yourself using cryptlib or by obtaining it from a commercial CA (it's generally much cheaper and easier to create it yourself than to obtain one from a third-party CA).

When you create or obtain the certificate for your server, you may need to specify the server name in the common name field of the certificate (how to create your own certificate is explained in "Certificates and Certificate Management" on page 234). For example if your server was `www.companyname.com` then the certificate for the server would contain this as its common name component (you can actually put in anything you like as the common name component, but this will result in some web browsers that use your server displaying a warning message when they connect).

SSH server sessions require a raw RSA (or optionally DSA for SSHv2) key, although you can also use one with a certificate or certificate chain attached. All other session types require one with certificate(s) attached. You add the server key as the `CRYPT_SESSINFO_PRIVATEKEY` attribute, for example to use a private key held in a crypto device as the server key you would use:

```
CRYPT_CONTEXT privateKey;

cryptGetPrivateKey( cryptDevice, &privateKey, CRYPT_KEYID_NAME,
    serverKeyName, NULL );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY,
    privateKey );
cryptDestroyContext( privateKey );
```

Note that, as with envelopes, the private key object can be destroyed as soon as it's added to the session, since the session maintains its own copy of the object internally.

If you're worried about some obscure (and rather unlikely) attacks on private keys, you can enable the `CRYPT_OPTION_MISC_SIDECHANNELPROTECTION` option as explained in "Working with Configuration Options" on page 359.

Establishing a Session

Much of the secure session process is identical to the enveloping process, so you should familiarise yourself with the general concept of enveloping as described in “Data Enveloping” on page 143 if you haven’t already done so. The secure session establishment process involves adding the information which is required to connect to the remote server as a client or to establish a server, and then activating the session to establish the secure session or wait for incoming connections. This process of activating the session has no real equivalent for envelopes because envelopes are activated automatically the first time data is pushed into them.

Client sessions can also be activated automatically, however the initial handshake process which is required to activate a session with a remote server is usually lengthy and complex so it’s generally better to explicitly activate the session under controlled conditions and have the ability to react to errors in an appropriate manner rather than to have the session auto-activate itself the first time that data is pushed. Server sessions that wait for an incoming connection must be explicitly activated, which causes them to wait for a client connection.

You can activate a session by setting its `CRYPT_SESSION_ACTIVE` attribute to true (any nonzero value). You can also determine the activation state of a session by reading this attribute, if it’s set to true then the session is active, otherwise it’s not active.

Persistent Connections

Some cryptlib session types such as CMP, SCEP, RTCS, OCSP, and TSP provide request/response protocols rather than continuous secure sessions like SSH and SSL/TLS. In many cases it’s possible to perform more than one request/response transaction per session, avoiding the overhead of creating a new connection for each transaction. To handle persistent connections for client sessions, cryptlib uses the `CRYPT_SESSION_CONNECTIONACTIVE` attribute to indicate that the connection is still active and is ready to accept further transactions. Transactions after the initial one are handled in exactly the same way as the first one, except that there’s no need to create a new session object for them:

```
CRYPT_SESSION cryptSession;
int connectionActive;

/* Create the session and add the server name */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_xxx );
cryptSetAttributeString( cryptSession, CRYPT_SESSION_SERVER_NAME,
    serverName, serverNameLength );

/* Perform the first transaction */
cryptSetAttribute( cryptSession, CRYPT_SESSION_REQUEST,
    cryptRequest1 );
cryptSetAttribute( cryptSession, CRYPT_SESSION_ACTIVE, 1 );
cryptGetAttribute( cryptSession, CRYPT_SESSION_RESPONSE,
    &cryptResponse1 );

/* Check whether the session connection is still open */
cryptGetAttribute( cryptSession, CRYPT_SESSION_CONNECTIONACTIVE,
    &connectionActive );
if( !connectionActive )
    /* The other side has closed the connection, exit */;

/* Perform the second transaction */
cryptSetAttribute( cryptSession, CRYPT_SESSION_REQUEST,
    cryptRequest2 );
cryptSetAttribute( cryptSession, CRYPT_SESSION_ACTIVE, 1 );
cryptGetAttribute( cryptSession, CRYPT_SESSION_RESPONSE,
    &cryptResponse2 );
```

Note the check of the `CRYPT_SESSION_CONNECTIONACTIVE` attribute. Since not all servers support persistent connections or may time out and close the connection after a period of inactivity, it’s a good idea to check that the connection is still open before trying to submit further transactions. Note also that there’s no need to explicitly delete the request from the first activation of the session, cryptlib

automatically does this for you once the session activation has completed. This does mean, however, that if you want to repeat the session transaction using the same data as before (which would be somewhat unusual), you need to re-add the request to the session, since the previous activation will have cleared it in preparation for the next activation.

The process on the server side is similar, after a successfully-completed client transaction you can either destroy the session or, if you want to support persistent connections, recycle the connection as for the client-side example above.

SSH Sessions

SSH is a secure data transfer protocol that provides confidentiality, integrity-protection, protection against replay attacks, and a variety of other services. The SSH server is authenticated via the server's public key and the client is authenticated either via a user name and password or (less frequently) a public key-based digital signature. cryptlib supports both versions 1 and 2 of the SSH protocol, although the obsolete version 1 is disabled by default.

The SSH protocol exhibits a design flaw (informally known as the SSH performance handbrake) that can lead to poor performance when transferring data, which is particularly noticeable with applications such as SFTP. Although cryptlib avoids the handbrake, many other implementations don't, restricting data transfer rates to as little as one tenth of the network link speed (the actual slowdown depends on the link characteristics and varies from one situation to another). In order to obtain the maximum performance from SSH, you need to either use cryptlib at both ends of the link (that is, both the client and server must be ones that avoid the performance handbrake), or use another protocol like SSL that doesn't have the handbrake.

SSH Client Sessions

Establishing a session with an SSH server requires adding the server name or IP address, an optional port number if it isn't using the standard SSH port, and the user name and password which is needed to log on to the server via the CRYPT_SESSINFO_USERNAME and CRYPT_SESSINFO_PASSWORD attributes (occasionally the server will use public-key based authentication instead of a password, which is covered later). Once you've added this information, you can activate the session and establish the connection:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSH );

/* Add the server name, user name, and password */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    serverName, serverNameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    username, usernameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password, passwordLength );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

The equivalent operation in Java or C# is:

```
/* Create the session */
int cryptSession = crypt.CreateSession( cryptUser,
    crypt.SESSION_SSH );

/* Add the server name, user name, and password */
crypt.SetAttributeString( cryptSession, crypt.SESSINFO_SERVER_NAME,
    serverName );
crypt.SetAttributeString( cryptSession, crypt.SESSINFO_USERNAME,
    username );
crypt.SetAttributeString( cryptSession, crypt.SESSINFO_PASSWORD,
    password );
```

```
/* Activate the session */  
crypt.SetAttribute( cryptSession, crypt.SESSINFO_ACTIVE, 1 );
```

In Visual Basic this is:

```
Dim cryptSession As Long  
  
' Create the session  
cryptCreateSession cryptSession, cryptUser, CRYPT_SESSION_SSH  
  
' Add the server name, user name, and password  
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_SERVER_NAME, _  
    serverName, Len( serverName )  
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_USERNAME, _  
    userName, Len( userName )  
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_PASSWORD, _  
    password, Len( password )  
  
' Activate the session  
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1
```

When it connects, cryptlib will automatically negotiate the highest protocol version supported by the server and use that to secure the session. You can determine which version is in use once the session has been established by reading the CRYPT_SESSINFO_VERSION attribute, a value of 1 indicates SSH version 1 and a value of 2 indicates SSH version 2. You can also force the use of a particular version (typically you'd want to ensure that SSHv2 is used) by setting the protocol version attribute before you activate the connection.

Activating a session results in cryptlib performing a lot of work in the background. For example when activating the SSH session shown above cryptlib will connect to the remote host, read the host and server keys used for authentication and encryption, generate a secret data value to exchange with the host using its host and server keys, create the appropriate encryption contexts and load keys based on the secret data value into them, negotiate general session parameters, and log on over the encrypted link using the given user name and password.

If the server that you're connecting to requires public-key authentication instead of password authentication, you need to provide a private key via the CRYPT_SESSINFO_PRIVATEKEY attribute to authenticate yourself to the server before you activate the session. The private key could be a native cryptlib key, but it could also be a key from a crypto device such as a smart card or Fortezza card:

```
CRYPT_SESSION cryptSession;  
  
/* Create the session */  
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSH );  
  
/* Add the server name, user name, and client key and activate the  
session */  
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,  
    serverName, serverNameLength );  
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,  
    username, usernameLength );  
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY,  
    cryptPrivateKey );  
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

When cryptlib connects to the server, it will use the provided private key as part of the SSH handshake to authenticate the client to the server, with the private key taking the place of the more usual password. If you're not sure which of the two options you need, you can provide both and cryptlib will use the appropriate one when it connects to the server.

SSH Server Sessions

Establishing an SSH server session requires specifying that the session is a server session and adding the SSH server key. Once you've added this information you can activate the session and wait for incoming connections:

```

CRYPT_SESSION cryptSession;
int bytesCopied;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_SSH_SERVER );

/* Add the server key and activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY,
    privateKey );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );

/* Process any remaining control messages from the client */
cryptPopData( cryptSession, buffer, bufferSize, &bytesCopied );

```

The Visual Basic form is:

```

Dim cryptSession As Long
Dim bytesCopied as Long

' Create the session
cryptCreateSession cryptSession, cryptUser, _
    CRYPT_SESSION_SSH_SERVER

' Add the server key and activate the session
cryptSetAttribute cryptSession, CRYPT_SESSINFO_PRIVATEKEY, privateKey
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1

' Process any remaining control messages from the client
cryptPopData cryptSession, buffer, bufferSize, bytesCopied

```

Note the use of the data pop call after the activation has been completed. SSH clients often send additional session control information such as channel requests or port forwarding information after the session has been activated. Telling cryptlib to try and read any additional messages that may have arrived from the client allows it to process these requests and respond to them as appropriate. In particular, your server shouldn't send data to the client immediately after the session has been established without first performing a data pop to respond to client requests, since the client may interpret the data that you send as an (incorrect) response to its request.

cryptlib supports both SSH version 1 and 2 (although the obsolete version 1 is disabled by default) and by default will function as a version 2 server. If you want to use the (obsolete) SSH version 1 protocol, you need to enable SSHv1 in the build and then set the CRYPT_SESSINFO_VERSION attribute to 1 to have the server respond as a version 1 rather than version 2 server.

Once you activate the session, cryptlib will block until an incoming client connection arrives, at which point it will negotiate a secure connection with the client. When the client connects, cryptlib will ask for a user name and password before it allows the connection to proceed. The handling of the user authentication process is controlled by the CRYPT_SESSINFO_AUTHRESPONSE attribute, by default cryptlib will return a CRYPT_ENVELOPE_RESOURCE status when it receives the user name and password, allowing you to verify the information before continuing. If it's valid, you should set the CRYPT_SESSINFO_AUTHRESPONSE attribute to true and resume the session activation by setting the CRYPT_SESSINFO_ACTIVE response to true again. If not, you can either set the CRYPT_SESSINFO_AUTHRESPONSE attribute to false and resume the session activation (which will give the user another chance to authenticate themselves), or close the session:

```

int status;

status = cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
if( status == CRYPT_ENVELOPE_RESOURCE )
{
    char username[ CRYPT_MAX_TEXTSIZE + 1 ];
    char password[ CRYPT_MAX_TEXTSIZE + 1 ];
    int usernameLength, passwordLength

```

```
/* Get the user name and password */
cryptGetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    username, &usernameLength );
cryptGetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password, &passwordLength );
username[ usernameLength ] = '\0';
password[ passwordLength ] = '\0';

/* Check the user details and allow or deny the response as
   appropriate */
if( checkUser( username, password ) )
    cryptSetAttribute( cryptSession, CRYPT_SESSINFO_AUTHRESPONSE,
        1 );
else
    cryptSetAttribute( cryptSession, CRYPT_SESSINFO_AUTHRESPONSE,
        0 );

/* Resume the session activation, sending the authentication
   response to the client and completing the handshake */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
}
```

To give the user the traditional three attempts at getting their name and password right, you would run the session activation code in a loop:

```
int status;

for( i = 0; i < 3; i++ )
{
    status = cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE,
        1 );
    if( cryptStatusOK( status ) )
        break;          /* User authenticated, exit */
    if( status == CRYPT_ENVELOPE_RESOURCE )
        /* Perform password check as before */;
    else
        break;          /* Some other type of error, exit */
}
```

Alternatively, you can set the `CRYPT_SESSINFO_AUTHRESPONSE` attribute to true before you activate the session and cryptlib will automatically allow the access and complete the activation, so you'll never need to handle the `CRYPT_ENVELOPE_RESOURCE` response. In this case you need to check the user details after the session has been activated and shut it down if the authorisation check fails.

SSH Channels

By default, cryptlib provides the most frequently-used SSH service, a direct encrypted connection from client to server. When you establish the SSH connection, cryptlib creates an SSH communications channel that's used to exchange data. This process is entirely transparent, and you don't have to worry about it if you don't want to — just treat the SSH session as a secure data pipe from one system to another.

There are however cases where you may need to explicitly deal with SSH channels, and that's when you're using special-purpose SSH facilities such as port forwarding, subsystems, or even user-defined channel types. In this case you need to explicitly create the special-purpose channel and add information describing its use before the channel can be activated. This process consists of three steps, creating the channel using the `CRYPT_SESSINFO_SSH_CHANNEL` attribute, specifying its type using the `CRYPT_SESSINFO_SSH_CHANNEL_TYPE` attribute, and finally specifying any optional channel arguments using the `CRYPT_SESSINFO_SSH_CHANNEL_ARG1` and `CRYPT_SESSINFO_SSH_CHANNEL_ARG2` attributes. For example to create a channel of the default type (which is normally done automatically by cryptlib, and that has no optional arguments) you would use:

```
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_SSH_CHANNEL,
    CRYPT_UNUSED );
cryptSetAttributeString( cryptSession,
    CRYPT_SESSINFO_SSH_CHANNEL_TYPE, "session", 7 );
```

Setting the `CRYPT_SESSINFO_SSH_CHANNEL` attribute to `CRYPT_UNUSED` tells cryptlib to create a new channel (rather than trying to select an existing one,

which is what the attribute is normally used for), and the `CRYPT_SESSINFO_SSH_CHANNEL_TYPE` attribute specifies its type. Once you've created a new channel in this manner you can read back the `CRYPT_SESSINFO_SSH_CHANNEL` attribute to get the channel ID that was assigned for the newly-created channel:

```
int channelID;

cryptGetAttribute( cryptSession, CRYPT_SESSINFO_SSH_CHANNEL,
                  &channelID );
```

This value is used to uniquely identify a particular channel, but it's only needed in the presence of multiple channels, which are described in "SSH Multiple Channels" on page 201.

On the server side, reading the details of a channel that's been opened by the client works similarly:

```
char channelType[ CRYPT_MAX_TEXTSIZE + 1 ];
char channelArg1[ CRYPT_MAX_TEXTSIZE + 1 ];
int channelID, channelTypeLength, channelArg1Length, status;

/* Get the channel ID and type */
cryptGetAttribute( cryptSession, CRYPT_SESSINFO_SSH_CHANNEL,
                  &channelID );
cryptGetAttributeString( cryptSession,
                        CRYPT_SESSINFO_SSH_CHANNEL_TYPE, channelType, &channelTypeLength );
channelType[ channelTypeLength ] = '\0';

/* Get the optional channel argument */
status = cryptGetAttributeString( cryptSession,
                                CRYPT_SESSINFO_SSH_CHANNEL_ARG1, channelArg1, &channelArg1Length );
if( cryptStatusOK( status ) )
    channelArg1[ channelArg1Length ] = '\0';
```

If you don't specify otherwise, cryptlib will open a channel of the default type when it connects. If you want to instead use a special-purpose SSH facility, you should provide the information necessary for creating it before you activate the connection. You can also open further channels after the connection has been completed, the process is described in "SSH Multiple Channels" on page 201. If you try to specify the use of more than one channel before the session has been activated, cryptlib will return `CRYPT_ERROR_INITED` when you try to create any channel after the first one, since it's only possible to request further channels once the initial channel has been successfully established.

SSH Subsystems

Alongside the default encrypted link service, SSH provides additional services such as SFTP, an application-level file transfer protocol tunnelled over the SSH link via a subsystem channel. If you plan to use SFTP, note the comment about the SSH performance handbrake at the start of this section. Although cryptlib avoids this problem, non-cryptlib implementations frequently don't, so that the performance of SFTP can be quite poor (as much as ten times slower than the network link speed) in some cases.

You can specify the use of a subsystem by setting the channel type to "subsystem" and the first channel argument to the subsystem name, in this case "sftp":

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSH );

/* Add the server name, user name, and password */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
                        serverName, serverNameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
                        username, usernameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
                        password, passwordLength );
```

```
/* Request the creation of the subsystem channel */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_SSH_CHANNEL,
    CRYPT_UNUSED );
cryptSetAttributeString( cryptSession,
    CRYPT_SESSINFO_SSH_CHANNEL_TYPE, "subsystem", 9 );
cryptSetAttributeString( cryptSession,
    CRYPT_SESSINFO_SSH_CHANNEL_ARG1, "sftp", 4 );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

Note that SFTP is not a part of the SSH protocol (it can also be run over SSL or IPsec, or directly over raw sockets), but simply an RPC mechanism for the Posix filesystem API. The handling of this RPC mechanism, and support for features such as translation of filenames, types, attributes, and operations to and from the Posix interface, is an application-specific issue outside the scope of cryptlib.

SSH Port Forwarding

Alongside standard SSH connections and SSH subsystems, it's also possible to perform port-forwarding using SSH channels. Port forwarding allows you to tunnel an arbitrary network connection over SSH to avoid having the data being sent over the network in the clear. For example you could use this to tunnel mail (SMTP to send, POP3 or IMAP to receive) over SSH to and from a remote host. SSH provides two types of port forwarding, forwarding from the client to the server, identified by a channel type of "direct-tcpip", and forwarding from the server to the client, identified by a channel type of "tcpip-forward". The only one that's normally used is client-to-server forwarding.

For client-to-server forwarding with a channel type of "direct-tcpip", the first channel argument is the remote host and port that you want to forward to. For example if you wanted to tunnel SMTP mail traffic to `mailserver.com` with SMTP being on port 25 (so the forwarding string would be `mailserver.com:25`), you would use:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSH );

/* Add the server name, user name, and password */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    serverName, serverNameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    username, usernameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password, passwordLength );

/* Request the creation of the port-forwarding channel */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_SSH_CHANNEL,
    CRYPT_UNUSED );
cryptSetAttributeString( cryptSession,
    CRYPT_SESSINFO_SSH_CHANNEL_TYPE, "direct-tcpip", 12 );
cryptSetAttributeString( cryptSession,
    CRYPT_SESSINFO_SSH_CHANNEL_ARG1, "mailserver.com:25", 17 );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

When cryptlib activates the connection, it will indicate to the remote SSH server that it should forward data sent over the SSH link to port 25 on `mailserver.com`. You can now either push data directly into the session to tunnel it to the remote server, or create a socket to listen on port 25 on the local machine and push data received on it into the session, creating a local to remote system port forwarding over the SSH channel.

Before you forward the data on the server as requested by the client, you should check to make sure that the requested forwarding is in fact permitted. For example a malicious user could use port forwarding to attack a machine inside your firewall by forwarding connections through the firewall over an SSH tunnel. Because of this, cryptlib will never open a forwarded connection by itself, but requires that you

explicitly forward the data. In other words it will indicate that port forwarding has been requested, but will never of its own volition open and/or forward arbitrary ports just because a client has requested it.

If you don't want to allow the port forwarding, all you need to do is ignore the port-forwarding channel. `cryptlib`'s default action is to not allow forwarded connections, making it impossible for a client to remotely access internal machines or ports unless you explicitly allow it.

SSH Multiple Channels

Although SSH is usually used to provide a straightforward secure link from one system to another, it's also possible to use it to multiplex multiple virtual sessions across a single logical session. This is done by tunnelling multiple data channels across the SSH link.

SSH implements this using in-band signalling, which means that control information and data share the same link. With a single data channel (the standard case) this isn't a problem, but with multiple data channels control information for one channel can be impeded by data being sent or received on another channel. For example if you need to send or receive control information (channel close/channel open/status information) and there's a data transfer in progress on another channel, the control information can't be sent or received until the data transfer has been completed. This is why virtually all networking protocols use out-of-band signalling, with a separate mechanism for control signalling that can't be impeded by data transfers on the link.

Because of the in-band signalling problem, there are a number of special-case considerations that you need to take into account when using multiple SSH data channels. The primary one is: Don't do it. Unless you really have a strong need to run with multiple channels, just stick to a single channel and everything will be OK.

If you really need to use multiple channels, your code will need to take some extra steps to handle the problems caused by SSH's in-band signalling. The standard approach to this problem is to run the SSH implementation as a standalone service or daemon, with a full-time thread or task dedicated to nothing but handling any control messages that may arrive. These standalone applications are capable of opening ports to local and remote systems, logging on users, initiating data transfers, and so on. Since it's probably not a good idea for `cryptlib` to open arbitrary ports or transfer files without any additional checking, your application needs to explicitly manage these control messages. This requires doing the following:

- Try and open all channels and send all control messages right after the connect, before any data transfers are initiated. This means that the control signalling won't be stalled behind data signalling.
- Avoid using the session in non-blocking mode or with a very small timeout. Using a very short timeout increases the chances of some data remaining unwritten or unread, which will cause control information to become stalled behind it.
- Periodically try and pop data to handle any new control messages that may have arrived on other channels. In standalone SSH implementations that run as services or daemons, this is handled by having a full-time thread or task dedicated to this function. If you want to take this approach in your application, you can use a user-supplied socket that you can wait on in your application as described in "Network Issues" on page 212.
- Trying to perform channel control actions can result in a `CRYPT_ERROR_INCOMPLETE` status if there's data still waiting to be read or written. This occurs because it's not possible to send or receive control information for another channel until the data for the current channel has been cleared. Since new data can arrive after you've cleared the existing data but before you can send the control message, you may need to run this portion of your code in a loop to ensure that the channel is clear so that you can send the control information. Note that both the send and receive sides of the channel have to be cleared to allow the control message to be sent and a response received.

If you've decided that you really do need to use multiple SSH channels, you can manage them using the `CRYPT_SESSIONINFO_SSH_CHANNEL` attribute, which contains an integer value that uniquely identifies each channel. You can select the channel to send on by setting this attribute before you push data, and determine the channel that data is being received on by reading it before you pop data:

```
int receiveChannelID, bytesCopied;

/* Send data over a given channel */
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_SSH_CHANNEL,
    sendChannelID );
cryptPushData( cryptSession, data, dataSize, &bytesCopied );

/* Receive data sent over a channel */
cryptGetAttribute( cryptSession, CRYPT_SESSIONINFO_SSH_CHANNEL,
    &receiveChannelID );
cryptPopData( cryptSession, buffer, bufferSize, &bytesCopied );
```

Read and write channels are distinct, so setting the write channel doesn't change the read channel, which is specified in incoming data messages that arrive.

If you're opening additional channels after the session handshake has completed, you need to tell cryptlib when to activate the newly-created channel. To do this, you set its `CRYPT_SESSIONINFO_SSH_CHANNEL_ACTIVE` attribute to true, which activates the channel by sending the details to the remote system. Using the previous example of a port-forwarding channel, if you wanted to open this additional channel after the session had already been established you would use:

```
/* Request the creation of the port-forwarding channel */
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_SSH_CHANNEL,
    CRYPT_UNUSED );
cryptSetAttributeString( cryptSession,
    CRYPT_SESSIONINFO_SSH_CHANNEL_TYPE, "direct-tcpip", 12 );
cryptSetAttributeString( cryptSession,
    CRYPT_SESSIONINFO_SSH_CHANNEL_ARG1, "mailserver.com:25", 17 );

/* Activate the newly-created channel */
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_SSH_CHANNEL_ACTIVE,
    1 );
```

If you want to close one of the additional channels, you can select it in the usual manner and then deactivate it by setting its `CRYPT_SESSIONINFO_SSH_CHANNEL_ACTIVE` attribute to false:

```
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_SSH_CHANNEL,
    channelID );
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_SSH_CHANNEL_ACTIVE,
    0 );
```

If you try to deactivate the last remaining channel, which corresponds to the session itself, cryptlib will return a `CRYPT_ERROR_PERMISSION` status. To close the final channel, you need to close the overall session.

SSL/TLS Sessions

SSL/TLS is a secure data transfer protocol that provides confidentiality, integrity-protection, protection against replay attacks, and a variety of other services. The SSL server is authenticated via a certificate, and the client isn't authenticated (in rare circumstances client certificates may be used, but these are usually avoided due to the high degree of difficulty involved in working with them). Alternatively, the client and server may be mutually authenticated via a secret-key mechanism such as a user name and password, which avoids the need for certificates altogether. cryptlib supports SSL version 3, TLS version 1.0 (a.k.a SSL version 3.1), and TLS version 1.1 (a.k.a SSL version 3.2).

SSL and TLS are actually variations of the same protocol, the protocol known by the generic term SSL is SSL v3.0 and TLS is SSL v3.1. A newer revision of TLS, TLS version 1.1, is SSL v3.2. cryptlib will automatically negotiate the highest protocol version supported by the other side and use that to secure the session. You can determine which version is in use once the session has been established by reading

the CRYPT_SESSINFO_VERSION attribute, a value of 0 indicates version 3.0 or SSL, a value of 1 indicates version 3.1 or TLS, and a value of 2 indicates version 3.2 or TLS version 1.1. You can also force the use of a particular version by setting the protocol version attribute before you activate the connection, for example you can have cryptlib function as an SSL-only server by setting the CRYPT_SESSINFO_VERSION to 0 to indicate the use of SSL version 3.0 rather than TLS version 3.1. A (fortunately) small number of buggy servers will fail the SSL handshake if the protocol version is advertised as TLS, if you receive a handshake failure alert when you try to activate the session (as indicated by the CRYPT_ATTRIBUTE_INT_ERRORMESSAGE attribute) you can try forcing the use of SSL to see if the server can handle a connect using only the older protocol version.

Because TLS v1.1 is relatively new and not widely supported yet (meaning that some clients and servers will break if they encounter a server or client that advertises this protocol version), cryptlib by default advertises TLS v1.0 as its highest protocol level. If you want to explicitly advertise TLS v1.1, you can set the CRYPT_SESSINFO_VERSION attribute to 2 before you activate the session to indicate the use of SSL v3.2 or TLS v1.1.

SSL/TLS Client Sessions

Establishing a session with an SSL/TLS server requires adding the server name or IP address and an optional port number if it isn't using the standard SSL/TLS port. Once you've added this information, you can activate the session and establish the connection:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSL );

/* Add the server name and activate the session */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    serverName, serverNameLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

The same operation in Java or C# is:

```
/* Create the session */
int cryptSession = crypt.CreateSession( cryptUser,
    crypt.SESSION_SSL );

/* Add the server name and activate the session */
crypt.SetAttributeString( cryptSession, crypt.SESSINFO_SERVER_NAME,
    serverName );
crypt.SetAttribute( cryptSession, crypt.SESSINFO_ACTIVE, 1 );
```

The Visual Basic form of the code is:

```
Dim cryptSession As Long

' Create the session
cryptCreateSession cryptSession, cryptUser, CRYPT_SESSION_SSL

' Add the server name and activate the session
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_SERVER_NAME, _
    serverName, Len( serverName )
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1
```

Activating a session results in cryptlib performing a lot of work in the background. For example when activating the SSL/TLS session shown above cryptlib will connect to the remote host, read the server's certificate, generate a secret data value to exchange with the server using the key contained in the certificate, create the appropriate encryption contexts and load keys based on the secret data value into them, negotiate general session parameters, and complete negotiating the encrypted link with the server.

SSL/TLS with Shared Keys

Note: The use of SSL/TLS sessions using shared keys is based on a draft standard from the TLS working group that may be subject to further changes before the final

standard is adopted. You should avoid deploying solutions based on this mechanism until the standard has been finalised by the TLS working group.

If the server you're connecting to uses shared keys (for example a user name and password), you need to provide this information via the CRYPT_SESSINFO_USERNAME and CRYPT_SESSINFO_PASSWORD attributes to authenticate yourself to the server before you activate the connection:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSL );

/* Add the server name, user name, and password */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    serverName, serverNameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    username, usernameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password, passwordLength );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

The equivalent operation in Java or C# is:

```
/* Create the session */
int cryptSession = crypt.CreateSession( cryptUser,
    crypt.SESSION_SSL );

/* Add the server name, user name, and password */
crypt.SetAttributeString( cryptSession, crypt.SESSINFO_SERVER_NAME,
    serverName );
crypt.SetAttributeString( cryptSession, crypt.SESSINFO_USERNAME,
    username );
crypt.SetAttributeString( cryptSession, crypt.SESSINFO_PASSWORD,
    password );

/* Activate the session */
crypt.SetAttribute( cryptSession, crypt.SESSINFO_ACTIVE, 1 );
```

In Visual Basic this is:

```
Dim cryptSession As Long

' Create the session
cryptCreateSession cryptSession, cryptUser, CRYPT_SESSION_SSL

' Add the server name, user name, and password
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_SERVER_NAME, _
    serverName, Len( serverName )
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_USERNAME, _
    userName, Len( userName )
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_PASSWORD, _
    password, Len( password )

' Activate the session
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1
```

Authenticating yourself using shared keys avoids the need for both server and client certificates, providing mutual authentication for both client and server (conventional SSL only authenticates the server using a server certificate). This type of key management also avoids the high CPU overhead of public-key encryption, making it ideal for use in resource-constrained environments or ones where you're charged for CPU usage.

SSL/TLS with Client Certificates

If the server you're connecting to requires a client certificate, you need to provide a private key with an attached signing certificate via the CRYPT_SESSINFO_PRIVATEKEY attribute to authenticate yourself to the server before you activate the session. The private key could be a native cryptlib key, but it could also be a key from a crypto device such as a smart card or Fortezza card. They both work in the same way for client authentication:

```

CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SSL );

/* Add the server name and client key/certificate and activate the
session */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    serverName, serverNameLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY,
    cryptPrivateKey );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );

```

When cryptlib connects to the server, it will use the provided private key and certificate as part of the SSL/TLS handshake to authenticate the client to the server. If the server doesn't require the use of a client certificate, cryptlib won't do anything with the private key, so it's OK to add this even if you're not sure whether it'll be needed or not.

Note that client certificates are very rarely used in practice because of the high level of difficulty involved in working with them. If you require client authentication, a far better solution is to either use a traditional authentication mechanism such as sending an authenticator like a password over the SSL connection, or to use SSL with shared keys, which provides mutual authentication of both client and server.

SSL/TLS Server Sessions

Establishing an SSL/TLS server session requires adding the server key/certificate, activating the session, and waiting for incoming connections:

```

CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_SSL_SERVER );

/* Add the server key/certificate and activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY,
    privateKey );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );

```

The same procedure in Visual Basic is:

```

Dim cryptSession As Long

' Create the session
cryptCreateSession cryptSession, cryptUser, _
    CRYPT_SESSION_SSL_SERVER

' Add the server key/certificate and activate the session
cryptSetAttribute cryptSession, CRYPT_SESSIONINFO_PRIVATEKEY, privateKey
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1

```

Once you activate the session, cryptlib will block until an incoming client connection arrives, at which point it will negotiate a secure connection with the client.

SSL/TLS Servers with Shared Keys

Note: The use of SSL/TLS sessions using shared keys is based on a draft standard from the TLS working group that may be subject to further changes before the final standard is adopted. You should avoid deploying solutions based on this mechanism until the standard has been finalised by the TLS working group.

If you're using shared keys (for example a user name and password) to provide security, you need to provide this information via the CRYPT_SESSINFO_USERNAME and CRYPT_SESSINFO_PASSWORD attributes. For example if you have a server that allows one of three users/clients to connect to it you would use:

```

CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_SSL_SERVER );

```

```
/* Add the user names and passwords */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    username1, username1Length );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password1, password1Length );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    username2, username2Length );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password2, password2Length );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    username3, username3Length );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password3, password3Length );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

Using shared keys in this manner avoids the need for both server and client certificates, providing mutual authentication for both client and server (conventional SSL only authenticates the server via a server certificate). This type of key management also avoids the high CPU overhead of public-key encryption, making it ideal for use in resource-constrained environments or ones where you're charged for CPU usage.

If you have clients who need to connect without providing a user name and password, you can still provide a server certificate in the usual manner using the `CRYPT_SESSINFO_PRIVATEKEY` attribute, and clients who don't provide a user name and password will connect using public-key encryption. Note though that a client that uses the server certificate rather than a user name and password loses the benefits of mutual client/server authentication, as well as incurring a higher CPU overhead due to the use of public-key encryption.

Once a client has authenticated themselves using a shared key, you can determine their identity by reading back the `CRYPT_SESSINFO_USERNAME` attribute:

```
char username[ CRYPT_MAX_TEXTSIZE + 1 ];
int usernameLength

/* Get the user name */
cryptGetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    username, &usernameLength );
username[ usernameLength ] = '\0';
```

If the attempt by the client to connect fails (typically due to the use of an incorrect password), the password information for that user will be reset to prevent password-guessing attacks in which an attacker repeatedly reconnects using every possible password until they succeed. If the password is reset, you need to re-add the user and password to the session before that particular user can connect again. In order to protect against password-guessing attacks you should employ standard precautions such as allowing a maximum of three incorrect attempts or inserting a time delay before another connect attempt is allowed.

SSL/TLS Servers with Client Certificates

If you want to use client certificates to authenticate incoming connections, you need to provide a public-key keyset or certificate store for cryptlib to use to check certificates provided by client connections. When a client tries to establish a connection, cryptlib will check that their certificate is present in the keyset. If it isn't present, the connection isn't permitted. This provides a very fine-grained level of access control through which individual end users can be permitted or denied access to the host. Since cryptlib uses the keyset to verify incoming connections, you can control who is allowed in by adding or removing their certificate to or from the keyset. Note that you must provide a public-key keyset that stores certificates (not a private-key keyset) to the session since SSL/TLS uses certificates for the access control functionality.

You can specify the public-key keyset to use for checking incoming connections with the `CRYPT_SESSINFO_KEYSET` attribute:

```

CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_SSL_SERVER );

/* Add the server key and public-key keyset and activate the
   session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY,
    privateKey );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_KEYSET, cryptKeyset );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );

```

When you set this attribute for a server session, cryptlib will require the use of client certificates for connections to the server, and won't allow connections from clients that aren't able to authenticate themselves using a certificate that was previously added to the keyset.

Request/Response Protocol Sessions

cryptlib supports a variety of request/response protocols including protocols such as the certificate management protocol (CMP), simple certificate enrolment protocol (SCEP), real-time certificate status protocol (RTCS), online certificate status protocol (OCSP), and timestamping protocol (TSP). CMP, SCEP, RTCS, and OCSP client sessions are certificate management services that are covered in “Obtaining Certificates using CMP”, “Obtaining Certificates using SCEP”, “Certificate Status Checking using RTCS”, and “Certificate Revocation Checking using OCSP” on pages 252, 247, 247, and 253, and a TSP client session is an S/MIME service which is covered in “Timestamping” on page 178. RTCS, OCSP and TSP server sessions are standard session types and are also covered here, CMP and SCEP server sessions are somewhat more complex and are covered in “Managing a CA using CMP or SCEP” on page 261.

RTCS Server Sessions

An RTCS server session is a protocol-specific session type that returns a real-time certificate status to a client. RTCS client sessions are used for certificate status checks and are described in “Certificate Status Checking using RTCS” on page 247.

Establishing an RTCS server session requires adding a certificate store that cryptlib can query for certificate status information, specified as the CRYPT_SESSINFO_KEYSET attribute, and an optional RTCS responder key/certificate if you want cryptlib to sign the responses it provides. Certificate stores are described in more detail in “Managing a Certification Authority” on page 256. Once you've added this information you can activate the session and wait for incoming connections:

```

CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_RTCS_SERVER );

/* Add the certificate store and activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_KEYSET,
    cryptCertStore );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );

```

Once you activate the session, cryptlib will block until an incoming client connection arrives, at which point it will read the RTCS request from the client and return a response optionally signed with the RTCS responder key.

OCSP Server Sessions

An OCSP server session is a protocol-specific session type that returns certificate revocation information to a client. OCSP client sessions are used for certificate revocation checks and are described in “Certificate Revocation Checking using OCSP” on page 247.

The difference between RTCS and OCSP is that RTCS provides real-time, live certificate status information while OCSP provides delayed revocation information, usually based on CRLs. In other words RTCS answers the question “Is this certificate OK to use right now?” while OCSP answers the question “Was this certificate revoked at some point in the past?”. OCSP can’t return true validity information, so that if fed a freshly-issued certificate and asked “Is this a valid certificate”, it can’t say “Yes” (a CRL can only answer “revoked”), and if fed a forged certificate it can’t say “No” (it won’t be present in any CRL). In addition OCSP will often return a status result drawn from stale information hours or even days old, while RTCS (as the name implies) will always return real-time information. Finally, OCSP uses a peculiar means of identifying certificates that some implementations disagree over, with the result that a certificate may be regarded as valid even if it isn’t because client and server are talking about different things. In contrast RTCS returns an unambiguous yes-or-no response drawn from live certificate data. For these reasons RTCS is the cryptlib preferred certificate status protocol.

Establishing an OCSP server session requires adding the OCSP responder key/certificate and a certificate store that cryptlib can query for certificate status information, specified as the `CRYPT_SESSIONINFO_KEYSET` attribute. Certificate stores are described in more detail in “Managing a Certification Authority” on page 256. Once you’ve added this information you can activate the session and wait for incoming connections:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_OCSP_SERVER );

/* Add the OCSP responder key/certificate and certificate store and
   activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_PRIVATEKEY,
    privateKey );
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_KEYSET,
    cryptCertStore );
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_ACTIVE, 1 );
```

Once you activate the session, cryptlib will block until an incoming client connection arrives, at which point it will read the OCSP request from the client and return a response signed with the OCSP responder key.

TSP Server Sessions

A TSP server session is a protocol-specific session type that returns timestamp information to a client. TSP client sessions are used with S/MIME and are described in “Timestamping” on page 178. Establishing a TSP server session requires adding the timestamping authority (TSA) key/certificate, activating the session, and waiting for incoming connections:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_TSP_SERVER );

/* Add the TSA key/certificate and activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_PRIVATEKEY, privateKey
);
cryptSetAttribute( cryptSession, CRYPT_SESSIONINFO_ACTIVE, 1 );
```

The TSA certificate must be one that has the `CRYPT_CERTINFO_EXTKEY_-TIMESTAMPING` extended key usage attribute set to indicate that it can be used for generating timestamps. Extended key usage attributes are described in “Key Usage, Extended Key Usage, and Netscape certificate type” on page 325. If you add a key/certificate without this attribute, cryptlib will return `CRYPT_ERROR_PARAM3` to indicate that the key parameter is invalid.

Once you activate the session, cryptlib will block until an incoming client connection arrives, at which point it will read the timestamp request from the client and return a timestamp signed with the TSA key.

Obtaining Session Status Information

When a session is established a lot of state information is exchanged between the client and server and status information is generated by both sides. After the session has been activated you can query the session object for information such as the session status, which security parameters are being used, the identity of the remote client that connected to your server (the identity of the remote server is already known if you're the client), and authentication and identification information that was obtained from the client or server during the session establishment process.

Obtaining Session Security Parameters

If you want to know the details of the encryption mechanism which is being used to protect the session, you can read various CRYPT_CTXINFO_XXX attributes from the session object, which will return information from the encryption context(s) that are being used to secure the session. For example once you've activated the session you can get the encryption algorithm, mode, and the key size being used with:

```
CRYPT_ALGO_TYPE cryptAlgo;
CRYPT_MODE_TYPE cryptMode;
int keySize;

cryptGetAttribute( cryptSession, CRYPT_CTXINFO_ALGO, &cryptAlgo );
cryptGetAttribute( cryptSession, CRYPT_CTXINFO_MODE, &cryptMode );
cryptGetAttribute( cryptSession, CRYPT_CTXINFO_KEYSIZE, &keySize );
```

Authenticating the Host with Key Fingerprints

Once you've connected to a server, you can verify the server's certificate or key fingerprint by reading the CRYPT_SESSINFO_SERVER_FINGERPRINT attribute, which contains a fingerprint value that uniquely identifies the server's certificate or key. You can compare this to a stored copy of the fingerprint, or format it for display to the user.

If you set the CRYPT_SESSINFO_SERVER_FINGERPRINT attribute before you connect to the server, cryptlib will verify it against the server key when it connects and break off the connection attempt with a CRYPT_ERROR_WRONGKEY status if the server's certificate or key doesn't match the fingerprint you've specified. This allows you to filter out bogus servers and/or keys before you try to send any sensitive information to them.

To determine the server's key fingerprint (without having to connect to it first), you can read the CRYPT_SESSINFO_SERVER_FINGERPRINT attribute from the SSH server session after you've added the server's private key, and the CRYPT_CERTINFO_FINGERPRINT attribute from the SSL/TLS server certificate.

Using fingerprints for authentication is the most reliable of the methods covered here, since it provides a guaranteed match to a known key that can't be spoofed or forged.

Authenticating the Host or Client using Certificates

In addition to providing integrity and privacy protection for a communications session, some session protocols also provide a means of verifying that the host or client you're connecting to really is who they claim to be. For everything but the SSH protocol this authentication is performed by having the host supply a certificate or certificate chain signed by a trusted CA which is used during the protocol initialisation phase to establish the session. The general idea is that the certificate contains the name of the host that you're connecting to or the name of the entity which is providing a particular service (for example an RTCS responder), so you can use the returned certificate to verify that you really are communicating with this host and not a machine that has been set up by an attacker to masquerade as the host. In addition if you're using SSL or TLS with client certificates, you can use the

certificate provided by the client when they connect to verify their identity, and if you're using SSL or TLS with shared keys you already have mutual authentication of client and server without the need for certificates.

In practice due to factors such as outsourcing of web hosting services and the relocation of servers, the host certificate frequently doesn't correspond to the server you're supposed to be connecting to (which is why most browsers only display a warning and then connect anyway, or don't even warn). cryptlib doesn't place any restrictions on what it will and won't connect to or accept responses from, leaving it up to you to determine whether you want to continue the session if the server doesn't match what's given in the host certificate or expected by the client.

Once the session has been activated, you can read the host or client's certificate chain as the CRYPT_SESSINFO_RESPONSE attribute:

```
CRYPT_CERTIFICATE cryptCertificate;

cryptGetAttribute( cryptSession, CRYPT_SESSINFO_RESPONSE,
                  &cryptCertificate );
```

You can then work with the certificate chain as usual, for example you can verify it using **cryptCheckCert** or fetch the subject name information as explained in "Certificate Identification Information" on page 299.

Authenticating the Client via Port and Address

In addition to the stronger fingerprint and certificate authentication mechanisms, you can also determine the IP address and port that a client is connecting from if you're running as a server (if you're the client, you already know which server and port you're connecting to). You can obtain this information by reading the CRYPT_SESSINFO_CLIENT_NAME and CRYPT_SESSINFO_CLIENT_PORT attributes, which work in a similar manner to the CRYPT_SESSINFO_SERVER_NAME and CRYPT_SESSINFO_SERVER_PORT attributes:

```
char name[ CRYPT_MAX_TEXTSIZE + 1 ];
int nameLength, port

cryptGetStringAttribute( cryptSession, CRYPT_SESSINFO_CLIENT_NAME,
                       name, &nameLength );
name[ nameLength ] = '\0';
cryptGetAttribute( cryptSession, CRYPT_SESSINFO_CLIENT_PORT, &port );
```

The same operation in Visual Basic is:

```
Dim name as String
Dim nameLength as Long
Dim port as Long

name = String( CRYPT_MAX_TEXTSIZE, vbNullChar );
cryptGetStringAttribute cryptSession, CRYPT_SESSINFO_CLIENT_NAME, _
    name, nameLength
name = Left( name, nameLength )
cryptGetAttribute cryptSession, CRYPT_SESSINFO_CLIENT_PORT, port
```

Note that cryptlib returns the client's IP address in dotted-decimal form (for IPv4) or colon-delimited form (for IPv6) rather than its full name, since a single IP address can be aliased to multiple names and may require complex name resolution strategies. If you require a full name rather than an IP address you'll need to resolve it yourself, taking into account the multiple hostname issue, the fact that the client may be using NAT, and the possibility of DNS spoofing.

Exchanging Data

Once a general-purpose secure communications session has been established, you can exchange data with the remote client or server over the encrypted, authenticated link that it provides. This works exactly like pushing and popping data to and from an envelope, except that the session is effectively a bottomless envelope that can accept or return (depending on the remote system) an endless stream of data. In many cases the overhead involved in wrapping up a block of data and exchanging it with a remote client or server can be noticeable, so you should always push and pop as much data at

once into and out of a session as you can. For example if you have a 100-byte message and communicate it to the remote host as 10 lots of 10 bytes, this is much slower than sending a single lot of 100 bytes. This behaviour is identical to the behaviour in applications like disk I/O, where writing a single big file to disk is a lot more efficient than writing 10 smaller files.

cryptlib helps to eliminate this problem as much as possible by not wrapping up and dispatching session data until you explicitly tell it to by flushing the data through just as you would with an envelope:

```
cryptPushData( cryptSession, data, dataSize, &bytesCopied );
cryptFlushData( cryptSession );
```

In Visual Basic this is:

```
cryptPushData cryptSession, data, dataSize, bytesCopied
cryptFlushData cryptSession
```

This means that cryptlib will accumulate as much data as possible in the session's internal buffer before encrypting and integrity-protecting it and sending it through to the remote system, avoiding the inefficiency of processing and sending many small blocks of data. Note that you only need to flush data through in this manner when you explicitly want to force all of the data in the session buffer to be sent to the remote system. If you don't force a flush cryptlib handles this automatically in the most efficient manner possible using its built-in buffering mechanisms.

When you close a session, cryptlib will immediately shut down the session as is, without flushing data in internal session buffers. This is done to handle cases where a session is aborted (for example because the user cancels the transaction or because of a network error), and it becomes necessary to exit without sending further data. If you want to send any remaining data before destroying the session, you need to explicitly flush the data through before you destroy the session object (remember to check the return status of the final flush to make sure that all of your data was indeed sent).

Reading the response from the remote client or server works in a similar manner:

```
cryptPopData( cryptSession, buffer, bufferSize, &bytesCopied );
```

Unless you specify otherwise, all of cryptlib's network operations are non-blocking and near-asynchronous, waiting only the minimum amount of time for data to be sent or received before returning to the caller (you can make this fully asynchronous if you want, see the next section). cryptlib will provide whatever data is available from the remote client or server or write whatever is possible to the remote client or server and then return, which is particularly important for interactive sessions where small amounts of data are flowing back and forth simultaneously. The amount of data which is returned is indicated by the `bytesCopied` parameter. If this value is zero then no data is available or was written at the current time. Since the interface is non-blocking, your application can perform other processing while it waits for data to arrive.

If you'd prefer to have cryptlib not block at all or block for a longer amount of time when waiting for data to arrive or be sent, you can enable this behaviour as described in "Network Issues" on page 212. With blocking network operations enabled, cryptlib will wait for a user-defined amount of time for data to arrive before returning to the caller. If data arrives or is sent during the timeout interval, cryptlib returns immediately. With non-blocking behaviour it will return immediately without waiting for data to become available.

Since cryptlib reads and writes are asynchronous, you shouldn't assume that all the data you've requested has been transferred when the push or pop returns. cryptlib will only transfer as much data as it can before the timeout, which may be less than the total amount. In particular if data is flowing in both directions at once (that is, both sides are writing data and not reading it), the network buffers will eventually fill up, resulting in no more data being written. If this happens you need to occasionally

interleave a read with the writes to drain the buffers and allow further data to be transferred.

Network Issues

Sometimes a machine won't connect directly to the Internet but has to access it through a proxy. cryptlib supports common proxy servers such as **socks**, and also supports HTTP proxies if the protocol being used is HTTP-based. In addition it may be necessary to adjust other network-related parameters such as timeout values in order to accommodate slow or congested network links or slow clients or servers. The following sections explain how to work with the various network-related options to handle these situations.

Secure Sessions with Proxies

Using a **socks** proxy requires that you tell cryptlib the name of the **socks** server and an optional user name (most servers don't bother with the user name). You can set the **socks** server name with the `CRYPT_OPTION_NET_SOCKS_SERVER` attribute and the optional **socks** user name with the `CRYPT_OPTION_NET_SOCKS_USERNAME` attribute. For example to set the **socks** proxy server to the host called **socks** (which is the name traditionally given to **socks** servers) you would use:

```
cryptSetAttributeString( CRYPT_UNUSED, CRYPT_OPTION_NET_SOCKS_SERVER,  
    "socks", 5 );
```

before activating the session. When you activate the session, cryptlib will communicate with the proxy using the **socks** protocol.

Using an HTTP proxy works in a similar manner, with the name of the proxy being specified with the `CRYPT_OPTION_NET_HTTP_PROXY` attribute. Note that HTTP proxies require that the protocol being used employs HTTP as its transport mechanism, so they're not used with any other protocol type.

However, it's also possible to move SSL/TLS traffic through most types of HTTP proxies, since SSL is frequently used to carry HTTP data. If you enable the use of an HTTP proxy, cryptlib will also use it for SSL/TLS sessions.

Under Windows, cryptlib provides automatic proxy discovery and support. You can enable this by setting the proxy server to **Autodetect**:

```
cryptSetAttributeString( CRYPT_UNUSED, CRYPT_OPTION_HTTP_PROXY,  
    "Autodetect", 10 );
```

which instructs cryptlib to automatically detect and use whatever proxy is being employed. Since the proxy-discovery process can take a few seconds, you should only enable autodetection if you're sure that a proxy is actually present. Enabling it unconditionally will result in cryptlib spending a lot of time trying to find a proxy that may not exist, which slows down the network connection setup process.

Network Timeouts

When connecting to a server and carrying out other portions of a protocol such as security parameter and session key negotiation (for which cryptlib knows that a response must arrive within a certain time) cryptlib sets an interval timer and reports a connect or read error if no response arrives within that time interval. This means that if there's a network problem such as a host being down or a network outage, cryptlib won't hang forever but will give up after a certain amount of time, defaulting to 30 seconds. You can change the connect timeout value using the `CRYPT_OPTION_NET_CONNECTTIMEOUT` attribute, which specifies the connect timeout delay in seconds. For example to set a longer timeout for a remote host or client which is slow in responding you would use:

```
cryptSetAttribute( CRYPT_UNUSED, CRYPT_OPTION_NET_CONNECTTIMEOUT,  
    60 );
```

to set a one minute timeout when activating the session. If you want to set the connect timeout for a specific session rather than system-wide for all sessions, you can set the attribute only for the session object in question:

```
cryptSetAttribute( cryptSession, CRYPT_OPTION_NET_CONNECTTIMEOUT,
60 );
```

In addition to the connect timeout cryptlib has a separate timeout setting for network communications, specified using the `CRYPT_OPTION_NET_READTIMEOUT` and `CRYPT_OPTION_NET_WRITETIMEOUT` attributes. Since cryptlib session objects normally use non-blocking I/O once the session has been established and data is being exchanged, the read and write timeouts are set to minimal values during any general data exchanges that occur after the connection negotiation process has completed. This means that all communications after that point are near-asynchronous and non-blocking, however by changing the read/write timeout settings you can make cryptlib wait for a certain amount of time for data to arrive or be written before returning. For example to wait up to 30 seconds for data to arrive you would use:

```
cryptSetAttribute( CRYPT_UNUSED, CRYPT_OPTION_NET_READTIMEOUT, 30 );
```

If data arrives during the wait interval, cryptlib will return as soon as the data becomes available, otherwise it'll wait for up to 30 seconds for data to arrive.

As with the connect timeout, you can also apply these options directly to session objects, which means that they'll only apply to that particular session rather than being a system-wide setting for all session objects:

```
cryptSetAttribute( cryptSession, CRYPT_OPTION_NET_READTIMEOUT, 30 );
```

If you need the quickest possible response (usually only interactive sessions need this), you can set network read/write timeouts of zero, which will result in cryptlib returning immediately if no data can be read or written. The downside to using a zero timeout is that it reduces data transfers to polled I/O, requiring repeated read or write attempts to transfer data. For write timeouts it's better to set at least a small non-zero timeout rather than a zero timeout to ensure that the data is successfully written. In almost all cases the write will complete immediately even with a non-zero timeout, only in very rare cases such as when network congestion occurs will it be necessary to wait for data to be sent. In other words during a read wait the session is frequently just idling waiting for something to happen, but during a write wait it's actively trying to move the data out, so setting a non-zero timeout will increase the chances of the network layer moving the data out during the current write attempt rather than having to retry the write later.

A second problem with very short timeouts occurs when you close a session. Since writes are fully asynchronous, the network session can be closed before all of the data is written. Although the network stack tries to flush the data through, if there's an error during transmission there's no way to indicate this since the session has already been closed. cryptlib tries to mitigate this by setting a minimum (non-zero) network timeout when it closes a session, but there's no way to guarantee that everything will be sent during the timeout interval (in general this is an unsolvable problem, for example if an intermediate router crashes and is rebooted or the routing protocols hunt around for an alternative route, the transfer will eventually complete, but it could take several minutes for this to happen, which would require an excessively long timeout setting).

To avoid this issue, you should avoid writing a large amount of data with a very small network timeout setting and then immediately closing the session. You can do this by writing data at a measured pace (with a non-zero timeout) during the session or by setting a reasonable write timeout before you flush the last lot of data through and close the session.

Managing your Own Network Connections and I/O

Instead of having cryptlib automatically manage network connections, it's possible for you to manage them yourself. This can be useful when you want to customise session establishment and connection management, for example to handle a STARTTLS upgrade during an SMTP or IMAP session, an STLS upgrade during a POP session, or an AUTH TLS upgrade during an FTP session. You can also use this facility if you want to use any high-performance I/O capabilities provided by your

system, for example asynchronous I/O or hardware-accelerated I/O in which a dedicated subsystem manages all network transfers and posts a completion notification to your application when the transfer is complete. This allows you to use your own connection-management/socket-multiplexing/read-write code rather than using the facilities provided by cryptlib.

The following discussion refers to network sockets because this is the most common abstraction that's used for network I/O, however cryptlib will work with any network I/O identifier that can be represented by an integer value or handle. If your network abstraction requires more than a straightforward handle, you can pass in a reference or index to an array of whatever data structures your system requires to handle network I/O.

You can handle your own network connections by adding them to a cryptlib session as the `CRYPT_SESSINFO_NETWORKSOCKET` attribute before you activate the session. When you activate the session, cryptlib will use the socket that you've supplied rather than opening its own connection. Once you shut down the session, you can continue to use the socket or close it as required:

```
int socket;

/* Connect the network socket */
socket = ...;

/* Add the socket to the session and activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_NETWORKSOCKET,
    socket );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

Before you hand the socket over to cryptlib, you should disable Nagle's algorithm, since cryptlib provides its own optimised packet management. cryptlib leaves this task to the caller to ensure that it doesn't have to make any changes to the socket settings itself. In other words, it will leave the socket exactly as it found it. In addition you need to use a blocking socket, since cryptlib implements its own non-blocking I/O layer for portability across different operating systems. This is particularly important for Windows, where the socket must be non-blocking to avoid false reports of the other side closing the connection due to bugs in some versions of Winsock. Note that if you use certain Winsock functions such as `WSAAsyncSelect` and `WSAEventSelect` on the socket, Windows will quietly switch the socket back to non-blocking mode, so you need to be careful about inadvertently changing the state of a socket behind cryptlib's back.

In addition to managing the connection process, you can also use externally-supplied sockets to handle network reads and writes. There are two general mechanisms used for external network I/O, the Berkeley sockets `select()`-style mechanism:

```
select( ... );
...
read( ... );
```

and the posted-read/posted-write mechanism used by high-performance and hardware-accelerated I/O subsystems:

```
read_async( ... );
...
wait_completion( ... );
```

An example of the latter is Windows' I/O completion ports, which allow a central dispatcher to initiate I/Os and a pool of client threads to manage them as required whenever a request completes. The equivalent under Unix (although it's less attuned towards high-performance server operation, being targeted mainly at file I/O) is Posix asynchronous I/O. Other operating systems provide similar facilities, for example Tandem NSK has the `RECV_NW` and `AWAITIOX` calls to perform posted reads and writes.

The more widely-used I/O model, using the `select()`-style mechanism, would wait until data is available to be read on the socket and then call **cryptPopData**:

```

/* Wait for data to become available */
select( ... );

/* Read data from the session */
cryptPopData( cryptSession, ... );

```

The posted-read/posted-write mechanism would have a read or write initiated by a central dispatcher (in the example below this is illustrated with Windows-style I/O handling):

```

/* Create an I/O completion port associated with the socket */
hCompletionPort = CreateIOCompletionPort( hSocket, ... );

/* Initiate the read request */
ReadFile( hSocket, ... );

```

Once the read request has been completed by the underlying I/O system, a thread from the thread pool that's waiting on the completion port is woken up and handles the result:

```

/* Wait for data to arrive */
GetQueuedCompletionStatus( hCompletionPort, ... );

/* Read data from the session */
cryptPopData( cryptSession, ... );

```

The Windows kernel contains a number of special optimisations to provide the best possible performance for this type of I/O. If you're running a high-performance server, you should consider using this style of I/O instead of the standard sockets interface for better performance. In fact this style of I/O is the one that's used by servers like IIS to maximise performance.

The Unix equivalent would be:

```

/* Initiate the read request */
aio_read( &aioch );

/* Wait for data to arrive */
aio_suspend( &aioch, ... );

/* Read data from the session */
cryptPopData( cryptSession, ... );

```

Unix asynchronous I/O is often used for high-performance I/O when the overhead of the standard BSD `select()` is unacceptable. A typical `select` implementation, for example, has to first copy and validate the socket descriptor masks for read, write, and exception conditions, then call the underlying device's poll routine for each socket descriptor in each mask to let the device know that an I/O operation is being requested for that descriptor, and finally wait for a notification on any of the descriptors from the lower-level device drivers. There's additional overhead created by the fact that the kernel can't afford to lock out I/O while all of this polling is taking place, so the `select` code has to be able to handle the case of I/O occurring during the polling process, usually by restarting the poll.

Asynchronous I/O, on the other hand, avoids all of this overhead by simply posting a read or write and then waiting for the kernel to notify it that the operation has completed. It therefore provides much better performance than an equivalent `select`-based implementation.

If you supply the network socket yourself and the socket is a server socket, you can no longer read the `CRYPT_SESSIONINFO_CLIENT_NAME` and `CRYPT_SESSIONINFO_CLIENT_PORT` attributes, since these are recorded when the incoming client connection is established, and won't be present with a user-supplied socket.

Key Generation and Storage

The previous sections on enveloping and secure sessions mentioned the use of encryption contexts containing public and private keys. The creation and generation of public/private keys in encryption contexts and the long-term storage of key data is covered in this section. Keys are stored in keysets, an abstract container that can hold one or more keys and that can be a cryptlib key file, a PGP/OpenPGP key ring, a database, an LDAP directory, or a URL accessed via HTTP. cryptlib accesses all of these keyset types using a uniform interface that hides all of the background details of the underlying keyset implementations. In addition you can generate and store keys in crypto devices such as smart cards, crypto accelerators, and Fortezza cards. Crypto devices are explained in more detail in “Encryption Devices and Modules” on page 350. The direct loading of key data into a context is covered in “Encryption and Decryption” on page 268.

Key Generation

To create an encryption context, you must specify the user who is to own the object or CRYPT_UNUSED for the default, normal user, and the algorithm you want to use for that context. The available public-key algorithms are given in “Algorithms” on page 380. For example, to create and destroy an RSA context you would use:

```
CRYPT_CONTEXT cryptContext;

cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_RSA );

/* Load key, perform en/decryption */

cryptDestroyContext( cryptContext );
```

Note that the CRYPT_CONTEXT is passed to **cryptCreateContext** by reference, as **cryptCreateContext** modifies it when it creates the encryption context. In almost all other cryptlib routines, CRYPT_CONTEXT is passed by value. The contexts that will be created are standard cryptlib contexts, to create a context which is handled via a crypto device such as a smart card or Fortezza card, you should use **cryptDeviceCreateContext**, which tells cryptlib to create a context in a crypto device. The use of crypto devices is explained in “Encryption Devices and Modules” on page 350.

Generating a Key Pair into an Encryption Context

Once you’ve created an encryption context the next step is to generate a public/private key pair into it. Before you can generate the key pair you need to set the CRYPT_CTXINFO_LABEL attribute which is later used to identify the key when it’s written to or read from a keyset or a crypto device such as a smart card or a Fortezza card using functions like **cryptAddPrivateKey** and **cryptGetPrivateKey**. If you try to generate a key pair into a context without first setting the key label, cryptlib will return CRYPT_ERROR_NOTINITED to indicate that the label hasn’t been set yet. The process of generating a public/private key pair is then:

```
CRYPT_CONTEXT privKeyContext;

cryptCreateContext( &privKeyContext, cryptUser, CRYPT_ALGO_RSA );
cryptSetAttributeString( privKeyContext, CRYPT_CTXINFO_LABEL, label,
    labelLength );
cryptGenerateKey( privKeyContext );
```

To do this in Java or C# you would use:

```
int privKeyContext = crypt.CreateContext( cryptUser, crypt.ALGO_RSA );
crypt.SetAttributeString( privKeyContext, crypt.CTXINFO_LABEL,
    label );
crypt.GenerateKey( privKeyContext );
```

The Visual Basic equivalent is:

```
Dim privKeyContext As Long

cryptCreateContext privKeyContext, cryptUser, CRYPT_ALGO_RSA
cryptSetAttributeString privKeyContext, CRYPT_CTXINFO_LABEL, label, _
    Len( label )
cryptGenerateKey privKeyContext
```

If you want to generate a key of a particular length, you can set the `CRYPT_CTXINFO_KEYSIZE` attribute before calling **cryptGenerateKey**. For example to generate a 1536-bit (192-byte) key you would use:

```
CRYPT_CONTEXT privKeyContext;

cryptCreateContext( &privKeyContext, cryptUser, CRYPT_ALGO_RSA );
cryptSetAttributeString( privKeyContext, CRYPT_CTXINFO_LABEL, label,
    labelLength );
cryptSetAttribute( cryptContext, CRYPT_CTXINFO_KEYSIZE, 1536 / 8 );
cryptGenerateKey( cryptContext );
```

You can also change the default encryption and signature key sizes using the cryptlib configuration options `CRYPT_OPTION_PKC_KEYSIZE` and `CRYPT_OPTION_SIG_KEYSIZE` as explained in “Working with Configuration Options” on page 359. Once a key is generated into a context, you can’t load or generate a new key over the top of it. If you try to do this, cryptlib will return `CRYPT_ERROR_INITED` to indicate that a key is already loaded into the context.

Although cryptlib can work directly with public/private key data held in an encryption context, you can’t communicate this key data to anyone else without first turning it into an encoded key object like a certificate. This is because the key consists of a (potentially large) number of abstract components that need to be encoded into a standard format in order to communicate them to someone else, with a certificate (or some equivalent object like a certificate request) being the standard way to do this.

Because of this key-encoding requirement, you can’t immediately use a newly-generated private key for anything other than signing a certification request or a self-signed certificate (although you can store a raw key in a file keyset for later use, see the next section for more details). To use the key for any other purpose, you need to convert it into a certificate and then store the certificate alongside the private key in a cryptlib private key file or crypto device. The process of obtaining a certificate and updating a keyset or device with it is covered in more detail in “Certificates and Certificate Management” on page 234. Once you’ve obtained the certificate, you can add it to the keyset or device in which the key is stored, and cryptlib will automatically associate it with the key when you read it.

Asynchronous Key Generation

Because the generation of larger public/private keys may take some time on a slower system, cryptlib provides an asynchronous key generation capability that allows the key to be generated as a background task or thread on those systems that provide this capability. You can generate a key asynchronously with **cryptGenerateKeyAsync**, which works in the same way as **cryptGenerateKey**. You can check the status of an asynchronous key generation with **cryptAsyncQuery**, which will return `CRYPT_ERROR_TIMEOUT` if the key generation operation is in progress or `CRYPT_OK` if the operation has completed. Any attempt to use the context while the key generation operation is still in progress will also return `CRYPT_ERROR_TIMEOUT`:

```
cryptGenerateKeyAsync( privKeyContext );
do
{
    /* Perform other task(s) */
    /* ... */
}
while( cryptAsyncQuery( privKeyContext ) == CRYPT_ERROR_TIMEOUT );
```

You can cancel the asynchronous key generation using **cryptAsyncCancel**.

Since the background key generation depends on how the operating system schedules threads, you shouldn't call **cryptAsyncQuery** immediately after calling **cryptGenerateKeyAsync** because the thread that performs the key generation may not have had time to run yet, making it appear as if the key generation has already completed since the context won't have started the generation process yet. The code example given above (which performs other work before querying the key generation progress) avoids any OS thread scheduling issues by performing another task while the OS starts the key generation thread in the background.

Keyset Types

cryptlib supports a wide variety of keyset types. Most of these are public-key keysets, which means that you can only store X.509 certificates (and by extension the public keys associated with them) in them, but not private keys. These keyset types include database keysets (the cryptlib native format for storing certificates), LDAP directories, and web pages accessed via HTTP.

In addition to the public-key keysets, cryptlib also supports the storage of private keys in cryptlib private key files (which use the PKCS #15 crypto token format) and crypto devices such as smart cards, Fortezza cards, and hardware crypto accelerators. cryptlib keysets can also be used to store certificates, but only those that already have a corresponding private key stored in the keyset. cryptlib private key keysets can't be used as general-purpose public-key or certificate stores, they can only store certificates associated with an existing private key.

The following table summarises the different keyset types and the operations that are possible with each one. Unless you have a strong reason not to do so, it's recommended that you use cryptlib private key files to store private keys and their associated certificates and database keysets to store standalone certificates.

Type	Access Allowed
cryptlib	Read/write access to public/private keys and any associated certificates stored in a file using the PKCS #15 crypto token format, with the private key portion encrypted. This is the cryptlib native keyset format for private keys.
Crypto device	Read access to public/private keys and read/write access to certificates stored in the device. Devices aren't general-purpose keysets but can act like them for keys contained within them. More information on crypto devices and on generating private keys in them is given in "Encryption Devices and Modules" on page 350.
Database	Read/write access to X.509 certificates stored in a database. This is the cryptlib native keyset format for public keys and certificates and provides a fast, scalable key storage mechanism. The exact database format used depends on the platform, but would typically include any ODBC database under Windows, and Informix, Ingres, Oracle, Postgres, and Sybase databases under other platforms.
HTTP	Read access to X.509 certificates and CRLs accessed via URLs.
LDAP	Read/write access to X.509 certificates and CRLs stored in an LDAP directory.
PGP	Read access to PGP/OpenPGP key rings.

The recommended method for certificate storage is to use a database keyset, which usually outperforms the other keyset types by a large margin, is highly scalable, and is well suited for use in cases where data is already administered through existing database servers.

Creating/Destroying Keyset Objects

Keysets are accessed as keyset objects that work in the same general manner as the other container objects used by cryptlib. You create a keyset object with **cryptKeysetOpen**, specifying the user who is to own the device object or CRYPT_UNUSED for the default, normal user, the type of keyset you want to attach it to, the location of the keyset, and any special options you want to apply for the keyset. This opens a connection to the keyset. Once you've finished with the keyset, you use **cryptKeysetClose** to sever the connection and destroy the keyset object:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, keysetType, keysetLocation,
  keysetOptions );

/* Load/store keys */

cryptKeysetClose( cryptKeyset );
```

The available keyset types are:

Keyset Type	Description
CRYPT_KEYSET_FILE	A flat-file keyset, either a cryptlib private key file or a PGP/-OpenPGP key ring.
CRYPT_KEYSET_HTTP	URL specifying the location of a certificate or CRL.
CRYPT_KEYSET_LDAP	LDAP directory.
CRYPT_KEYSET_PLUGIN	Generic RDBMS accessed via the database network plugin interface
CRYPT_KEYSET_DATABASE	Generic RDBMS interface.
CRYPT_KEYSET_ODBC	Generic ODBC RDBMS interface.
CRYPT_KEYSET_DATABASE_-STORE	As for the basic keyset types, but representing a certificate store for use by a CA rather than a simple keyset. The user who creates and updates these keyset types must be a CA user.
CRYPT_KEYSET_PLUGIN_STORE	
CRYPT_KEYSET_ODBC_STORE	

These keyset types and any special conditions and restrictions on their use are covered in more detail below.

The keyset location varies depending on the keyset type and is explained in more detail below. Note that the CRYPT_KEYSET is passed to **cryptKeysetOpen** by reference, as the function modifies it when it creates the keyset object. In all other routines, CRYPT_KEYSET is passed by value.

The keyset options are:

Keyset Option	Description
CRYPT_KEYOPT_-CREATE	Create a new keyset. This option is only valid for writeable keyset types, which includes keysets implemented as databases and cryptlib key files.
CRYPT_KEYOPT_NONE	No special access options (this option implies read/write access).
CRYPT_KEYOPT_-READONLY	Read-only keyset access. This option is automatically enabled by cryptlib for keyset types that have read-only restrictions enforced by the nature of the keyset, the operating

Keyset Option	Description
	system, or user access rights.
	Unless you specifically require write access to the keyset, you should use this option since it allows cryptlib to optimise its buffering and access strategies for the keyset.

These options are also covered in more detail below.

File Keysets

For cryptlib private key files and PGP/OpenPGP key rings, the keyset location is the path to the disk file. For example to open a connection to a cryptlib key file `key.p15` located in `/users/dave/`, you would use:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE,
    "/users/dave/keys.p15", CRYPT_KEYOPT_READONLY );
```

cryptlib will automatically determine the file type and access it in the appropriate manner. Since cryptlib uses the PKCS #15 crypto token format to store private keys, the files are given a `.p15` extension or an appropriate equivalent as dictated by the operating system being used. As another example, to open a connection to a cryptlib private key file located in the **Keys** share on the Windows server **FileServer**, you would use:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE,
    "\\FileServer\\Keys\\key.p15", CRYPT_KEYOPT_READONLY );
```

The same operation in Visual Basic is:

```
Dim cryptKeyset As Long

cryptKeysetOpen cryptKeyset, cryptUser, CRYPT_KEYSET_FILE, _
    "\\FileServer\\Keys\\key.p15", CRYPT_KEYOPT_READONLY
```

When you open a PGP/OpenPGP keyset, cryptlib will automatically set the access mode to read-only even if you don't specify the `CRYPT_KEYOPT_READONLY` option, since writes to this keyset type aren't supported. If you try to write a key to this keyset type, cryptlib will return `CRYPT_ERROR_PERMISSION` to indicate that you don't have permission to write to the file. The only file keyset type that can be written to is a cryptlib private key file. This keyset contains one or more encrypted private keys and any associated certificates. To create a new cryptlib keyset you would use:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE,
    "Private key file.p15", CRYPT_KEYOPT_CREATE );
```

The equivalent in Java or C# is:

```
int cryptKeyset = crypt.KeysetOpen( cryptUser, crypt.KEYSET_FILE,
    "Private key file.p15", crypt.KEYOPT_CREATE );
```

If a cryptlib keyset of the given name already exists and you open it with `CRYPT_KEYOPT_CREATE`, cryptlib will erase it before creating a new one in its place. The erasure process involves overwriting the original keyset with random data and committing the write to disk to ensure that the data really is overwritten, truncating its length to 0 bytes, resetting the file timestamp and attributes, and deleting the file to ensure that no trace of the previous key remains. The new keyset is then created in its place.

For security reasons, cryptlib won't write to a file if it isn't a normal file (for example if it's a hard or symbolic link, if it's a device name, or if it has other unusual properties such as having a stream `fattach()` 'd to it).

Where the operating system supports it, cryptlib will set the security options on the keyset so that only the person who created it (and, in some cases, the system administrator) can access it. For example under Unix the file access bits are set to allow only the file owner to access the file, and under Windows NT/2000/XP/Vista the file's access control list is set so that only the user who owns the file can access or change it. Since not even the system administrator can access the keyset under Windows NT/2000/XP/Vista, you may need to manually enable access for others to allow the file to be backed up or copied.

If your application is running as another user (for example if it's running as a daemon under Unix or a service under Windows), the keyset will be owned by the daemon or service that creates it, following standard security practice. If you want to make the keyset accessible to standard users, you need to either change the security options to allow the required user access (for example by changing the file access permissions or running in the context of the intended user when you create it), or provide an interface to your daemon/service to allow access to the keyset. The latter is generally the preferred option, since it allows your daemon/service to control exactly what the user can do with the keyset.

In addition if you're installing or configuring cryptlib as one user for use by another user, you'll need to adjust the access for any files that are created during the install or configuration process to allow access by the target user. For example if you install and configure cryptlib as a Windows administrator to run as a system service, you'll need to change the ownership of any key and configuration files to the system account:

```
cacls filename /e /g system:f
```

If you don't do this then the service (running under the system account) can't access the key/configuration files created under the administrator account.

When you open a keyset that contains private keys, you should bind it to the current thread for added security to ensure that no other threads can access the file or the keys read from it:

```
CRYPT_KEYSET cryptKeyset;

/* Open a keyset and claim it for exclusive use */
cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE,
    "Private key file.pl5", CRYPT_KEYOPT_READONLY );
cryptSetAttribute( cryptKeyset, CRYPT_PROPERTY_OWNER, threadID );
```

You can find out more about binding objects to threads in "Object Security" on page 42.

HTTP Keysets

For keys accessed via an HTTP URL, the keyset name is the URL:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_HTTP, url,
    CRYPT_KEYOPT_READONLY );
```

HTTP keysets normally behave just like any other keysets, however if you're reading a key from a fixed URL (with no per-key ID) you need to use the special ID `[none]` to indicate that the keyset URL points directly at the certificate. For example to read a certificate from the static URL `http://www.server.com/cert.der` you would use:

```
CRYPT_KEYSET cryptKeyset;
CRYPT_HANDLE publicKey;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_HTTP,
    "http://www.server.com/cert.der", CRYPT_KEYOPT_READONLY );
cryptGetPublicKey( cryptKeyset, &cryptCertificate, CRYPT_KEYID_NAME,
    "[none]" );
```

The CRLs provided by some CAs can become quite large, so you may need to play with timeouts in order to allow the entire CRL to be downloaded if the link is slow or congested.

If you want to publish certificates online, the best way to do this is with an HTTP keyset. The server side of HTTP certificate access is handled as a standard cryptlib session, and is covered in “Making Certificates Available Online” on page 262.

Database Keysets

For keys (strictly speaking, X.509 certificates) that are stored in a database, the keyset location is the access path to the database. The nature of the access path depends on the database type, and ranges from an alias or label that identifies the database (for example an ODBC data source) through to a complex combination of the name or address of the server that contains the database, the name of the database on the server, and the user name and password required to access the database.

The exact keyset type also depends on the operating system with which cryptlib is being used. Under Windows, all database keyset types are accessed as ODBC data sources with the keyset type CRYPT_KEYSET_ODBC. The ODBC interface is also available for most database types under Unix through various Unix ODBC drivers. For the few system that don't provide a vendor-independent database access system, database keysets are accessed either directly or via a generic network plugin interface that allows cryptlib to communicate with any type of database backend. The direct database interface, which compiles the database interface into cryptlib, has a keyset type CRYPT_KEYSET_DATABASE. All other databases are accessed through an RPC mechanism specified using a keyset type of CRYPT_KEYSET_PLUGIN. With some systems that don't support any type of database access (for example some embedded systems have no database capability), cryptlib can't be used with a database keyset and is restricted to the simpler keyset types such as cryptlib private key files.

The simplest type of keyset to access is a local database that requires no extra parameters such as a user name or password. An example of this is an ODBC data source on the local machine. For example if the keyset is stored in a database such as Ingres, MySQL, Oracle, SQL Server, Sybase, or Postgres, which is accessed through the “PublicKeys” data source, you would access it with:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_ODBC,
    "PublicKeys", CRYPT_KEYOPT_READONLY );
```

The same operation in Visual Basic is:

```
Dim cryptKeyset As Long

cryptKeysetOpen cryptKeyset, cryptUser, CRYPT_KEYSET_ODBC,
    "PublicKeys", CRYPT_KEYOPT_READONLY
```

The second type of database keyset is one which is accessed through a plugin that converts cryptlib data accesses to the format used by the database backend. The generic plugin interface takes as parameters the name of the server that cryptlib is to connect to and an optional port number separated by a colon. For example if the database ran on the server **keyserver.company.com**, the keyset would be accessed with:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_PLUGIN,
    "keyserver.company.com", CRYPT_KEYOPT_READONLY );
```

Through the use of the plugin interface, cryptlib can access any type of database across any OS platform. Details on writing the required plugin are given in “Database and Networking Plugins” on page 380.

The database name parameter used above was a simple ODBC data source or database name, but this can also contain a user name, password, and server name, in

the format `user:pass@server`. For example, you can specify a combination of database user name and password as `user:pass`, and a user name and server as `user@server`. Other, database-specific combinations and parameters may also be possible, depending on the database backend you're using.

In the examples shown above, the keyset was opened with the `CRYPT_KEYOPT_READONLY` option. The use of this option is recommended when you'll use the keyset to retrieve a certificate but not store one (which is usually the case) since it allows cryptlib to optimise its transaction management with the database backend. This can lead to significant performance improvements due to the different data buffering and locking strategies that can be employed if the back-end knows that the database won't be updated. If you try to write a certificate to a keyset that has been opened in read-only mode, cryptlib will return `CRYPT_ERROR_PERMISSION` to indicate that you don't have permission to write to the database.

To create a new certificate database, you can use the `CRYPT_KEYOPT_CREATE` option. If a keyset of the given name already exists, cryptlib will return `CRYPT_ERROR_DUPLICATE`, otherwise it will create a new certificate database ready to have certificates added to it.

Database keysets can also be used as certificate stores, an extended type of keyset which is required in order to perform CA operations such as issuing certificates and CRLs. In order to create this type of keyset instead of a conventional one you must be a CA user and you need to specify its type as `CRYPT_KEYSET_DATABASE_STORE`, `CRYPT_KEYSET_ODBC_STORE`, or `CRYPT_KEYSET_PLUGIN_STORE` instead of the basic database keyset type. Certificate stores have a higher overhead than normal keysets because they meet a number of special CA-specific requirements, so you should only create one if you are using it to run a CA. In addition, certificates and CRLs can't be directly added to or deleted from a certificate store but have to be processed using cryptlib's certificate management functionality. More information on certificate stores and their use is given in "Managing a Certification Authority" on page 256.

In order to create or open a certificate store, you must be a CA user. If you try to access a certificate store and aren't a CA user, cryptlib will return `CRYPT_ERROR_PARAM2` to indicate that the user type isn't valid for accessing this type of keyset. Normal users can't update a certificate store in any way, however they can access them in read-only mode as normal database keysets. For example while a CA could open a certificate store as:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_PLUGIN_STORE,
    "certstore.company.com", CRYPT_KEYOPT_NONE );
```

and perform updates on the store, a non-CA user could only access it in read-only mode as a standard database keyset:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_PLUGIN,
    "certstore.company.com", CRYPT_KEYOPT_READONLY );
```

When opened in this manner the certificate store appears as a standard database keyset rather than as a full certificate store.

To provide additional security alongside the precautions taken by cryptlib, you should apply standard database security measures to ensure that all database keyset accesses are carried out with least privileges. For example if your application only needs read access to a keyset, you can use the SQL GRANT/REVOKE mechanism to allow read-only access of the appropriate kind for the application. An SQL statement like `REVOKE ALL ON certificates FROM user; GRANT SELECT ON certificates TO user` would allow only read accesses to the certificate keyset. You can also use server-specific security measures such as accessing the keyset through SQL Server's built-in `db_datareader` account, which only allows read access to tables, and the

ability to run the application under a dedicated low-privilege account (a standard feature of Unix systems).

LDAP Keysets

For keys stored in an LDAP directory, the keyset location is the name of the LDAP server, with an optional port if access is via a non-standard port. For example if the LDAP server was called `directory.ldapservers.com`, you would access the keyset with:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_LDAP,
    "directory.ldapservers.com", CRYPT_KEYOPT_READONLY );
```

If the server is configured to allow access on a non-standard port, you can append the port to the server name in the usual manner for URL's. For example if the server mentioned above listened on port 8389 instead of the usual 389 you would use:

```
CRYPT_KEYSET cryptKeyset;

cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_LDAP,
    "directory.ldapservers.com:8389", CRYPT_KEYOPT_READONLY );
```

You can also optionally include the `ldap://` protocol specifier in the URL, this is ignored by cryptlib.

The storage of certificates in LDAP directories is haphazard and vendor-dependent, and you may need to adjust cryptlib's LDAP configuration options to work with a particular vendor's idea of how certificates and CRLs should be stored on a server. In order to make it easier to adapt cryptlib to work with different vendor's ways of storing information in a directory, cryptlib provides various LDAP-related configuration options that allow you to specify the X.500 objects and attributes used for certificate storage. These options are:

Configuration Option	Description
CRYPT_OPTION_KEYS - LDAP_CERTNAME	The X.500 attribute that certificates are stored as. For some reason certificates belonging to certification authorities (CAs) are stored under their own attribute type, so if a search for a certificate fails cryptlib will try again using the CA certificate attribute (there's no easy way to tell in advance how a certificate will be stored, so it's necessary to do it this way). In addition a number of other attribute types have been invented to hide certificates under, it may require a bit of experimentation to determine how the server you're using stores things.
CRYPT_OPTION_KEYS - LDAP_CACERTNAME	
	The default settings for these options are <code>userCertificate;binary</code> and <code>cACertificate;binary</code> , a variety of other choices also exist. Note the use of the <code>binary</code> qualifier, this is required for a number of directories that would otherwise try and encode the returned information as text rather than returning the raw certificate.
CRYPT_OPTION_KEYS - LDAP_CRLNAME	The X.500 attribute that certificate revocation lists (CRLs) are stored as, defaulting to <code>certificateRevocationList;binary</code> .
CRYPT_OPTION_KEYS - LDAP_EMAILNAME	The X.500 attribute that email addresses are stored as, defaulting to <code>mail</code> . Since X.500 never defined an email address attribute,

Configuration Option	Description
	various groups defined their own ones, mail is the most common one but there are a number of other alternatives around, including emailAddress, rfc822Name, rfc822MailBox, and email. As usual, some experimentation will be necessary to find out what works.
CRYPT_OPTION_KEYS - LDAP_FILTER	The filter used to selected returned LDAP attributes during a query, defaulting to (objectclass=*) . Experimentation will be necessary to determine what's required for this value.
CRYPT_OPTION_KEYS - LDAP_OBJECTCLASS	The X.500 object class, defaulting to inetOrgPerson. Again, there is no consistency among servers, the usual amount of guesswork will be required to find out what works.
CRYPT_OPTION_KEYS - LDAP_OBJECTTYPE	The object type to fetch, defaulting to CRYPT_CERTTYPE_NONE to fetch all object types. Setting this to CRYPT_CERTTYPE_CERTIFICATE or CRYPT_CERTTYPE_CRL will fetch only certificates or CRLs.

These configuration options apply to all LDAP keysets, you can also apply them to an individual keyset object rather than as a general configuration option, which means that they'll affect only the one LDAP keyset object.

There is no consistency in the configuration of LDAP directories, and since the query used to retrieve a certificate depends on how the directory is configured, it's often impossible to tell what to submit without asking the directory administrators for the correct formula. Since the actual values depend on the server configuration, there is no way that cryptlib can determine which ones to use for a given server.

Two examples of magic formulae that are required by different CAs running LDAP directories are "searchDN = CN=Norway Post Organizational CA, O=CA, C=NO, filter = (&(objectclass=*)(pssSubjectDNString=CN=RTV EDI-server 2, O=RTV, C=NO)), attributes = certificateRevocationList;binary" and "(&((&(objectclass=inetorgperson)(objectclass=organizationalperson)) (objectClass=Strong-AuthenticationUser))(usercertificate;binary=*)(|(commonname=name)(rfc822-mailbox=email address)))". In order to handle some of these combinations you will have to set a selection of the CRYPT_OPTION_KEYSET_LDAP_... attributes as well as modifying the key ID you use when you actually read a key.

To allow even more flexibility in specifying LDAP access parameters, cryptlib will also accept RFC 1959 LDAP URLs as key IDs (see "Obtaining a Key for a User" on page 226). These have the general form `ldap://host:port/dn?-attributes?scope?filter`, and can be used to specify arbitrarily complex combinations of DN components (see RFC 1485), search scope, and filter (see RFC 1558). For example to specify the Norway post magic formula above as a key ID the LDAP URL would be `ldap:///CN=Norway%20Post%20Organizational%20CA,%20O=CA,%20C=NO?certificate-RevocationList;binary??(&(objectclass=*)(pssSubjectDNString=%20CN=RTV%20EDI-server%202,%20O=RTV,%20C=NO))`. Note that the ability to use an LDAP URL for lookup in this manner may not be available in some LDAP client implementations.

The default settings used by cryptlib have been chosen to provide the best chance of working, however given that everyone who stores certificates in an LDAP server

configures it differently it's almost guaranteed that trying to use LDAP to store certificates will require reconfiguration of the client, the server, the certificates being stored, or several of the above in order to function. In effect the LDAP configuration acts as a form of access control mechanism that makes it impossible to access certificates or CRLs until the CA reveals the correct magic formula. For this reason the use of LDAP is not recommended for storing certificates.

Reading a Key from a Keyset

Once you've set up a connection to a keyset, you can read one or more keys from it. Some keysets such as HTTP URLs can contain only one key, whereas cryptlib private key files, PGP/OpenPGP key rings, databases, and LDAP keysets may contain multiple keys.

You can also use a crypto device such as a smart card, Fortezza card, or crypto hardware accelerator as a keyset. Reading a key from a device creates an encryption context which is handled via the crypto device, so that although it looks just like any other encryption context it uses the device to perform any encryption or signing.

The two functions that are used to read keys are **cryptGetPublicKey** and **cryptGetPrivateKey**, which get a public and private key respectively. The key to be read is identified through a key identifier, either the name or the email address of the key's owner, specified as `CRYPT_KEYID_NAME` and `CRYPT_KEYID_EMAIL`, or the label assigned to the key as the `CRYPT_CTXINFO_LABEL` attribute when it's generated or loaded into a context, also specified as `CRYPT_KEYID_NAME`.

cryptGetPublicKey returns a generic `CRYPT_HANDLE` that can be either a `CRYPT_CONTEXT` or a `CRYPT_CERTIFICATE` depending on the keyset type. Most public-key keysets will return an X.509 certificate, but some keysets (like PGP/OpenPGP key rings) don't store the full certificate information and will return only an encryption context rather than a certificate. You don't have to worry about the difference between the two, they are interchangeable in most cryptlib functions.

Obtaining a Key for a User

The rules used to match the key ID to a key depend on the keyset type, and are as follows:

Type	User ID Handling
Cryptlib	The key ID is a label attached to the key via the <code>CRYPT_CTXINFO_LABEL</code> attribute when it's generated or loaded into the context, and is specified using <code>CRYPT_KEYID_NAME</code> . Alternatively, if a certificate is associated with the key, the key ID can also be the name or email address indicated in the certificate.
Crypto device	The key ID is matched in full in a case-insensitive manner.
Database	The key ID is either the name or the email address of the key owner, and is matched in full in a case-insensitive manner.
HTTP	The key ID is either the name or the email address of the key owner, and is matched in full in a case-sensitive manner. The one exception is when the location is specified by a static URL, in which case the key ID has the special value <code>[none]</code> .
LDAP	The key ID is an X.500 distinguished name (DN), which is neither a name nor an email address but a peculiar construction that (in theory) uniquely identifies a key in the X.500 directory. Since a DN isn't really a name or an email address, it's possible to match an entry using either <code>CRYPT_KEYID_NAME</code> or <code>CRYPT_KEYID_EMAIL</code> . The key ID is matched in a manner which is controlled by the way the LDAP server is configured (usually the match is case-

Type	User ID Handling
	insensitive).
	You can also specify an LDAP URL as the key ID as described in “LDAP Keysets” on page 224.
PGP	<p>The key ID is a name with an optional email address which is usually given inside angle brackets. Since PGP keys usually combine the key owner’s name and email address into a single value, it’s possible to match an email address using CRYPT_KEYID_NAME, and vice versa.</p> <p>The key ID is matched as a substring of any of the names and email addresses attached to the key, with the match being performed in a case-insensitive manner. This is the same as the matching performed by PGP.</p> <p>Note that, like PGP, this will return the first key in the keyset for which the name or email address matches the given key ID. This may result in unexpected matches if the key ID that you’re using is a substring of a number of names or email addresses that are present in the key ring. Since email addresses are more likely to be unique than names, it’s a good idea to specify the email address to guarantee a correct match.</p>

Assuming that you wanted to read Noki Crow’s public key from a keyset, you would use:

```
CRYPT_HANDLE publicKey;

cryptGetPublicKey( cryptKeyset, &publicKey, CRYPT_KEYID_NAME,
    "Noki S.Crow" );
```

In Java or C# this is:

```
int publicKey = crypt.GetPublicKey( cryptKeyset, crypt.KEYID_NAME,
    "Noki S.Crow" );
```

In Visual Basic the operation is:

```
Dim publicKey As Long

cryptGetPublicKey cryptKeyset, publicKey, CRYPT_KEYID_NAME, _
    "Noki S.Crow"
```

Note that the CRYPT_HANDLE is passed to **cryptGetPublicKey** by reference, as the function modifies it when it creates the public key context. Reading a key from a crypto device works in an identical fashion:

```
CRYPT_HANDLE publicKey;

cryptGetPublicKey( cryptDevice, &publicKey, CRYPT_KEYID_NAME,
    "Noki S.Crow" );
```

The only real difference is that any encryption performed with the key is handled via the crypto device, although cryptlib hides all of the details so that the key looks and functions just like any other encryption context.

You can use **cryptGetPublicKey** not only on straight public-key keysets but also on private key keysets, in which case it will return the public portion of the private key or the certificate associated with the key.

The other function which is used to obtain a key is **cryptGetPrivateKey**, which differs from **cryptGetPublicKey** in that it expects a password alongside the user ID if the key is being read from a keyset. This is required because private keys are usually stored encrypted and the function needs a password to decrypt the key. If the key is held in a crypto device (which requires a PIN or password when you open a session with it, but not when you read a key), you can pass in a null pointer in place of the password. For example if Noki Crow’s email address was `noki@crow.com`

and you wanted to read their private key, protected by the password “Password”, from a keyset, you would use:

```
CRYPT_CONTEXT privKeyContext;

cryptGetPrivateKey( cryptKeyset, &privKeyContext, CRYPT_KEYID_EMAIL,
    "noki@crow.com", "Password" );
```

The same operation in Visual Basic is:

```
Dim privKeyContext As Long

cryptGetPrivateKey cryptKeyset, privKeyContext, CRYPT_KEYID_EMAIL, _
    "noki@crow.com", "Password"
```

If you supply the wrong password to **cryptGetPrivateKey**, it will return **CRYPT_ERROR_WRONGKEY**. You can use this to automatically handle the case where the key might not be protected by a password (for example if it’s stored in a crypto device or a non-cryptlib keyset that doesn’t protect private keys) by first trying the call without a password and then retrying it with a password if the first attempt fails with **CRYPT_ERROR_WRONGKEY**. cryptlib caches key reads, so the overhead of the second key access attempt is negligible:

```
CRYPT_CONTEXT privKeyContext;

/* Try to read the key without a password */
if( cryptGetPrivateKey( cryptKeyset, &privKeyContext,
    CRYPT_KEYID_NAME, name, NULL ) == CRYPT_ERROR_WRONGKEY )
{
    /* Ask the user for the keys' password and retry the read */
    password = ...;
    cryptGetPrivateKey( cryptKeyset, &privKeyContext, CRYPT_KEYID_NAME,
        name, password );
}
```

cryptGetPrivateKey always returns an encryption context.

General Keyset Queries

Where the keyset is implemented as a standard database, you can use cryptlib to perform general queries to obtain one or more certificates that fit a given match criterion. For example you could retrieve a list of all the keys that are set to expire within the next fortnight (to warn their owners that they need to renew them), or that belong to a company or a division within a company. You can also perform more complex queries such as retrieving all certificates from a division within a company that are set to expire within the next fortnight. cryptlib will return all certificates that match the query you provide, finally returning **CRYPT_ERROR_COMPLETE** once all matching certificates have been obtained.

The general strategy for performing queries is as follows:

```
submit query
repeat
    read query result
while query status != CRYPT_COMPLETE
```

You can cancel a query in progress at any time by submitting a new query consisting of the command “cancel”.

Queries are submitted by setting the **CRYPT_KEYINFO_QUERY** attribute for a keyset, which tells it how to perform the query. Let’s look at a very simple query which is equivalent to a straight **cryptGetPublicKey**:

```
CRYPT_CERTIFICATE certificate;

cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY,
    "$email='noki@crow.com'", 22 );
do
    status = cryptGetPublicKey( keyset, &certificate, CRYPT_KEYID_NONE,
        NULL );
while( cryptStatusOK( status ) );
```

This will read each certificate corresponding to the given email address from the database. Note that the key ID is unused because the keys that are returned are selected by the initial query and not by the key identifier.

This example is an artificially simple one, it's possible to submit queries of arbitrary complexity in the form of full SQL queries. Since the key databases that are being queried can have arbitrary names for the certificate attributes (corresponding to database columns), cryptlib provides a mapping from certificate attribute to database field names. An example of this mapping is shown in the code above, in which `$email` is used to specify the email address attribute, which may have a completely different name once it reaches the database backend. The certificate attribute names are as follows:

Attribute	Field
\$C, \$SP, \$L, \$O, \$OU, \$CN	Certificate country, state or province, locality, organisation, organisational unit, and common name.
\$date	Certificate expiry date
\$email	Certificate email address

You can use these attributes to build arbitrarily complex queries to retrieve particular groups of certificates from a key database. For example to retrieve all certificates issued for US users (obviously this is only practical with small databases) you would use:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$C='US'", 7 );
```

Extending this one stage further, you could retrieve all certificates issued to Californian users with:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$C='US' AND  
$SP='CA'", 20 );
```

Going another step beyond this, you could retrieve all certificates issued to users in San Francisco:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$C='US' AND  
$SP='CA' AND $L='San Francisco'", 43 );
```

Going even further than this, you could retrieve all certificates issued to users in San Francisco whose names begin with an 'a':

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$C='US' AND  
$SP='CA' AND $L='San Francisco' AND $CN LIKE 'A%'", 61 );
```

These queries will return the certificates in whatever order the underlying database returns them in. You can also specify that they be returned in a given order, for example to order the certificates in the previous query by user name you would use:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$C='US' AND  
$SP='CA' AND $L='San Francisco' ORDER BY $CN", 56 );
```

To return them in reverse order, you would use:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$C='US' AND  
$SP='CA' AND $L='San Francisco' ORDER BY $CN DESCENDING", 67 );
```

The ability to selectively extract collections of certificates provides a convenient mechanism for implementing a hierarchical certificate database browsing capability. You can also use it to perform general-purposes queries and certificate extractions, for example to return all certificates that will expire within the next week (and that therefore need to be replaced or renewed), you would use:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$date < today +  
1 week", length );
```

To sort the results in order of urgency of replacement, you would use:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$date < today +  
1 week ORDER BY $date", length );
```

To retrieve all certificates that don't need replacement within the next week, you could negate the previous query to give:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "NOT $date <
    today + 1 week", length );
```

As these examples show, cryptlib's keyset query capability provides the ability to perform arbitrary general-purpose queries on keysets.

Once a query has begun running, it can return a considerable number of certificates. If you try to initiate another query while the first one is in progress or perform a standard read, write, or delete operation, cryptlib will return a `CRYPT_ERROR_INCOMPLETE` error to indicate that the query is still active. You can cancel the currently active query at any point by setting the `CRYPT_KEYINFO_QUERY` attribute to "cancel":

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "cancel", 6 );
```

This will clear the current query and prepare the keyset for another query or an alternative operation such as a key read, write, or delete.

Handling Multiple Certificates with the Same Name

Sometimes a keyset may contain multiple certificates issued to the same person. Whether this situation will occur varies by CA, some CAs won't issue multiple certificates with the same name, some will, and some may modify the name to eliminate conflicts, for example by adding unique ID values to the name or using middle initials to disambiguate names. If multiple certificates exist, you can perform a keyset query to read each in turn and try and find one that matches your requirements, for example you might be able to filter them based on key usage or some other parameter held in the certificate. The general idea is to issue a query based on the name and then read each certificate that matches the query until you find an appropriate one:

```
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "...", ... );
while( cryptGetPublicKey( &certificate, keyset, ... ) == CRYPT_OK && \
    certificate doesn't match required usage )
    /* Continue */;
cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "cancel", 6 );
```

This use of general queries allows the maximum flexibility in selecting certificates in cases when multiple choices are present.

Key Group Management

Sometimes it may be desirable to treat a group of keys in the same way. For example if a collection of servers use keys to protect their communications with each other then compromise of one key may require the revocation of all keys in the group and the issuance of a new group of keys. The easiest way to handle key groups is by assigning a common identifier to all the keys in the group when you issue certificates for them, and then replacing all keys with that identifier when it comes time to update the key group.

The first part of the process involves assigning a key group identifier to certificates. The easiest way to do this is to specify it as part of the PKI user information that's used with the CMP and SCEP protocols. For example to specify that a PKI user belongs to the remote access users key group using the organisational unit portion of the user DN, you would use:

```

/* ... */

/* Add PKI user identification information */
cryptSetAttributeString( cryptPKIUser, CRYPT_CERTINFO_COUNTRYNAME,
    countryName, 2 );
cryptSetAttributeString( cryptPKIUser,
    CRYPT_CERTINFO_ORGANIZATIONNAME, organizationName,
    organizationNameLength );
cryptSetAttributeString( cryptPKIUser,
    CRYPT_CERTINFO_ORGANIZATIONALUNITNAME, "Remote access key group",
    23 );
cryptSetAttributeString( cryptPKIUser, CRYPT_CERTINFO_COMMONNAME,
    commonName, commonNameLength );

/* ... */

```

When the user requests their certificate, the key group will be given as the organisational unit component (alongside the other components such as the organisation name and country) in their DN. More information on working with PKI users is given in “Initialising PKI User Information” on page 258. Alternatively, you can manually set the key group identifier when you issue a certificate to someone in the key group if you’re manually issuing certificates rather than using an automated mechanism like CMP or SCEP.

The second part of the process involves identifying all of the certificates in a key group that need to be revoked or replaced. This is handled through cryptlib’s keyset query capability, retrieving each certificate in the group in turn:

```

CRYPT_CERTIFICATE certificate;

cryptSetAttributeString( keyset, CRYPT_KEYINFO_QUERY, "$OU='Remote
access key group'", 30 );
do
    status = cryptGetPublicKey( keyset, &certificate, CRYPT_KEYID_NONE,
        NULL );
while( cryptStatusOK( status ) );

```

Once the certificate has been fetched, you can revoke it or notify the owner that they need to replace it as required. More information on keyset queries is given in “General Keyset Queries” on page 228.

Writing a Key to a Keyset

Writing a key to a keyset isn’t as complex as reading it since there’s no need to specify the key identification information which is needed to read a key, however there are some restrictions on the type of key you can write to a keyset. Public-key keysets such as database and LDAP keysets store full certificates, so the object that you write to these keysets must be a `CRYPT_CERTIFICATE` and not just a `CRYPT_CONTEXT`. In contrast, keysets such as cryptlib private key files primarily store public/private key pairs but can also store the certificate or certificates that are associated with the private key. If you try to write the incorrect type of object to a keyset (for example a private key to a certificate keyset), cryptlib will return a `CRYPT_ERROR_PARAM2` error to indicate that the object you are trying to add is of the incorrect type for this keyset.

If you try to write a key to a read-only keyset, cryptlib will return `CRYPT_ERROR_PERMISSION` to indicate that you can’t write to the keyset. If you try to write a certificate to a cryptlib private key file or a crypto device that doesn’t already have a corresponding private key present, cryptlib will return `CRYPT_ERROR_PARAM2` to indicate that you can’t add this type of object if there isn’t already a matching private key present. If you just want to write a certificate to a file, you can use **cryptExportCert** to obtain the certificate and then write that to a file.

You can write a certificate to a public key keyset with **cryptAddPublicKey**, which takes as parameters the keyset and the key certificate to write:

```

cryptAddPublicKey( cryptKeyset, cryptCertificate );

```

Since all identification information is contained in the certificate, there’s no need to specify any extra data such as the certificate owner’s name or email address.

Writing a private key requires one extra parameter, the password which is used to encrypt the private key components. cryptlib will use the default encryption method (usually three-key triple DES) to encrypt the key with the given password. If you're writing the private key to a crypto device, the password parameter should be set to NULL since the device provides its own protection for the key (not all devices support direct key loading, some require the key to be generated inside the device).

To write a private key to a keyset you would use the corresponding **cryptAddPrivateKey** function:

```
cryptAddPrivateKey( cryptKeyset, privKeyContext, password );
```

If the certificate you are trying to write is already present in the keyset, cryptlib will return CRYPT_ERROR_DUPLICATE. If the keyset is a public-key keyset, you can use **cryptDeleteKey** to delete the existing certificate so you can write the new one in its place. If the keyset is a cryptlib key file or crypto device, this would delete both the certificate and the key it corresponds to. Finally, certificate stores can't be directly manipulated by adding or deleting certificates and CRLs but must be managed using cryptlib's certificate management functionality. If you try to directly insert or delete a certificate or CRL, cryptlib will return CRYPT_ERROR_PERMISSION to indicate that this operation isn't allowed.

There is one instance in which it's possible to add a new certificate to a cryptlib private key file when there's already an existing certificate present, and that's when the new certificate updates the existing one. For example some CAs will re-issue a certificate with a newer expiry date (rather than using a new key and certificate), if you add this new certificate to the keyset cryptlib will replace the existing, older certificate with the newer one and use the newer one in all future operations.

You can't create a key inside a standard cryptlib context and then move it to the device later on since the security features of the device won't allow this, and you can't take a key created via a crypto device and write it to a keyset, because it can't be exported from the device. By using crypto hardware to handle your keys you're guaranteeing that the key is never exposed outside the hardware, keeping it safe from any malicious code that might be present in your system.

Although cryptlib can work directly with private keys, other formats like X.509 certificates, S/MIME messages, and SSL require complex and convoluted naming and identification schemes for their keys. Because of this, you can't immediately use a newly-generated private key with these formats for anything other than signing a certification request or a self-signed certificate. To use it for any other purpose, you need to obtain an X.509 certificate that identifies the key and then store the certificate alongside the private key in a cryptlib private key file or crypto device. The process of obtaining a certificate and updating a keyset or device with it is covered in more detail in "Certificates and Certificate Management" on page 234. Once you've obtained the certificate, you can add it to the keyset or device and cryptlib will automatically associate it with the key when you read the key.

If you are working with a database keyset, you can also add a certificate revocation list (CRL) to the keyset. Since a CRL isn't an actual key, you can't read it back out of the keyset (there's nothing to read), but you can use it to check the revocation state of certificates. CRLs and their uses are explained in more detail in "Certificate Revocation using CRLs" on page 314.

Changing a Private Key Password

Changing the password on a private key file involves reading the key from a keyset using the old password, deleting the key from the keyset, and writing the in-memory copy back again using the new password:

```
read key from keyset using old password;  
delete key from keyset;  
re-write key to keyset using new password;
```

All cryptlib key file updates are atomic all-or-nothing operations, which means that if the computer crashes between deleting the old key and writing the new one, the old

key will still be present when the machine is rebooted (specifically, all changes are committed when the keyset is closed, which minimises the risk of losing data due to a system crash or power outage in the middle of a long sequence of update operations).

To update a private key with a new password, you would use:

```
CRYPT_KEYSET cryptKeyset;
CRYPT_CONTEXT cryptKey;

/* Read the key from the keyset using the old password */
cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE,
    keysetName, CRYPT_KEYOPT_NONE );
cryptGetPrivateKey( cryptKeyset, &cryptKey, CRYPT_KEYID_NAME, label,
    oldPassword );

/* Delete the current copy of the key from the keyset */
cryptDeleteKey( cryptKeyset, label );

/* Write the key back to the keyset using the new password */
cryptAddPrivateKey( cryptKeyset, cryptKey, newPassword );
cryptKeysetClose( cryptKeyset );
```

The same operation in Visual Basic is:

```
Dim cryptKeyset As Long
Dim cryptKey As Long

' Read the key from the keyset using the old password
cryptKeysetOpen cryptKeyset, cryptUser, CRYPT_KEYSET_FILE, keysetName,
    CRYPT_KEYOPT_NONE
cryptGetPrivateKey cryptKeyset, cryptKey, CRYPT_KEYID_NAME, label,
    oldPassword

' Delete the current copy of the key from the keyset
cryptDeleteKey cryptKeyset, label

' Write the key back to the keyset using the new password
cryptAddPrivateKey cryptKeyset, cryptKey, newPassword
cryptKeysetClose cryptKeyset
```

Deleting a Key

Deleting a key with **cryptDeleteKey** works in the same manner as reading a key, with the key to delete being identified by a key ID in the usual manner. For example if you wanted to delete S.Crow's key from a keyset, you would use:

```
cryptDeleteKey( cryptKeyset, CRYPT_KEYID_NAME, "S.Crow" );
```

Deleting a key from a crypto device is identical:

```
cryptDeleteKey( cryptDevice, CRYPT_KEYID_NAME, "S.Crow" );
```

In the case of an LDAP directory, this will delete the entire entry, not just the certificate attribute or attributes for the entry. In the case of a cryptlib private key file or crypto device, this will delete the key and any certificates that may be associated with it. If you try to delete a key from a read-only keyset, cryptlib will return **CRYPT_ERROR_PERMISSION**. If the key you're trying to delete isn't present in the keyset, cryptlib will return **CRYPT_ERROR_NOTFOUND**.

Certificates and Certificate Management

Although cryptlib can work directly with private keys, other formats like X.509 certificates, S/MIME messages, and SSL require complex and convoluted naming and identification schemes for their keys. Because of this, you can't immediately use a newly-generated private key with these formats for anything other than signing a certification request or a self-signed certificate. To use it for any other purpose, you need to obtain an X.509 certificate that identifies the key. Once you've obtained the certificate, you can update the keyset or device that contains the basic public/private key data with additional certificate information. This additional information can be a standalone certificate or a full certificate chain from a trusted root CA down to the end user certificate. This chapter covers the details of obtaining a certificate or certificate chain and attaching it to a private key.

The certificate management message exchange is usually carried out via HTTP or email or through some other unspecified mechanism, however cryptlib also supports the Certificate Management Protocol (CMP) and Simple Certificate Enrolment Protocol (SCEP), which define a mechanism for communicating with a CA to obtain certificates and request the revocation of existing certificates. This chapter explains how to use CMP and SCEP to obtain a certificate or request a revocation from a CA. In order to check a certificate's status, you can use the real-time certificate status protocol (RTCS) to perform a certificate status check, or the online certificate status protocol (OCSP) to perform a certificate revocation check only. The RTCS and OCSP checking processes are also covered in this chapter.

High-level vs. Low-level Certificate Operations

As with the general cryptlib programming interface, cryptlib supports certificate management operations at three levels:

Plug-and-play PKI

The highest level is the plug-and-play PKI level, which is the easiest one to use and therefore the recommended one. At this level, cryptlib handles all certificate processing and management operations for you, requiring no special knowledge of certificate formats, protocols, or operations. Because of its simplicity and ease of use, it's strongly recommended that you use this interface if at all possible.

Mid-level Certificate Management

The intermediate level requires some knowledge of key generation procedures and certificate management operations. This level involves the use of CMP and SCEP to obtain certificates and manage a CA, and RTCS and OCSP for certificate status checking. Most of the details of certificate management are taken care of for you by cryptlib, but you'll need to perform some manual handling of certificate management operations.

Low-level Certificate Management

The lowest level involves manually managing certificates and certificate revocations, and requires dealing with an entire range of arcane, difficult-to-use, and largely dysfunctional mechanisms such as Distinguished Names, X.500 directories, certificate revocation lists, and assorted other paraphernalia. Working with certificates at this level is extraordinarily difficult, and you should be absolutely certain that you're prepared for the large amount of effort that will be required to make anything work. At a minimum, you should read through and understand the certificate tutorials mentioned in "Recommended Reading" on page 15 before trying to do anything with low-level certificate operations.

If you're absolutely certain that you must work with certificates at a low level, and that you understand just how much effort will be involved, you can find out more about low-level certificate operations in "Certificates in Detail" on page 288 and "Certificate Extensions" on page 320.

Certificates and Keys

Once a public/private key pair is saved to a private key keyset, cryptlib allows extra certificate information to be added to the keyset. For example the process of creating a keyset containing a certificate and private key is:

```
generate public/private key pair;
write key pair to keyset;
submit certification request to CA;
receive certificate from CA;
update keyset to include certificate;
```

If the certificate is a self-signed CA certificate, there's no need to obtain the certificate from an external CA and you can add it directly to the keyset after you create it. If the key pair is being generated in a crypto device such as a smart card or Fortezza card, this process is:

```
generate public/private key pair;
submit certification request to CA;
receive certificate from CA;
update device to include certificate;
```

This example assumes that the certificate is immediately available from a CA, which is not always the case. The full range of possibilities are covered in more detail further on.

Once you've updated the private key with a certificate (which is the only time you can write a public key certificate to a private key keyset), cryptlib will automatically associate the certificate with the private key so that when you read it with **cryptGetPrivateKey** cryptlib will recreate the certificate alongside the key and attach it to the key. You can then use the combined certificate and key to perform operations that require the use of certificates such as certificate signing, S/MIME email decryption and signing, and user authentication. If you update the private key with a complete certificate chain instead of just a single certificate, cryptlib will attach the full certificate chain to the key when you read it with **cryptGetPrivateKey**.

The update process involves adding the certificate information to the keyset or device, which updates it with the certificate object (either a certificate or a certificate chain):

```
cryptAddPublicKey( cryptKeyset, cryptCertificate );
```

The certificate object which is being written must match a private key stored in the keyset or device. If it doesn't match an existing private key, cryptlib will return a **CRYPT_ERROR_PARAM2** error to indicate that the information in the certificate object being added is incorrect. If there's already a certificate for this key present, cryptlib will return a **CRYPT_ERROR_DUPLICATE** error to indicate that one key can't have two different certificates associated with it. See "Writing a Key to a Keyset" on page 231 for more on writing keys to keysets.

Using Separate Signature and Encryption Certificates

It's good security practice to use different keys for signing and encryption, and most digital signature laws contain some requirement that the two capabilities be implemented with separate keys. cryptlib supports the use of two (or more) keys belonging to a single user, the only issue to be aware of is that you should give each key a distinct label to allow it to be selected with **cryptGetPrivateKey**. For example the process of creating a keyset containing separate signature and encryption keys with the signature key labelled "My signature key" and the encryption key labelled "My encryption key" would be:

```
set key label to "Signature key";
generate public/private signature key pair;
set key label to "Encryption key";
generate public/private encryption key pair;
write key pairs to keyset;
submit certification requests to CA;
receive signature certificate from CA;
receive encryption certificate from CA;
update keyset to include certificates;
```

When you want to sign data, you would call **cryptGetPrivateKey** specifying the use of “Signature key”; when you want to decrypt data you would call **cryptGetPrivateKey** specifying the use of “Encryption key” (or cryptlib’s automatic key management will find it for you if you’re using it with a cryptlib envelope).

Plug-and-play PKI

The easiest way to set up keys and certificates is through cryptlib’s plug-and-play PKI facility, which performs the operations described above for you. To set up keys and certificates in this manner, cryptlib requires a private-key keyset or crypto token such as a smart card or Fortezza card to store keys and certificates in, the URL of a plug-and-play PKI-capable CA, and a user name and password to authorise the issuing of the certificates. The session type for the plug-and-play PKI is **CRYPT_SESSION_CMP**, the same type as a standard CMP session except that cryptlib manages everything for you. The private-key keyset or crypto token is specified as **CRYPT_SESSINFO_CMP_PRIVKEYSET**, and the user name and password to authorise the operation are provided as the **CRYPT_SESSINFO_USERNAME** and **CRYPT_SESSINFO_PASSWORD**:

```
CRYPT_SESSION cryptSession;
CRYPT_KEYSET cryptKeyset;

/* Create the CMP session and private-key keyset */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_CMP );
cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE,
    keysetName, CRYPT_KEYOPT_CREATE );

/* Add the server name/address */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER, server,
    serverLength );

/* Add the username, password, and private-key keyset */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    userName, userNameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password, passwordLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_CMP_PRIVKEYSET,
    cryptKeyset );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, TRUE );
```

The same operation in Visual Basic is:

```
Dim cryptSession As Long
Dim cryptKeyset As Long

' Create the CMP session and private-key keyset
cryptCreateSession cryptSession, cryptUser, CRYPT_SESSION_CMP
cryptKeysetOpen cryptKeyset, cryptUser, CRYPT_KEYSET_FILE, _
    keysetName, CRYPT_KEYOPT_CREATE

' Add the server name/address
cryptSetAttributeString cryptSession CRYPT_SESSINFO_SERVER, _
    server, Len( server )

' Add the username, password, and private-key keyset
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_USERNAME, _
    userName, Len( userName )
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_PASSWORD, _
    password, Len( password )
cryptSetAttribute cryptSession, CRYPT_SESSINFO_CMP_PRIVKEYSET, _
    cryptKeyset
```

```
' Activate the session
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1
```

Once this process has been completed, the private-key keyset or crypto token that you provided will contain a signature key identified by the label “Signature key”, and an encryption key identified by the label “Encryption key” if the public-key algorithm being used is capable of encryption, along with any additional certificates and CA certificates that are required to use the keys. Both keys will be protected using the password that you provided to authenticate the certification process. If you want to change the password for either of the keys you can do so as described in “Changing a Private Key Password” on page 232 before you close the keyset and commit the data to disk. Alternatively, if you want to retain the password that you used for the certificate issue to protect the keys and certificates, you can close the keyset immediately after you add it to the session and cryptlib will manage it for you.

If the CA is issuing you a CA certificate of your own, the keyset or crypto token will contain a single CA signing key identified by the label “Signature key”. Since CA keys can’t be used for encryption or general-purpose signing but only for signing other certificates, only the single CA signing key is created.

In addition to returning your own certificates, the plug-and-play PKI mechanism also performs a PKIBoot certificate bootstrap operation that downloads an initial trusted certificate set for you to use. This trusted certificate set only contains a small number of known-good certificates trusted by the CA that provided you with your own certificates, rather than the 100+ certificates that you’d be forced to automatically trust when you use a web browser (some of these browser certificates have weak 512-bit keys, or are owned by CAs that have gone out of business, or whose private keys have been on-sold to third parties when the original owner went bankrupt, sometimes passing through multiple owners). The PKIBoot operation allows an end user, starting with nothing more than the user name and password used for the plug-and-play PKI operation to acquire all of the information necessary to use the PKI, without having to manually download and install certificates, or being forced to trust a large collection of certificates from unknown CAs.

Once the PKIBoot process has completed, the trusted certificates will be present in memory as standard cryptlib trusted certificates (see “Certificate Trust Management” on page 317). To commit them to permanent storage and make them available for future cryptlib sessions, you need to save the cryptlib configuration data as explained in “Working with Trust Settings” on page 318:

```
cryptSetAttribute( CRYPT_UNUSED, CRYPT_OPTION_CONFIGCHANGED, FALSE );
```

If you don’t want to rely on the PKIBoot trusted certificates, don’t commit the configuration data to permanent storage and they’ll be deleted from memory the next time cryptlib is restarted.

At this point the keys are ready for use for encryption, signing, email protection, authentication, and so on. Because of the ease of use provided by the plug-and-play PKI facility, it’s strongly recommended that you use this in place of any other certificate management process, since the alternatives require significantly larger amounts of effort in order to do more or less the same thing.

Simple Certificate Creation

The process of creating a certificate is a rather complicated task that can be somewhat daunting when all you want to do is exchange a public key with someone. In order to simplify the process, cryptlib provides a facility to create simplified certificates that don’t require you to go through all of the steps outlined in the following sections. These simplified certificates are valid for any type of usage (including encryption, signing, use in SSL servers and S/MIME, and issuing other certificates and CRLs) and have a long enough lifetime that you don’t have to worry about them expiring or becoming invalid while you’re still using them.

To create one of these simplified certificates, you set the CRYPT_CERTINFO_XYZZY attribute after creating the certificate object to tell cryptlib to create a simplified certificate, add a name via the CRYPT_CERTINFO_COMMONNAME attribute (and an email address via the CRYPT_CERTINFO_RFC822NAME attribute if you plan to use the certificate for email purposes), and sign it. The name is usually the name of the certificate owner, but if you want to use it with an SSL server then it's the name of the SSL server. For example to create a simplified certificate for Dave Smith you would use:

```
CRYPT_CERTIFICATE cryptCertificate;

/* Create a simplified certificate */
cryptCreateCert( &cryptCertificate, cryptUser,
    CRYPT_CERTTYPE_CERTIFICATE );
cryptSetAttribute( cryptCertificate, CRYPT_CERTINFO_XYZZY, 1 );

/* Add the public key and certificate owner name and sign the
   certificate with the private key */
cryptSetAttribute( cryptCertificate,
    CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO, pubKeyContext );
cryptSetAttributeString( cryptCertificate, CRYPT_CERTINFO_COMMONNAME,
    "Dave Smith", 10 );
cryptSignCert( cryptCertificate, cryptContext );
```

To create a simplified certificate for the SSL server www.sslserver.com you would go through the same steps but give the server name instead of the user's name:

```
/* ... */
cryptSetAttributeString( cryptCertificate, CRYPT_CERTINFO_COMMONNAME,
    "www.sslserver.com", 17 );
/* ... */
```

Finally, if you wanted to use the certificate for email purposes you also need to add the certificate owner's email address:

```
/* ... */
cryptSetAttributeString( cryptCertificate, CRYPT_CERTINFO_RFC822NAME,
    "dave@smith.com", 14 );
/* ... */
```

The same operation in Java or C# is:

```
/* Create a simplified certificate */
int cryptCertificate = crypt.CreateCert( cryptUser,
    crypt.CERTTYPE_CERTIFICATE );
crypt.SetAttribute( cryptCertificate, crypt.CERTINFO_XYZZY, 1 );

/* Add the public key and certificate owner name and sign the
   certificate with the private key */
crypt.SetAttribute( cryptCertificate,
    crypt.CERTINFO_SUBJECTPUBLICKEYINFO, pubKeyContext );
crypt.SetAttributeString( cryptCertificate, crypt.CERTINFO_COMMONNAME,
    "Dave Smith" );
crypt.SignCert( cryptCertificate, cryptContext );
```

The Visual Basic version is:

```
Dim cryptCertificate As Long

' Create a simplified certificate
cryptCreateCert cryptCertificate, cryptUser,
    CRYPT_CERTTYPE_CERTIFICATE
cryptSetAttribute cryptCertificate, CRYPT_CERTINFO_XYZZY, 1

' Add the public key and certificate owner name and
' sign the certificate with the private key
cryptSetAttribute cryptCertificate,
    CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO, _ pubKeyContext
cryptSetAttributeString cryptCertificate, CRYPT_CERTINFO_COMMONNAME, _
    "Dave Smith", 10
cryptSignCert cryptCertificate, cryptContext
```

Since these certificates can be used for any purpose and (effectively) never expire, you can use them without having to worry about certificate requests, communicating with (and paying money to) a CA, proof of possession protocols, X.500 distinguished

names, key usages, certificate extensions, and all the other paraphernalia that comes with X.509 certificates.

In order to distinguish these simplified certificates from normal certificates, cryptlib indicates that they were issued under a simplified-certificate policy using the `certificatePolicies` attribute, which is described in more detail in “Certificate Policies, Policy Mappings, Policy Constraints, and Policy Inhibiting” on page 323.

The Certification Process

Creating a private key and an associated certificate involves two separate processes: generating the public/private key pair, and obtaining a certificate for the public key which is then attached to the public/private key. The key generation process is:

```
generate public/private key pair;
write key pair to keyset;
```

For a crypto device such as a smart card or Fortezza card, the key is generated inside the device, so this step simplifies to:

```
generate public/private key pair;
```

The generated key is already stored inside the device, so there's no need to explicitly write it to any storage media.

The certification process varies somewhat, a typical case has already been presented earlier:

```
create certification request;
submit certification request to CA;
receive certificate from CA;
update keyset or device to include certificate;
```

Now that the general outline has been covered, we can look at the individual steps in more detail. Generating a public/private key pair and saving it to a keyset is relatively simple:

```
CRYPT_CONTEXT cryptContext;
CRYPT_KEYSET cryptKeyset;

/* Create an RSA public/private key context, set a label for it, and
   generate a key into it */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_RSA );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_LABEL,
    "Private key", 11 );
cryptGenerateKey( cryptContext );

/* Save the generated public/private key pair to a keyset */
cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE, fileName,
    CRYPT_KEYOPT_CREATE );
cryptAddPrivateKey( cryptKeyset, cryptContext, password );
cryptKeysetClose( cryptKeyset );

/* Clean up */
cryptDestroyContext( cryptContext );
```

The same operation in Java or C# is:

```
/* Create an RSA public/private key context, set a label for it, and
   generate a key into it */
int cryptContext = crypt.CreateContext( cryptUser, crypt.ALGO_RSA );
crypt.SetAttributeString( cryptContext, crypt.CTXINFO_LABEL, "Private
    key" );
crypt.GenerateKey( cryptContext );

/* Save the generated public/private key pair to a keyset */
int cryptKeyset = crypt.KeysetOpen( cryptUser, crypt.KEYSET_FILE,
    fileName, crypt.KEYOPT_CREATE );
crypt.AddPrivateKey( cryptKeyset, cryptContext, password );
crypt.KeysetClose( cryptKeyset );

/* Clean up */
crypt.DestroyContext( cryptContext );
```

The Visual Basic equivalent is:

```
Dim cryptContext As Long
Dim cryptKeyset As Long

' Create an RSA public/private key context, set a label for it,
' and generate a key into it
cryptCreateContext cryptContext, cryptUser, CRYPT_ALGO_RSA
cryptSetAttributeString cryptContext, CRYPT_CTXINFO_LABEL, _
    "Private key", 11
cryptGenerateKey cryptContext

' Save the generated public/private key pair to a keyset
cryptKeysetOpen cryptKeyset, cryptUser, CRYPT_KEYSET_FILE, filename, _
    CRYPT_KEYOPT_CREATE
cryptAddPrivateKey cryptKeyset, cryptContext, password

' Clean up
cryptKeysetClose cryptKeyset
cryptDestroyContext cryptContext
```

The process for a crypto device is identical except that the keyset write is omitted, since the key is already held inside the device.

In practice you'd probably use **cryptGenerateKeyAsync** so that the user can perform other actions while the key is being generated, although for typical key sizes on a modern PC the key generation is practically instantaneous. If you want to use **cryptGenerateKeyAsync**, you'd run the key generation and the certification request creation in parallel so that by the time the certificate details have been filled in the key is ready for use.

At the same time that you create and save the public/private key pair, you would create a certification request:

```
CRYPT_CERTIFICATE cryptCertRequest;

/* Create a certification request */
cryptCreateCert( &cryptCertRequest, cryptUser,
    CRYPT_CERTTYPE_CERTREQUEST );

/* Fill in the certification request details */
/* ... */

/* Sign the request */
cryptSignCert( cryptCertRequest, cryptContext );
```

The equivalent in Visual Basic is:

```
Dim cryptCertRequest As Long

' Create a certification request
cryptCreateCert cryptCertRequest, cryptUser, _
    CRYPT_CERTTYPE_CERTREQUEST

' Fill in the certification request details
' ...

' Sign the request
cryptSignCert cryptCertRequest, cryptContext
```

The certificate request details vary depending on what you'll want in the certificate that you're requesting. At a minimum, you need to supply the certificate identification information described in "Certificate Identification Information" on page 299 and "Extended Certificate Identification Information" on page 304. Depending on the situation, you may also be able to specify additional certificate components of the type described in "Certificate Extensions" on page 320.

The next step depends on the speed with which the certification request can be turned into a certificate. If the CA's turnaround time is very quick (for example if it's operated in-house) then you can submit the request directly to the CA to convert it into a certificate. In this case you can keep the keyset that you wrote the key to open and update it immediately with the certificate:

```

CRYPT_CERTIFICATE cryptCertificate;

/* Send the certification request to the CA and obtain the returned
   certificate */
/* ... */

/* Import the certificate and check its validity */
cryptImportCert( cert, certLength, cryptUser, &cryptCertificate );
cryptCheckCert( cryptCertificate, caCertificate );

/* Update the still-open keyset with the certificate */
cryptAddPublicKey( cryptKeyset, cryptCertificate );

/* Clean up */
cryptKeysetClose( cryptKeyset );
cryptDestroyCert( cryptCertificate );

```

Again, the Visual Basic equivalent for this is:

```

Dim cryptCertificate As Long

' Send the certification request to the CA and obtain the
' returned certificate
' ...

' Import the certificate and check its validity
cryptImportCert cert, certLength, cryptUser, cryptCertificate
cryptCheckCert cryptCertificate, caCertificate

' Update the still-open keyset with the certificate
cryptAddPublicKey cryptKeyset, cryptCertificate

' Clean up
cryptKeysetClose cryptKeyset
cryptDestroyCert cryptCertificate

```

Since a device acts just like a keyset for certificate updates, you can write a certificate to a device in the same manner.

If, as will usually be the case, the certification turnaround time is somewhat longer, you will need to wait awhile to receive the certificate back from the CA. Once the certificate arrives from the CA, you update the keyset as before:

```

CRYPT_CERTIFICATE cryptCertificate;
CRYPT_KEYSET cryptKeyset;

/* Obtain the returned certificate from the CA */
/* ... */

/* Import the certificate and check its validity */
cryptImportCert( cert, certLength, cryptUser, &cryptCertificate );
cryptCheckCert( cryptCertificate, caCertificate );

/* Open the keyset for update and add the certificate */
cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE, fileName,
    CRYPT_KEYOPT_NONE );
cryptAddPublicKey( cryptKeyset, cryptCertificate );
cryptKeysetClose( cryptKeyset );

/* Clean up */
cryptDestroyCert( cryptCertificate );

```

The Visual Basic equivalent is:

```

Dim cryptCertificate As Long
Dim cryptKeyset As Long

' Obtain the returned certificate from the CA
' ...

' Import the certificate and check its validity
cryptImportCert cert, certLength, cryptUser, cryptCertificate
cryptCheckCert cryptCertificate, caCertificate

```

```
' Open the keyset for update and add the certificate
cryptKeysetOpen cryptKeyset, cryptUser, CRYPT_KEYSET_FILE, fileName, _
    CRYPT_KEYOPT_NONE
cryptAddPublicKey cryptKeyset, cryptCertificate
cryptKeysetClose cryptKeyset

' Clean up
cryptDestroyCert cryptCertificate
```

Again, device updates work in the same manner.

A final case involves self-signed certificates that are typically CA root certificates. Since self-signed CA certificates can be created on the spot, you can immediately update the still-open keyset with the self-signed certificate without any need to go through the usual certification process. When you create a CA certificate you need to set the `CRYPT_CERTINFO_CA` attribute to true (any nonzero value) to indicate that the certificate (and by extension the private key associated with it) is a CA certificate. If you don't do this and then try to sign a certificate using the key, cryptlib will return `CRYPT_ERROR_INVALID` to indicate that the key can't sign certificates because it isn't a CA key. To create a self-signed CA certificate you would do the following:

```
CRYPT_CERTIFICATE cryptCertificate;

/* Create a self-signed CA certificate */
cryptCreateCert( &cryptCertificate, cryptUser,
    CRYPT_CERTTYPE_CERTIFICATE );
cryptSetAttribute( cryptCertificate, CRYPT_CERTINFO_SELSIGNED, 1 );
cryptSetAttribute( cryptCertificate, CRYPT_CERTINFO_CA, 1 );
/* ... */

/* Sign the certificate with the private key and update the still-open
   keyset with it*/
cryptSignCert( cryptCertificate, cryptContext );
cryptAddPublicKey( cryptKeyset, cryptCertificate );

/* Clean up */
cryptKeysetClose( cryptKeyset );
cryptDestroyCert( cryptCertificate );
```

When you sign a certificate for which the `CRYPT_CERTINFO_CA` attribute has been set, cryptlib will enable the key usages `CRYPT_KEYUSAGE_KEYCERTSIGN` and `CRYPT_KEYUSAGE_CRLSIGN` to indicate that the key can be used to sign certificates and CRLs. Since this is a CA key it will by default only be usable for these purposes and not for any other purpose such as encryption or general-purpose signing. You can override this by setting the key usage yourself, however CA keys shouldn't really be used for a purpose other than one or both of certificate and/or CRL signing.

Obtaining Certificates using CMP

The discussion so far has covered the means of communicating with the CA in very general terms. Typically the message exchange is carried out via HTTP or email or through some other, unspecified mechanism. In addition to these very flexible communications options, cryptlib also supports the Certificate Mismanagement Protocol (CMP), which defines a mechanism for communicating with a CA to obtain certificates and request the revocation of existing certificates. CMP makes use of session objects as described in "Secure Sessions" on page 190, the following description assumes that you're familiar with the operation and use of cryptlib session objects.

The general process involved in a CMP session is a two-step one of which the first step is creating the appropriate request, for example a request for a new, updated, or additional certificate or a revocation of an existing certificate, and the second step is submitting it to a CA for processing. The result of the processing (typically a signed certificate) is returned at the end of the session:

```
create a CMP request;
fill in the request details;
sign the request;
```



```

create a CMP session;
add the CMP server address and request type;
add user name and password or signature key;
add the issuing CA's certificate;
add the CMP request;
activate the CMP session;
obtain the result from the CMP session;
destroy the CMP session;

```

The process involved in creating a request for use in CMP is mostly identical to creating a normal certification request (although the formats are incompatible cryptlib hides the details so the programming interface is identical) and is covered below.

cryptlib also implements a full CMP server that allows you to issue certificates using CMP. This process is described in “Managing a CA using CMP or SCEP” on page 261.

CMP Certificate Requests

CMP uses a generic certificate request object to handle requests for new certificates and certificate renewals and updates. The creation of a CMP certificate request of type CRYPT_CERTTYPE_REQUEST_CERT is as follows:

```

CRYPT_CERTIFICATE cryptCMPRequest;

/* Create a certification request */
cryptCreateCert( &cryptCMPRequest, cryptUser,
    CRYPT_CERTTYPE_REQUEST_CERT );

/* Fill in the standard certification request details */
/* ... */

/* Sign the request */
cryptSignCert( cryptCMPRequest, cryptContext );

```

If you’re requesting a new certificate, you generally only need to provide the public key to be certified. Since cryptlib will only copy across the appropriate key components, there’s no need to have a separate public and private key context, you can add the same private key context that you’ll be using to sign the certification request to supply the CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO information and cryptlib will use the appropriate data from it. If the CA doesn’t handle the certificate identification information for you, you’ll also need to provide that. This is rather more complex, and is explained in “Certificate Identification Information” on page 299.

If you’re requesting an update of an existing certificate, you can add information from the existing certificate to the request for use in the new certificate. If you want to renew only the public key in the existing certificate, you should add it as CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO, if you want to renew the entire certificate you should add it as a CRYPT_CERTINFO_CERTIFICATE. For example to renew an entire certificate you would use:

```

CRYPT_CERTIFICATE cryptCMPRequest;

/* Create a certification request and add the existing certificate
details */
cryptCreateCert( &cryptCMPRequest, cryptUser,
    CRYPT_CERTTYPE_REQUEST_CERT );
cryptSetAttribute( cryptCMPRequest, CRYPT_CERTINFO_CERTIFICATE,
    cryptCertificate );

/* Sign the request */
cryptSignCert( cryptCMPRequest, cryptContext );

```

When you add a CRYPT_CERTINFO_CERTIFICATE cryptlib only copies across the public key and certificate owner DN, but not any other attributes such as key usage information (if everything was copied across then the new certificate would be identical to the existing one). This allows you to configure the new certificate in whichever manner you choose, for example to set new or different options from those present in the original certificate.

Requesting the revocation of an existing certificate is very similar to requesting a certificate using a CMP request, the only difference being that the request type is now `CRYPT_CERTTYPE_REQUEST_REVOCATION`. Creating a revocation request involves adding the certificate to be revoked to the request and adding any extra information such as the revocation reason that must be present in the CRL which is issued by the CA:

```
CRYPT_CERTIFICATE cryptCMPRequest;

/* Create a revocation request */
cryptCreateCert( &cryptCMPRequest, cryptUser,
    CRYPT_CERTTYPE_REQUEST_REVOCATION );

/* Fill in the revocation request details */
cryptSetAttribute( cryptCMPRequest CRYPT_CERTINFO_CERTIFICATE,
    certToRevoke );
cryptSetAttribute( cryptCMPRequest, CRYPT_CERTINFO_CRLREASON,
    revocationReason );
```

Note that a revocation request isn't signed since the key required to sign it may not be available any more (loss of the corresponding private key is one of the reasons for revoking a certificate). Once the revocation request has been completed you can submit it to the CA as usual.

CMP Operation Types

The CMP protocol provides for a confusing variety of certificate issue operations with in some cases no clear distinction as to which request type is appropriate for which situation. Because of this, cryptlib will always generate the most generic request type possible, as with other certificate operations it may be necessary to experiment with request types in order to determine the type which is being expected by a CA (some CAs may behave differently for different request types even if the request data is otherwise identical). Since the same uncertainty over which CMP request type to use exists among CAs, it's quite likely that the CAs you'll be interacting with will also accept a variety of requests for a particular situation, so that the generic type generated by cryptlib should work in most cases.

The different CMP certificate request operations are:

Operation	Description
<code>CRYPT_REQUESTTYPE_-INITIALISATION</code>	Initial request to a CA, protected by a user name and password supplied by the CA.
<code>CRYPT_REQUESTTYPE_-CERTIFICATE</code> <code>CRYPT_REQUESTTYPE_-KEYUPDATE</code>	Subsequent requests to the CA, protected by a signature created with an existing CA-certified key. The message contents for these two request types are identical, the only difference is that one is called a certificate request and the other a key update request.
<code>CRYPT_REQUESTTYPE_-REVOCATION</code>	Request for revocation of an existing certificate, protected either by a password supplied by the CA or by a signature created with an existing CA certified key.

When you submit a CMP request, you need to specify the request type before you activate the session. If it's an initialisation or (for some CAs) revocation request the session is authenticated using a user name and password that was previously obtained from the CA. If it's a certificate or key update or (for some CAs) revocation request, the session is authenticated using a signature created with a key that was previously certified by the CA.

Note that some CAs will treat the password which is used during the initialisation stage as a one-time password, so that all subsequent requests have to be signed certificate or key update requests. In addition some CAs require the DN used in subsequent certificates to be the same as the one used in the initialisation request while others don't, some CAs allow a user-specified DN while others require the use

of a fixed DN or set it themselves (overriding the user-supplied value), and some CAs require revocation requests to be protected by a signature rather than a password, which means that if no signature certificate is available (for example you want to revoke a certificate because you've lost the private key, or you have an encryption-only certificate), the certificate can't be revoked. CAs will also perform CA policy-specific operations during the certificate issue process, for example some CAs will automatically revoke a certificate which is superseded by a new one via an update request to prevent a situation in which two otherwise identical certificates exist at the same time.

CMP Sessions

Once a CMP request has been prepared, it's ready for submission to the CA. This is done via a CMP session object, which manages the details of communicating with the CA, authenticating the user, and verifying the data being exchanged. You need to provide the CA server using the CRYPT_SESSINFO_SERVER attribute and either a user name and password using the CRYPT_SESSINFO_USERNAME and CRYPT_SESSINFO_PASSWORD attributes (for an initialisation or revocation request) or a private signing key using the CRYPT_SESSINFO_PRIVATEKEY attribute (for a certificate or key update or revocation request). Finally, you need to provide the certificate of the issuing CA and the request type and data. Once all of this is done, you can activate the session to request the certificate or revocation.

You can submit an initialisation request and obtain an initial certificate from a CA as follows:

```
CRYPT_SESSION cryptSession;

/* Create the CMP session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_CMP );

/* Add the server name/address and request type */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER, server,
    serverLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_CMP_REQUESTTYPE,
    CRYPT_REQUESTTYPE_INITIALISATION );

/* Add the username and password or private signing key. Since this
   is an initialisation request, we add the user name and password */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    userName, userNameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password, passwordLength );

/* Add the certificate of the CA who is to issue the certificate or
   revocation and the request itself */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_CACERTIFICATE,
    cryptCACert );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_REQUEST,
    cryptCmpRequest );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, TRUE );
```

The same operation in Visual Basic is:

```
Dim cryptSession As Long

' Create the CMP session
cryptCreateSession cryptSession, cryptUser, CRYPT_SESSION_CMP

' Add the server name/address and request type
cryptSetAttributeString cryptSession CRYPT_SESSINFO_SERVER, _
    server, Len( server )
cryptSetAttribute cryptSession CRYPT_SESSINFO_CMP_REQUESTTYPE, _
    CRYPT_REQUESTTYPE_INITIALIZATION
```

```

' Add the username and password or private signing key. Since this
' is an initialisation request, we add the user name and password.
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_USERNAME, _
    userName, Len( userName )
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_PASSWORD, _
    password, Len( password )

' Add the certificate of the CA who is to issue the certificate or
' revocation and the request itself
cryptSetAttribute cryptSession, CRYPT_SESSINFO_CACERTIFICATE, _
    cryptCACert
cryptSetAttribute cryptSession, CRYPT_SESSINFO_REQUEST, _
    cryptCmpRequest

' Activate the session
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1

```

If the server that you're communicating with is a cryptlib CMP server, the username and password contain a built-in checksum mechanism which is used by cryptlib to check for data entry errors. If cryptlib returns a CRYPT_ERROR_BADDATA when you set the CRYPT_SESSINFO_USERNAME or CRYPT_SESSINFO_PASSWORD attributes then the user has made a mistake when they entered the name or password. More details on the format and error checking process used for user names and passwords is given in "PKI User IDs" on page 260.

You can submit subsequent certificate or key update requests to obtain further certificates from a CA as follows:

```

CRYPT_SESSION cryptSession;

/* Create the CMP session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_CMP );

/* Add the server name/address and request type */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER, server,
    serverLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_CMP_REQUESTTYPE,
    CRYPT_REQUESTTYPE_CERTIFICATE );

/* Add the username and password or private signing key. Since this
   is a certification request, we add the private key */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY,
    privateKey );

/* Add the certificate of the CA who is to issue the certificate or
   revocation and the request itself */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_CACERTIFICATE,
    cryptCACert );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_REQUEST,
    cryptCmpRequest );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, TRUE );

```

The Java or C# equivalent is:

```

/* Create the CMP session */
int cryptSession = crypt.CreateSession( cryptUser,
    crypt.SESSION_CMP );

/* Add the server name/address and request type */
crypt.SetAttributeString( cryptSession, crypt.SESSINFO_SERVER,
    server );
crypt.SetAttribute( cryptSession, crypt.SESSINFO_CMP_REQUESTTYPE,
    crypt.REQUESTTYPE_CERTIFICATE );

/* Add the username and password or private signing key. Since this is
   a certification request, we add the private key */
crypt.SetAttribute( cryptSession, crypt.SESSINFO_PRIVATEKEY,
    privateKey );

```

```

/* Add the certificate of the CA who is to issue the certificate or
   revocation and the request itself */
crypt.SetAttribute( cryptSession, crypt.SESSINFO_CACERTIFICATE,
    cryptCACert );
crypt.SetAttribute( cryptSession, crypt.SESSINFO_REQUEST,
    cryptCmpRequest );

/* Activate the session */
crypt.SetAttribute( cryptSession, crypt.SESSINFO_ACTIVE, 1 );

```

Submitting a request for a certificate revocation works in an identical manner, with authentication being performed using a user name and password as it is for an initialisation request or a private key as it is for a certification request.

If the session is successfully activated the CMP object will contain the response from the CA, typically a newly-issued certificate. Revocation requests return no data except the status code resulting from the activation of the session. If you're requesting a certificate you can read it from the session as a CRYPT_SESSINFO_RESPONSE attribute:

```

CRYPT_CERTIFICATE cryptCertificate;
int status;

/* Activate the session */
status = cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE,
    TRUE );
if( cryptStatusError( status ) )
    /* Couldn't obtain certificate from CA */;

/* Get the returned certificate */
cryptGetAttribute( cryptSession, CRYPT_SESSINFO_RESPONSE,
    &cryptCertificate );

```

Once you've obtained the certificate, you should save it with the private key it's associated with as described in "Certificates and Keys" on page 235. Because CMP is a complex protocol with a large number of variations and options, it can fail for a variety of reasons. The error-handling techniques described in "Secure Sessions" on page 190 may be useful in determining the exact nature of the problem.

Obtaining Certificates using SCEP

Obtaining a certificate using the Simple Certificate Enrolment Protocol (SCEP) works much like it does for CMP. The general process involved in an SCEP session is a two-step one of which the first step is creating a certification request and the second step is submitting it to a CA for processing. The result of the processing (typically a signed certificate) is returned at the end of the session. SCEP makes use of session objects as described in "Secure Sessions" on page 190, the following description assumes that you're familiar with the operation and use of cryptlib session objects:

```

create a PKCS #10 request;
fill in the request details;

create an SCEP session;
add the SCEP server address;
add user name and password;
add the issuing CA's certificate;
add the PKCS #10 request;
add the private key matching the PKCS #10 request;
activate the SCEP session;
obtain the result from the SCEP session;
destroy the SCEP session;

```

The process involved in creating a request for use in SCEP is mostly identical to the one for CMP, with a few differences as noted below. cryptlib also implements a full SCEP server that allows you to issue certificates using SCEP. This process is described in "Managing a CA using CMP or SCEP" on page 261.

SCEP Certificate Requests

SCEP uses a PKCS #10 certificate request object to handle requests for certificates. The creation of a PKCS #10 certificate request of type CRYPT_CERTTYPE_CERTREQUEST is as follows:

```
CRYPT_CERTIFICATE cryptCertRequest;

/* Create a certification request */
cryptCreateCert( &cryptCertRequest, cryptUser,
    CRYPT_CERTTYPE_CERTREQUEST );

/* Fill in the standard certification request details */
/* ... */
```

Note that, unlike CMP requests, the SCEP request isn't signed. This is because cryptlib has to fill in further details in the request as part of the SCEP message exchange process.

SCEP Sessions

Once a PKCS #10 request has been prepared, it's ready for submission to the CA. This is done via a SCEP session object, which manages the details of communicating with the CA, authenticating the user, and verifying the data being exchanged. You need to specify the CA server and a user name and password using the CRYPT_SESSINFO_SERVER, CRYPT_SESSINFO_USERNAME and CRYPT_SESSINFO_PASSWORD attributes in the usual manner. In addition you need to supply the private key that was used to create the request using the CRYPT_SESSINFO_PRIVATEKEY attribute. The private key is never sent to the server, but is used to for signing and encryption purposes by the SCEP client. Finally, you need to provide the certificate of the issuing CA and the request data. Once all of this is done, you can activate the session to obtain the certificate:

```
CRYPT_SESSION cryptSession;

/* Create the SCEP session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_SCEP );

/* Add the server name/address */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER, server,
    serverLength );

/* Add the username, password, and private key */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_USERNAME,
    userName, userNameLength );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_PASSWORD,
    password, passwordLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY, privateKey
    );

/* Add the certificate of the CA who is to issue the certificate and
the request itself */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_CACERTIFICATE,
    cryptCACert );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_REQUEST, cryptRequest
    );

/* Activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, TRUE );
```

The same operation in Visual Basic is:

```
Dim cryptSession As Long

' Create the SCEP session
cryptCreateSession cryptSession, cryptUser, CRYPT_SESSION_SCEP

' Add the server name/address
cryptSetAttributeString cryptSession CRYPT_SESSINFO_SERVER, _
    server, Len( server )
```

```

' Add the username, password, and private key.
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_USERNAME, _
    userName, Len( userName )
cryptSetAttributeString cryptSession, CRYPT_SESSINFO_PASSWORD, _
    password, Len( password )
cryptSetAttribute cryptSession, CRYPT_SESSINFO_PRIVATEKEY, _
    privateKey

' Add the certificate of the CA who is to issue the certificate and
the request itself
cryptSetAttribute cryptSession, CRYPT_SESSINFO_CACERTIFICATE, _
    cryptCACert
cryptSetAttribute cryptSession, CRYPT_SESSINFO_REQUEST, _
    cryptRequest

' Activate the session
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1

```

If the server that you're communicating with is a cryptlib SCEP server, the username and password contain a built-in checksum mechanism which is used by cryptlib to check for data entry errors. If cryptlib returns a `CRYPT_ERROR_BADDATA` when you set the `CRYPT_SESSINFO_USERNAME` or `CRYPT_SESSINFO_PASSWORD` attributes then the user has made a mistake when they entered the name or password. More details on the format and error checking process used for user names and passwords is given in "Managing a CA using CMP or SCEP" on page 261.

Unlike CMP, SCEP only recognises a basic certification request for a new certificate, so there's no need to specify a request type before you activate the session. In addition, SCEP can only certify keys capable of both encryption and signing, which means that you can only certify RSA keys with no usage restrictions that would limit them to being used only for encryption or only for signing. The returned certificate will contain a combined key usage allowing both encryption and signing.

The SCEP CA certificate must also be capable of encryption and signing, which isn't normally done with a CA certificate but is required by the SCEP protocol. If you add a CA certificate or private key that isn't capable of both encryption and signing, cryptlib will return a `CRYPT_ERROR_PARAM3` to indicate that the CA certificate or key can't be used for SCEP.

If the session is successfully activated the SCEP object will contain the response from the CA, which will be a newly-issued certificate that you can read from the session as a `CRYPT_SESSINFO_RESPONSE` attribute:

```

CRYPT_CERTIFICATE cryptCertificate;
int status;

/* Activate the session */
status = cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, TRUE
    );
if( cryptStatusError( status ) )
    /* Couldn't obtain certificate from CA */;

/* Get the returned certificate */
cryptGetAttribute( cryptSession, CRYPT_SESSINFO_RESPONSE,
    &cryptCertificate );

```

Once you've obtained the certificate, you should save it with the private key it's associated with as described in "Certificates and Keys" on page 235. Because SCEP is a complex protocol with a large number of variations and options, it can fail for a variety of reasons. The error-handling techniques described in "Secure Sessions" on page 190 may be useful in determining the exact nature of the problem.

Certificate Status Checking using RTCS

In order to check the validity of a certificate, cryptlib supports the real-time certificate status protocol (RTCS). The simplest way to use RTCS is with `cryptCheckCert`, which returns a straightforward valid/not valid status and is described in the next section. More complex RTCS usage, including obtaining detailed status information

and querying the status of multiple certificates at once is covered in the sections that follow.

Basic RTCS Queries

The simplest way to work with RTCS is to use it with **cryptCheckCert** to check the validity of a certificate. Since RTCS is an online protocol, communicating with the responder requires the use of a cryptlib session object which is described in more detail in “Secure Sessions” on page 190, the following description assumes that you’re familiar with the operation and use of cryptlib session objects. Establishing an RTCS client session requires adding the RTCS responder name or IP address and an optional port number if it isn’t using the standard port. Once this is done, you can check the certificate using **cryptCheckCert**, with the second parameter being the RTCS responder.

```
CRYPT_SESSION cryptSession;
int status;

/* Create the RTCS session and add the responder name */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_RTCS );
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
    serverName, serverNameLength );

/* Check the certificate */
status = cryptCheckCert( cryptCertificate, cryptSession );
if( cryptStatusOK( status ) )
    /* Certificate is OK */;

/* Clean up the session object */
cryptDestroySession( cryptSession );
```

Note that the RTCS session isn’t activated in the usual manner by setting the **CRYPT_SESSINFO_ACTIVE** attribute to true, since this is done by **cryptCheckCert** when it performs the validity check.

If **cryptCheckCert** returns OK this means that the certificate is valid right now. If it returns **CRYPT_ERROR_INVALID** (or some other error) the certificate isn’t valid, either because it has expired, has been revoked, is a forged certificate, or for some other reason. Usually all that matters is whether a certificate is OK to use or not, but if you require detailed information as to why a certificate isn’t OK to use you need to perform a manual RTCS check without the help of **cryptCheckCert**, as described below.

Creating an RTCS Request

Performing an RTCS status check without the help of **cryptCheckCert** involves creating an RTCS request object, adding a copy of the certificate to be checked to the request, submitting the request to the RTCS responder and receiving the responder’s reply, and finally checking the certificate’s status in the RTCS reply:

```
create RTCS request;
add certificate to be checked to request;
exchange data with RTCS responder;
check certificate using RTCS response;
```

An RTCS request is a standard certificate object of type **CRYPT_CERTTYPE_RTCS_REQUEST**. You create this in the usual manner and add the certificate as a **CRYPT_CERTINFO_CERTIFICATE** attribute. Since RTCS queries don’t have to be signed, there’s no need to perform any further operations on the request object, and it’s ready for submission to the responder:

```
CRYPT_CERTIFICATE cryptRTCSRequest;

/* Create the RTCS request */
cryptCreateCert( &cryptRTCSRequest, cryptUser,
    CRYPT_CERTTYPE_RTCS_REQUEST );

/* Add the certificate to be queried to the request */
cryptSetAttribute( cryptRTCSRequest, CRYPT_CERTINFO_CERTIFICATE,
    cryptCertificate );
```


Sometimes a user's certificate will contain the information required for cryptlib to communicate with the responder, but often this is missing or incorrect. You can check for the presence of RTCS information in the certificate by checking for the existence of the CRYPT_CERTINFO_AUTHORITYINFO_RTCS attribute, which contains information about the RTCS responder, usually in the form of a URL. If you want to read the location of the responder, you can obtain it by reading the CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER attribute from within the RTCS information. Since the RTCS attribute is a composite GeneralName field, you need to first select it and then read the URL from within the GeneralName:

```
char url[ CRYPT_MAX_TEXTSIZE + 1 ];
int urlLength;

cryptSetAttribute( cryptCertificate,
    CRYPT_CERTINFO_AUTHORITYINFO_RTCS, CRYPT_UNUSED );
cryptGetAttributeString( cryptCertificate,
    CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER, url, &urlLength );
url[ urlLength ] = '\0';
```

If the RTCS responder location isn't present or is incorrect, you'll need to add the responder URL manually before you can submit the request, as explained in the next section.

Communicating with an RTCS Responder

Since RTCS is an online protocol, communicating with the responder requires the use of a cryptlib session object which is described in more detail in "Secure Sessions" on page 190, the following description assumes that you're familiar with the operation and use of cryptlib session objects. If the name of the RTCS responder is specified in the certificate which is being checked you can directly submit the request to an RTCS session object as a CRYPT_SESSINFO_REQUEST attribute without requiring any further setup of the session object. If the responder isn't specified in the certificate, you'll have to specify it yourself as described further on. In either case cryptlib will contact the responder, submit the status query, and obtain the response from the responder. If the query was successful, the session object will contain the RTCS response object in the form of a CRYPT_SESSINFO_RESPONSE that contains the reply from the server:

```
CRYPT_SESSION cryptSession;
CRYPT_CERTIFICATE cryptRTCSResponse;
int status;

/* Create the RTCS session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_RTCS );

/* Add the RTCS request and activate the session with the RTCS
responder */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_REQUEST,
    cryptRTCSRequest );
status = cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE,
    TRUE );
if( cryptStatusError( status ) )
    /* Couldn't establish session with RTCS responder */;

/* Clean up the RTCS request object, which isn't needed any more */
cryptDestroyCert( cryptRTCSRequest );

/* Obtain the response information */
status = cryptGetAttribute( cryptSession, CRYPT_SESSINFO_RESPONSE,
    &cryptRTCSResponse );
if( cryptStatusError( status ) )
    /* No response available from responder */;

/* Clean up the session object */
cryptDestroySession( cryptSession );
```

Once you've got the response from the server, you can get the certificate status from it by reading the CRYPT_CERTINFO_CERTSTATUS attribute:

```
int certStatus;

cryptGetAttribute( cryptRTCSResponse, CRYPT_CERTINFO_CERTSTATUS,
                  &certStatus );
if( certStatus == CRYPT_CERTSTATUS_VALID )
    /* Certificate is valid */;

/* Clean up the RTCS response */
cryptDestroyCert( cryptRTCSResponse );
```

The possible certificate status values are `CRYPT_CERTSTATUS_VALID`, `CRYPT_CERTSTATUS_NOTVALID`, and `CRYPT_CERTSTATUS_UNKNOWN`, with obvious meanings.

As mentioned above, you may need to set the RTCS responder URL if it isn't present in the certificate or if the value given in the certificate is incorrect. You can set the responder URL as the `CRYPT_SESSINFO_SERVER_NAME`:

```
CRYPT_SESSION cryptSession;

/* Create the RTCS session */
cryptCreateSession( &cryptSession, cryptUser, CRYPT_SESSION_RTCS );

/* Add the responder URL and RTCS request and activate the session
   with the RTCS responder */
cryptSetAttributeString( cryptSession, CRYPT_SESSINFO_SERVER_NAME,
                        serverName, serverNameLength );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_REQUEST,
                  cryptRTCSRequest );
/* ... */
```

Advanced RTCS Queries

In addition to querying the status of individual certificates, you can query the status of a number of certificates at once by adding more than one certificate to the RTCS request. The response will contain information for each certificate in the query, which you can use to verify each certificate using **`cryptCheckCert`**. If the response information indicates that the certificate is invalid, cryptlib will return `CRYPT_ERROR_INVALID` and leave the entry for the certificate in the RTCS response as the selected one, allowing you to obtain further information about the certificate if any is available:

```
CRYPT_CERTIFICATE cryptRTCSResponse;
time_t revocationTime;
int revocationReason;

/* Check the certificate against the RTCS response */
cryptCheckCert( cryptCertificate, cryptRTCSResponse );
if( status == CRYPT_ERROR_INVALID )
{
    int revocationTimeLength;

    /* The certificate has been revoked, get the revocation time and
       reason */
    cryptGetAttributeString( cryptRTCSResponse,
                            CRYPT_CERTINFO_REVOCATIONDATE, &revocationTime,
                            &revocationTimeLength );
    cryptGetAttribute( cryptRTCSResponse, CRYPT_CERTINFO_CRLREASON,
                      &revocationReason );
}
```

If all you're interested in is an overall validity indication for a collection of certificates then an alternative technique that doesn't require calling **`cryptCheckCert`** for each certificate is to step through the responses using the extension cursor management, checking the status for each certificate and recording whether any one indicates that the certificate is invalid:

```

int certsValid = TRUE;

cryptSetAttribute( cryptRTCSResponse,
    CRYPT_CERTINFO_CURRENT_CERTIFICATE, CRYPT_CURSOR_FIRST );
do
{
    int certStatus;

    /* Check the status of the currently selected certificate */
    cryptGetAttribute( cryptRTCSResponse, CRYPT_CERTINFO_CERTSTATUS,
        &certStatus );
    if( certStatus != CRYPT_CERTSTATUS_VALID )
        certsValid = FALSE;
}
while( certsValid &&
    cryptSetAttribute( cryptRTCSResponse,
        CRYPT_CERTINFO_CURRENT_CERTIFICATE, CRYPT_CURSOR_NEXT ) ==
        CRYPT_OK );

if( !certsValid )
    /* At least one certificate is invalid */;

```

This will step through all of the responses checking for an indication that a certificate is invalid. Once the loop terminates, the `certsValid` variable will contain the composite status of the complete set of certificates.

Certificate Revocation Checking using OCSP

In order to check whether a certificate is present in a CRL, cryptlib supports the online certificate status protocol (OCSP). Unlike RTCS, OCSP can't be used with **cryptCheckCert**, requiring the use of the more complex interface described below. Note that OCSP doesn't return a proper certificate status (it can't truly determine whether a certificate is really valid), and will often return a response based on out-of-date CRL information. If you require a true online certificate validity check, you should use the real-time certificate status protocol (RTCS) as described in "Certificate Status Checking using RTCS" on page 247.

Creating an OCSP Request

OCSP requests work just like RTCS requests described in "Creating an RTCS Request" on page 250, except that the request type is `CRYPT_CERTTYPE_OCSP_REQUEST` instead of `CRYPT_CERTTYPE_RTCS_REQUEST`, however in addition to the certificate being queried an OCSP request also needs to have the CA certificate that issued the certificate being queried added to the request before the certificate itself is added. The CA certificate is added as a `CRYPT_CERTINFO_CACERTIFICATE` attribute:

```

CRYPT_CERTIFICATE cryptOCSPRequest;

/* Create the OCSP request */
cryptCreateCert( &cryptOCSPRequest, cryptUser,
    CRYPT_CERTTYPE_OCSP_REQUEST );

/* Add the certificate to be queried and the CA certificate that
   issued it to the request */
cryptSetAttribute( cryptOCSPRequest, CRYPT_CERTINFO_CACERTIFICATE,
    cryptCACert );
cryptSetAttribute( cryptOCSPRequest, CRYPT_CERTINFO_CERTIFICATE,
    cryptCertificate );

```

As with RTCS requests, the certificate being queried may contain responder details in the `CRYPT_CERTINFO_AUTHORITYINFO_OCSP` attribute, or you may need to add them manually as explained in "Creating an RTCS Request" on page 250.

OCSP requests can also be signed, if you're working with a CA that uses this capability then you can sign the request before submitting it in the standard way using **cryptSignCert**:

```
CRYPT_CERTIFICATE cryptOCSPRequest;

/* Create the OCSF request */
cryptCreateCert( &cryptOCSPRequest, cryptUser,
    CRYPT_CERTTYPE_OCSP_REQUEST );

/* Add the certificate to be queried to the request and sign it */
cryptSetAttribute( cryptOCSPRequest, CRYPT_CERTINFO_CERTIFICATE,
    cryptCertificate );
cryptSignCert( cryptOCSPRequest, privateKey );
```

OCSP requests can also include signing certificates alongside the signature, you can specify the amount of additional information to include with the signature by setting the `CRYPT_CERTINFO_SIGNATURELEVEL` attribute as described in “Signing/Verifying Certificates” on page 308.

Communicating with an OCSP Responder

Communicating with an OCSP responder works in exactly the same way as communicating with an RTCS responder described in “Communicating with an RTCS Responder” on page 251, except that the session type is `CRYPT_SESSION_OCSP` rather than `CRYPT_SESSION_RTCS`. Once you’ve successfully activated the session, you can read the certificate revocation status from the returned OCSP response by reading the `CRYPT_CERTINFO_REVOCATIONSTATUS` attribute:

```
int revocationStatus;

cryptGetAttribute( cryptOCSPResponse, CRYPT_CERTINFO_REVOCATIONSTATUS,
    &revocationStatus );
if( revocationStatus == CRYPT_OCSPSTATUS_NOTREVOKED )
    /* Certificate hasn't been revoked */;

/* Clean up the OCSP response */
cryptDestroyCert( cryptOCSPResponse );
```

The possible certificate status values are `CRYPT_OCSPSTATUS_NOTREVOKED`, `CRYPT_OCSPSTATUS_REVOKED`, and `CRYPT_OCSPSTATUS_UNKNOWN`. Note that since OCSP is purely a revocation checking protocol, `CRYPT_OCSPSTATUS_NOTREVOKED` means exactly that, that the certificate hasn’t been revoked. This doesn’t mean the same as saying that the certificate is OK, a bogus certificate that exists but isn’t recognised by the CA as having been issued (for example a forged certificate created by an attacker), or an expired certificate, or a certificate which is invalid for some other reason or isn’t even a certificate (for example an Excel spreadsheet) would also be given a status of “not revoked” since that’s all that the responder is capable of saying about it. In addition OCSP responders are often fed from stale CRL information, so a not-revoked response doesn’t necessarily mean that the certificate is really not revoked, merely that at the time the information was last updated it hadn’t been revoked. OCSP is purely an online CRL query mechanism, not a general-purpose certificate validity checker.

In addition to the certificate status, the OCSP response also contains information relating to the CRL that the responder used to create the response, including `CRYPT_CERTINFO_THISUPDATE`, the time of the current CRL, an optional `CRYPT_CERTINFO_NEXTUPDATE`, the time of the next CRL, and `CRYPT_CERTINFO_REVOCATIONDATE`, the time at which the certificate was revoked. If the OCSP responder is using a direct query of a certificate store rather than assembling the information indirectly using CRLs then the current CRL time will usually be set to the current time even if it’s assembled from stale information hours or days old. In addition the next update time may be set to the current time, or to a future time. None of these fields are particularly useful and different CAs assign different meanings to them, so they can be ignored in most circumstances, they relate mainly to the CRL-based origins of certain portions of OCSP. In addition, while RTCS uses times relative to the local system time, OCSP uses the absolute time on the responder, so time values will vary based on time differences between the OCSP responder and the local machine.

Advanced OCSP Queries

Some OCSP responders can resolve multiple certificate status queries in a single request, however because of the data format used in OCSP this doesn't work properly for OCSP version 1 responders so it's better to submit a number of separate queries rather than trying to query the status of a set of certificates in a single request. In addition some responders can't handle multiple certificates, or will ignore all but the first certificate, making it even more advisable to restrict queries to a single certificate. Although a planned future revision of OCSP may not have this problem, it's still prudent to only query a single certificate per request unless you're sure that the responder you're using will handle multi-certificate queries correctly.

If you submit a query containing multiple certificates, the response from the responder constitutes a mini-CRL that contains revocation information only for the certificates submitted in the query (assuming that the responder can handle multiple certificates in a query). Because of this you can treat the response as if it were a normal CRL and check the certificates you submitted against it with **cryptCheckCert** just like a CRL. If the certificate has been revoked, cryptlib will return **CRYPT_ERROR_INVALID** and leave the certificate's revocation entry in the OCSP response as the selected one, allowing you to obtain further information on the revocation (for example the revocation date or reason):

```
CRYPT_CERTIFICATE cryptOCSPResponse;
time_t revocationTime;
int revocationReason;

/* Check the certificate against the OCSP response */
cryptCheckCert( cryptCertificate, cryptOCSPResponse );
if( status == CRYPT_ERROR_INVALID )
{
    int revocationTimeLength;

    /* The certificate has been revoked, get the revocation time and
       reason */
    cryptGetAttributeString( cryptOCSPResponse,
        CRYPT_CERTINFO_REVOCATIONDATE, &revocationTime,
        &revocationTimeLength );
    cryptGetAttribute( cryptOCSPResponse, CRYPT_CERTINFO_CRLREASON,
        &revocationReason );
}
```

Note that, as with standard CRLs, the revocation reason is an optional component and may not be present in the OCSP response. If the revocation reason isn't present, cryptlib will return **CRYPT_ERROR_NOTFOUND**. If all you're interested in is a revoked/not revoked status for a collection of certificates then you can step through the responses checking the status for each one in turn in the same way as for RTCS.

Managing a Certification Authority

Although it's possible to manually manage the operation of a CA and issue certificates and CRLs using **cryptSignCert**, it's much easier to use cryptlib's built-in CA management capabilities to do this for you. In order to use the CA management capabilities you need to create a certificate store as explained in "Creating/Destroying Keyset Objects" on page 219. The keyset type for a certificate store can only be `CRYPT_KEYSET_DATABASE_STORE`, `CRYPT_KEYSET_ODBC_STORE`, or `CRYPT_KEYSET_PLUGIN_STORE`, since cryptlib requires a full relational database with transaction processing capabilities in order to manage the CA operations. The use of a transaction-capable certificate store results in a high degree of scalability and provides the level of reliability, availability, and error recovery required of such an application and stipulated in a number of standards that cover CA operation.

Once you've created a certificate store, you can open a connection to it like a normal keyset. Since all accesses that open the keyset for write access are logged, it's better to open the connection to the keyset once and then leave it open for ongoing operations than to open and close it for each operation, since this would lead to an excessive number of log entries.

A certificate store doesn't work like a standard keyset in which it's possible to insert and delete certificates and CRLs at random. Instead, it's used in combination with various certificate management functions that use the certificate store as a mechanism for managing the operations performed by a CA. The CA operations consist of recording incoming certificate requests, converting them into certificates, and issuing CRLs for revoked certificates. All of these operations are managed automatically for you by cryptlib using the transaction processing capabilities of the certificate store to handle the data storage, reliability, and auditing requirements of the CA.

There are two ways in which you can run a CA. The easiest option is to use cryptlib's built-in CMP or SCEP servers to handle all CA operations. The more complex option is to use cryptlib's CA management functions to handle the CA operations yourself. Of the two CA management protocols, CMP is the more complete, allowing you to request new certificates, update/replace existing ones, and revoke existing certificates, works with special-purpose certificates such as signing-only or encryption-only types, and provides flexibility in the authorisation mechanisms used, with the request authorised either with a user name and password or signed with an existing certificate. SCEP on the other hand is a relatively simple protocol that allows for a single type of operation, issuing a new certificate, and a single certificate type, an RSA certificate capable of both encryption and signing, with the request authorised with a user name and password.

Before you begin you'll need to decide which of the two best meets your needs. Usually it'll be CMP, which is more flexible than SCEP. Alternatively, you can run both a CMP and SCEP server, although you'll have to run them on different ports since both protocols use HTTP for their communications.

Creating the Top-level (Root) CA Key

The first thing that you need to do when you set up your CA is to create your top-level (root) CA key. This involves creating the public/private key pair, adding identification information to it, signing it to create the CA root certificate, and optionally storing it to disk if you're not holding it in a crypto token such as a smart card or hardware security module (HSM). You can generate the root CA key as follows:

```
CRYPT_CONTEXT cryptContext;

/* Create an RSA public/private key context, set a label for it, and
   generate a key into it */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_RSA );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_LABEL,
    "Private key", 11 );
cryptGenerateKey( cryptContext );
```

More details on keys and key generation are given in “Key Generation and Storage” on page 216.

Once you’ve generated the key, you can create the root CA certificate and add the CA’s identification information to it, which usually consists of the country, organisation name, organisational unit name, and finally the actual CA name, referred to as the common name in PKI terminology:

```
CRYPT_CERTIFICATE cryptCertificate;

/* Create the CA certificate and add the public key */
cryptCreateCert( &cryptCertificate, cryptUser,
    CRYPT_CERTTYPE_CERTIFICATE );
cryptSetAttribute( cryptCertificate,
    CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO, cryptContext );

/* Add identification information */
cryptSetAttributeString( cryptCertificate, CRYPT_CERTINFO_COUNTRYNAME,
    countryName, 2 );
cryptSetAttributeString( cryptCertificate,
    CRYPT_CERTINFO_ORGANIZATIONNAME, organizationName,
    organizationNameLength );
cryptSetAttributeString( cryptCertificate,
    CRYPT_CERTINFO_ORGANIZATIONALUNITNAME, organizationalUnitName,
    organizationalUnitNameLength );
cryptSetAttributeString( cryptCertificate, CRYPT_CERTINFO_COMMONNAME,
    commonName, commonNameLength );
```

More details on certificate naming are given in “Certificate Identification Information” on page 299.

Once the CA name is set, you need to mark the certificate as a self-signed CA certificate:

```
cryptSetAttribute( cryptCertificate, CRYPT_CERTINFO_SELFSIGNED, 1 );
cryptSetAttribute( cryptCertificate, CRYPT_CERTINFO_CA, 1 );
```

Finally, you may want to add two URLs that indicate to users where further CA services may be found, in particular CRYPT_CERTINFO_AUTHORITYINFO_-CERTSTORE to tell users where to go to find further certificates and CRYPT_CERTINFO_AUTHORITYINFO_RTCS to tell users where to go for real-time certificate status information: Since these attributes are a composite GeneralName field, you need to first select them and then add the URL as a CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER attribute within the GeneralName:

```
cryptSetAttribute( cryptCertificate,
    CRYPT_CERTINFO_AUTHORITYINFO_CERTSTORE, CRYPT_UNUSED );
cryptSetAttributeString( cryptCertificate,
    CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER, certstoreUrl,
    certstoreUrlLength );
cryptSetAttribute( cryptCertificate,
    CRYPT_CERTINFO_AUTHORITYINFO_RTCS, CRYPT_UNUSED );
cryptSetAttributeString( cryptCertificate,
    CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER, rtcsUrl, rtcsUrlLength );
```

With the URLs present in the resulting certificate, users will automatically know where to go to obtain further certificate-related information.

You can also set these URLs on a per-user basis when you set up each user’s information, although putting it in the CA certificate allows you to set it just once without having to set it up for each user (cryptlib will automatically propagate it from the CA certificate to the user certificates when they’re issued). More details on certificate store access are given in “HTTP Keysets” on page 221, and details on real-

time certificate status checking are given in “Certificate Status Checking using RTCS” on page 249.

Your root CA certificate is now ready to be signed:

```
cryptSignCert( cryptCertificate, cryptContext );
```

If you’re storing the CA information on disk, you now need to save the keys and certificates to a password-protected private-key file:

```
CRYPT_KEYSET cryptKeyset;

/* Save the generated public/private key pair to a keyset */
cryptKeysetOpen( &cryptKeyset, cryptUser, CRYPT_KEYSET_FILE, fileName,
    CRYPT_KEYOPT_CREATE );
cryptAddPrivateKey( cryptKeyset, cryptContext, password );
cryptAddPublicKey( cryptKeyset, cryptCertificate );
cryptKeysetClose( cryptKeyset );

/* Clean up */
cryptDestroyContext( cryptContext );
cryptDestroyCert( cryptCertificate );
```

If you’re storing the information in a crypto device, the keys will already be in the device, and all you need to do is update it with the newly-created certificate:

```
cryptAddPublicKey( cryptDevice, cryptCertificate );

/* Clean up */
cryptDestroyCert( cryptCertificate );
```

At this point your root CA key is ready to use for issuing certificates.

Initialising PKI User Information

In order to be able to issue certificates to an end user (called a PKI user in CMP terminology), cryptlib first needs to know various pieces of information about them. You supply this information via a PKI user certificate object, providing a partial or complete DN for the issued certificate, as well as any other information that’s required for the certificate such as an email address or URL, an indication as to whether the user is a CA capable of issuing their own certificates, and so on. Once you’ve provided the information for the PKI user, you add it to the certificate store that will be used by the CMP or SCEP CA session, after which the CA server will consult the certificate store when it needs to issue a certificate. cryptlib will automatically generate the user ID and password for you when you’ve finished creating the PKI user object.

When you add the DN information to the PKI user object, you can specify either a complete DN or a partial DN that omits the user’s common name. The PKI user object acts both as a template for the DN in the user’s certificate and as a constraint on the actual DN that a user can choose, preventing them from choosing an arbitrary DN for their certificate. It’s strongly recommended that you specify the user’s full DN in the PKI user object, so that they aren’t required to know the DN but can simply submit a request and have the CA take care of assigning a DN for them.

Alternatively, you can specify all DN components except the common name and let the user specify the common name in the request. The least preferable option, since it both requires that the user know their full DN and specify it in the request, and allows them to request any type of DN, is to omit setting a DN in the PKI user object, which allows the user to specify any DN value. However, omitting the DN from the PKI user template can lead to problems later if you want to read the PKI user object back from the certificate store, since there’s no name present to identify it.

Taking the simplest option, in which the CA supplies the full DN and the user doesn’t need to know any DN details, you would use:

```
CRYPT_CERTIFICATE cryptPKIUser;

/* Create the PKI user */
cryptCreateCert( &cryptPKIUser, cryptUser, CRYPT_CERTTYPE_PKIUSER );
```



```

/* Add identification information */
cryptSetAttributeString( cryptPKIUser, CRYPT_CERTINFO_COUNTRYNAME,
    countryName, 2 );
cryptSetAttributeString( cryptPKIUser,
    CRYPT_CERTINFO_ORGANIZATIONNAME, organizationName,
    organizationNameLength );
cryptSetAttributeString( cryptPKIUser,
    CRYPT_CERTINFO_ORGANIZATIONALUNITNAME, organizationalUnitName,
    organizationalUnitNameLength );
cryptSetAttributeString( cryptPKIUser, CRYPT_CERTINFO_COMMONNAME,
    commonName, commonNameLength );

/* Add the user information to the certificate store */
cryptCAAddItem( cryptCertStore, cryptPKIUser );

/* Clean up */
cryptDestroyCert( cryptPKIUser );

```

The same operation in Visual Basic is:

```

Dim cryptPKIUser As Long

' Create the PKI user
cryptCreateCert cryptPKIUser, cryptUser, CRYPT_CERTTYPE_PKIUSER

' Add identification information
cryptSetAttributeString cryptPKIUser, CRYPT_CERTINFO_COUNTRYNAME, _
    countryName, 2
cryptSetAttributeString cryptPKIUser, _
    CRYPT_CERTINFO_ORGANIZATIONNAME, organizationName, _
    organizationNameLength
cryptSetAttributeString cryptPKIUser, _
    CRYPT_CERTINFO_ORGANIZATIONALUNITNAME, organizationalUnitName, _
    organizationalUnitNameLength
cryptSetAttributeString cryptPKIUser, CRYPT_CERTINFO_COMMONNAME, _
    commonName, commonNameLength

' Add the user information to the certificate store
cryptCAAddItem cryptCertStore, cryptPKIUser

' Clean up
cryptDestroyCert cryptPKIUser

```

A simple way to handle this type of operation is to automatically populate the certificate store with information from a source such as a personnel database containing all of the required user information.

Other PKI User Information

In addition to the user DN, you can may also want to add further information to allow the user to automatically locate resources such as further certificates issued by the CA and RTCS responders. By adding these URLs to the PKI user information (which ensures that it'll be present in the certificate once it's issued), anyone using the certificate can automatically determine where to go to find further certificates and certificate status information without requiring any manual configuration.

The easiest way to get this information into user certificates is to add it to the issuing CA's certificate, from which it'll be automatically propagated into any certificates that the CA issues. You can however also add this information on a per-user basis as the CRYPT_CERTINFO_AUTHORITYINFO_CERTSTORE and CRYPT_CERTINFO_AUTHORITYINFO_RTCS attributes, which contain information about the location of the certificate store and RTCS responder, usually in the form of a URL. Since these attributes are composite GeneralName fields, you need to first select them and then add the URL as a CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER attribute within the GeneralName:

```
cryptSetAttribute( cryptPKIUser,
    CRYPT_CERTINFO_AUTHORITYINFO_CERTSTORE, CRYPT_UNUSED );
cryptSetAttributeString( cryptPKIUser,
    CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER, certstoreUrl,
    certstoreUrlLength );
cryptSetAttribute( cryptPKIUser, CRYPT_CERTINFO_AUTHORITYINFO_RTCS,
    CRYPT_UNUSED );
cryptSetAttributeString( cryptPKIUser,
    CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER, rtcsUrl, rtcsUrlLength );
```

With the URL present in the resulting certificate, users will automatically know where to go to obtain further certificates and certificate status information.

In addition to the CA-related information, you can also specify additional user information that will appear in the issued certificate. The most common additional information would be an email address that's used to identify the user alongside their DN:

```
cryptSetAttributeString( cryptPKIUser, CRYPT_CERTINFO_RFC822NAME,
    emailAddr, emailAddrLength );
```

although since this may change over time you may want to let the user specify it in their certificate request. A downside of this flexibility is that the user can then request a certificate with any email address they want rather than the one that you've got recorded for them.

In addition to the standard identification information, you can also specify other information that should appear in all certificates issued to this particular user. One piece of certificate information that can *only* be specified in the PKI user data is whether the user is to be a CA or not. To create a CA user, you set the CA flag for the user:

```
cryptSetAttribute( cryptPKIUser, CRYPT_CERTINFO_CA, 1 );
```

This is the only way in which a CA certificate can be issued, since allowing a user to specify the issuing of a CA certificate in a user request would allow any user to make themselves a CA. If cryptlib receives a request from a user for the creation of a CA certificate it will either reject the request, since the CA capability can only be permitted by the issuing CA and not the requesting user.

Because a CA-enabled user has special privileges, you should take extra care in managing passwords and related information for them, and may want to delete the user after their CA certificate has been issued to prevent them from being re-used to obtain further CA certificates. This makes the sub-CA creation capability a one-shot process that requires explicit manual intervention by the issuing CA every time a sub-CA is created.

PKI User IDs

Certificate initialisation requests are identified through a user ID (to locate the appropriate PKI user information) and a password (to authenticate the request). Once the user information has been entered into the certificate store, you can read back the PKI user ID, identified by CRYPT_CERTINFO_PKIUSER_ID, the password used to authenticate the initialisation operation, identified by CRYPT_CERTINFO_PKIUSER_ISSUEPASSWORD, and the password used to authenticate certificate revocation (if you're using CMP), CRYPT_CERTINFO_PKIUSER_REVPASSWORD. Use of the revocation password is optional, the CA may use signed revocation requests rather than password-protected ones:

```
char userID[ CRYPT_MAX_TEXTSIZE + 1 ];
char issuePW[ CRYPT_MAX_TEXTSIZE + 1 ];
char revPW[ CRYPT_MAX_TEXTSIZE + 1 ];
int userIDlength, issuePWlength, revPWlength;

cryptGetAttributeString( cryptPKIUser, CRYPT_CERTINFO_PKIUSER_ID,
    userID, &userIDlength );
userID[ userIDlength ] = '\0';
cryptGetAttributeString( cryptPKIUser,
    CRYPT_CERTINFO_PKIUSER_ISSUEPASSWORD, issuePW, &issuePWlength );
issuePW[ issuePWlength ] = '\0';
```

```
cryptGetAttributeString( cryptPKIUser,
    CRYPT_CERTINFO_PKIUSER_REVPASSWORD, revPW, &revPWlength );
revPW[ revPWlength ] = '\0';
```

The CA needs to communicate this information to the user via some out-of-band means, typically through the use of a PIN mailer or via some other direct communication means during the certificate sign-up process. Once this information is communicated, the user can use it to obtain their initial certificate. Any further certificates are typically obtained by signing the request with the initial certificate or with subsequently-obtained certificates.

cryptlib uses a standard format for the user ID and password that follows the style used for software registration codes and serial numbers. The user ID is in the form XXXXX-XXXXX-XXXXX and the password is in the form XXXXX-XXXXX-XXXXX-XXXXX. Characters that might cause confusion (for example O and 0 or I and l) aren't present, and the data contains a checksum which is used to catch typing errors when the user enters the information. An example of a user ID and password is:

```
user ID = 293XU-NZMSN-DC5J3
password = G3DKZ-DR79M-L6AGY-X6H6X
```

If the user enters either of these incorrectly, the cryptlib client will return `CRYPT_ERROR_BADDATA` when you try to set the user name or password attribute for the CMP or SCEP client session.

Managing a CA using CMP or SCEP

CMP and SCEP servers that allow you to issue certificates to a CMP or SCEP client make use of session objects as described in “Secure Sessions” on page 190, the following description assumes that you're familiar with the operation and use of cryptlib session objects. Once the PKI user information has been set up for each user, there isn't anything further that needs to be done. Because the CA management process is completely automated and entirely handled by cryptlib, the CA more or less runs itself. The only operations that you still need to perform yourself are periodic ones such as expiring old certificates with `CRYPT_CERTACTION_EXPIRE_CERT`, issuing CRLs with `CRYPT_CERTACTION_ISSUE_CRL` (assuming you're not using the much more sensible option of allowing online queries of the certificate store which is used by the CA), and handling restart recoveries with `CRYPT_CERTACTION_CLEANUP` (the manual certificate management operations are described in “CA Management Operations” on page 265). All other operations are handled for you by the CMP or SCEP server.

Establishing a CMP or SCEP server session requires adding the CA certificate store and CA server key/certificate as the `CRYPT_SESSINFO_KEYSET` and `CRYPT_SESSINFO_PRIVATEKEY` attributes, activating the session, and waiting for incoming connections. The CMP server session is denoted by `CRYPT_SESSION_CMP_SERVER`, the SCEP server session is denoted by `CRYPT_SESSION_SCEP_SERVER`:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_CMP_SERVER );

/* Add the CA certificate store and CA server key and activate the
   session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_KEYSET,
    cryptCertStore );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_PRIVATEKEY,
    privateKey );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

The same operation in Java or C# is:

```
/* Create the session */
int cryptSession = crypt.CreateSession( cryptUser,
    crypt.SESSION_CMP_SERVER );

/* Add the CA certificate store and CA server key and activate the
session */
crypt.SetAttribute( cryptSession, crypt.SESSINFO_KEYSET,
    cryptCertStore );
crypt.SetAttribute( cryptSession, crypt.SESSINFO_PRIVATEKEY,
    privateKey );
crypt.SetAttribute( cryptSession, crypt.SESSINFO_ACTIVE, 1 );
```

The Visual Basic equivalent is:

```
' Create the session
cryptCreateSession cryptSession, cryptUser, CRYPT_SESSION_CMP_SERVER

' Add the CA certificate store and CA server key and activate the
' session
cryptSetAttribute cryptSession, CRYPT_SESSINFO_KEYSET, cryptCertStore
cryptSetAttribute cryptSession, CRYPT_SESSINFO_PRIVATEKEY, privateKey
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1
```

Once you activate the session, cryptlib will block until an incoming client connection arrives, at which point it will negotiate the certificate issue or revocation process with the client. All checking and certificate processing operations are taken care of by cryptlib. There is no need for you to perform any further processing operations when running a CA in this way, although you may want to occasionally perform some of the maintenance operations described in “Managing a CA Directly” on page 264.

If you plan to use the PKIBoot certificate bootstrap mechanism to communicate trusted certificates to the user, you need to mark the certificates that you want cryptlib to supply to the user as trusted certificates as described in “Certificate Trust Management” on page 317. At a minimum, you should mark your CA’s certificates as trusted to ensure that the user will get the CA certificates alongside their own certificates when they have a certificate issued for them. In addition you can supply additional certificates (for example ones for certificate status responders or timestamp servers) to the user by marking them as trusted by the CA.

The cryptlib CMP and SCEP implementations run on top of a certificate store that implements consistent transactions (as far as the underlying software and hardware allows it), so that any incomplete CA transaction which is aborted by a software or hardware failure or network error will be either cleanly rolled back if it hasn’t been confirmed yet (for example a certificate issue request for which no acknowledgement was received from the user) or completed if it was confirmed (for example a revocation request that has been validated by cryptlib). This means that if (for example) the server on which the CA is running crashes halfway through a revocation operation, the revocation will be cleanly completed after the server is restarted. This behaviour may differ from the behaviour exhibited by other CAs, which (depending on CA policy) may simply abort all incomplete transactions, or may try and complete some transactions.

In addition to ensuring transactional integrity, cryptlib also enforces certificate status integrity constraints, which means that if it receives and successfully processes an update request for a certificate, it will revoke the certificate that was being updated to prevent two otherwise identical certificates from existing at the same time. As with the other transaction types, the replacement operation is atomic so that either the new certificate will cleanly replace the old one, or no overall change will take place.

Making Certificates Available Online

Once you’ve issued a certificate, you can make it available online using a standard HTTP keyset. This allows users to fetch certificates over the Internet by performing a standard keyset access. Although the interface is to a keyset, it’s handled as a cryptlib session of type `CRYPT_SESSION_CERTSTORE_SERVER` because it works with a variety of session interfaces and attributes that aren’t normally used with keysets.

Since a cert store session doesn't perform any crypto operations like the other session types, all that you need to add before you activate the session is the cert store keyset:

```
CRYPT_SESSION cryptSession;

/* Create the session */
cryptCreateSession( &cryptSession, cryptUser,
    CRYPT_SESSION_CERTSTORE_SERVER );

/* Add the CA certificate store and activate the session */
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_KEYSET,
    cryptCertStore );
cryptSetAttribute( cryptSession, CRYPT_SESSINFO_ACTIVE, 1 );
```

The Visual Basic equivalent is:

```
' Create the session
cryptCreateSession cryptSession, cryptUser, _
    CRYPT_SESSION_CERTSTORE_SERVER

' Add the CA certificate store and activate the
' session
cryptSetAttribute cryptSession, CRYPT_SESSINFO_KEYSET, cryptCertStore
cryptSetAttribute cryptSession, CRYPT_SESSINFO_ACTIVE, 1
```

Since the client-side of this session is a standard HTTP keyset, you can use it directly in crypto operations like signed or encrypted enveloping:

```
CRYPT_ENVELOPE cryptEnvelope;
int bytesCopied;

cryptCreateEnvelope( &cryptEnvelope, cryptUser, CRYPT_FORMAT_SMIME );

/* Add the encryption keyset and recipient email address */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_KEYSET_ENCRYPT,
    cryptKeyset );
cryptSetAttributeString( cryptEnvelope, CRYPT_ENVINFO_RECIPIENT,
    "person@company.com", 18 );

/* Add the data size information and data, wrap up the processing, and
   pop out the processed data */
cryptSetAttribute( cryptEnvelope, CRYPT_ENVINFO_DATASIZE,
    messageLength );
cryptPushData( cryptEnvelope, message, messageLength, &bytesCopied );
cryptFlushData( cryptEnvelope );
cryptPopData( cryptEnvelope, envelopedData, envelopedDataBufferSize,
    &bytesCopied );

cryptDestroyEnvelope( cryptEnvelope );
```

Although the interface is identical to the standard enveloping interface with a local keyset, in this case cryptlib is fetching the certificate that's required for encryption from the remote CA. Having the keyset available online and managed directly by the CA avoids requiring each user to manage their own individual store of certificates, and allows a single consistent certificate collection to be maintained at a central location.

For both security and performance reasons, you should always open the keyset in read-only mode and access it as a general certificate keyset (CRYPT_KEYSET_DATABASE, CRYPT_KEYSET_ODBC, or CRYPT_KEYSET_PLUGIN) rather than a CA certificate store (CRYPT_KEYSET_DATABASE_STORE, CRYPT_KEYSET_ODBC_STORE, or CRYPT_KEYSET_PLUGIN_STORE). cryptlib will check to make sure that it's a read-only standard keyset when you add it to the session, and return a CRYPT_ERROR_PARAM3 error if it's of the incorrect type.

For additional security, you can apply standard database security measures to protect the certificate database against (potentially malicious) access. Some ways of doing this include using the database's REVOKE/GRANT capability to allow only SELECT access (read-only, no write or update capability), and accessing the database as a low-privilege user with only read access. cryptlib will automatically use the lowest level of access available to perform the task, in this case minimal read-only access combined with basic SELECT point queries (no views, joins, or other

complexity). Finally, cryptlib both filters its input data and uses parameterised queries/bound query data to prevent hostile users from inserting malicious escape sequences into the query.

The CRYPT_SESSION_CERTSTORE_SERVER server type employs cryptlib as little more than a web interface to a certificate store. Since most databases are web-enabled, a simpler option is to use the database itself to provide certificate access to users — it's just a straight HTTP query of the database. This means that you can create standalone HTTP certificate store servers using nothing more than the database engine that you use to store the certificates.

Managing a CA Directly

In addition to the mostly-automated process of running a CA via CMP or SCEP, cryptlib also lets you manage a CA directly using various certificate management operations. This process isn't as convenient as using CMP or SCEP since a lot of the automation provided by cryptlib's automated CA handling is lost by working at this lower level.

A CA issues certificates and certificate revocations in response to requests from users, so that when an incoming request arrives the first thing you need to do is store it in the certificate store so that cryptlib can work with it. After that you can use the CA management functions to convert the request into a certificate or revocation and optionally return the result of the operation to the user.

Recording Incoming Requests

To store an incoming request you use **cryptCAAddItem**, which takes the request and adds it to the store, updating the audit log and performing any other necessary management operations. Once it's stored, cryptlib generates a log entry recording the arrival of the request and can use it to recover the request or any subsequent data such as certificates created from it even in the event of a system crash or failure, so that no information will be lost once it has entered the store:

```
CRYPT_CERTIFICATE cryptCertRequest;

/* Obtain the cert request from the user */
cryptCertRequest = ...;

/* Verify that the request is in order */
/* ... */

/* Add the request to the cert store */
cryptCAAddItem( cryptCertStore, cryptCertRequest );
```

Once this process has been completed the request has been entered into the store and will be subject to the CA management operations provided by cryptlib. This step must be completed before the certificate management process can be applied to the request, even if it'll immediately be used to generate a certificate or revocation, since it's needed to ensure that the operation of the CA can be recovered and continued in the event of a software or system failure.

Retrieving Stored Requests

Once a request has been recorded in the store, some time may elapse before it can be processed, during which time the certificate object that contains the request may be destroyed. When the certificate is ready for issue, you can recreate the request by retrieving it from the store using **cryptCAGetItem** in the same way that you can use **cryptGetPublicKey** to obtain a certificate from a standard certificate store:

```
CRYPT_CERTIFICATE cryptCertRequest;

/* Obtain the cert request from the user */
cryptCertRequest = ...;

/* Verify that the request is in order */
/* ... */
```

```

/* Add the request to the cert store and destroy it */
cryptCAAddItem( cryptCertStore, cryptCertRequest );
cryptDestroyCert( cryptCertRequest );

/* Perform other operations */
/* ... */

/* Recreate the request so that it can be processed */
cryptCAGetItem( cryptCertStore, &cryptCertRequest,
    CRYPT_CERTTYPE_REQUEST_CERT, CRYPT_CERTINFO_CRYPT_KEYID_NAME,
    name );

```

Once the request has been recreated, you can subject it to the CA management process in the usual manner.

CA Management Operations

cryptlib provides a wide variety of CA management operations that include issuing and revoking certificates and creating CRLs, as well as general management operations such as clearing up expired certificates and CRL entries. All of these operations are performed by cryptlib using **cryptCACertManagement** with no further input necessary from the user. The general concept of the certificate management function is:

```

CRYPT_CERTIFICATE cryptCertificate;

cryptCACertManagement( &cryptCertificate, action, cryptCertStore,
    cryptCAKey, cryptCertRequest );

```

with some of the parameters being optional depending on the type of action being performed. The certificate management actions that can be performed are:

Cert Management Action	Description
CRYPT_CERTACTION_- EXPIRE_CERT	Remove all expired certificates from the active certificate collection and remove all expired CRL entries from the active CRL entry collection in the certificate store.
CRYPT_CERTACTION_- CLEANUP	Perform certificate store cleanup/recovery actions after a restart (for example a system crash), processing or deleting any leftover incomplete actions as appropriate.
CRYPT_CERTACTION_- ISSUE_CERT	Issue a certificate by signing a certificate request with the given CA key, updating the certificate store to contain the newly-issued certificate.
CRYPT_CERTACTION_- ISSUE_CRL	Issue a CRL for the CA indicated by the given CA key.
CRYPT_CERTACTION_- REVOKE_CERT	Revoke the certificate indicated in the revocation request. Since submitting the corresponding revocation request requires interaction with the CMP protocol this action can't be performed directly but is initiated in conjunction with CMP.

The first parameter for the function can optionally return the newly-issued certificate or CRL, if you don't want to do anything with this at the current time you can set it to null and read it later with **cryptGetPublicKey**. In all cases cryptlib will carry out the operations in a safe, all-or-nothing manner that leaves the certificate store in a consistent state after the operation has completed. This guarantees the reliable operation of the CA even in the presence of hardware or software failures in the underlying components.

The details of each type of CA management operation are given in the following sections.

Issuing and revoking a Certificate

The process of issuing a certificate converts a previously stored certificate request into a certificate via the certificate store. To issue a certificate, you need to provide a certificate store, a CA key to use to sign the certificate, and a copy of the (previously stored) certificate request:

```
CRYPT_CERTIFICATE cryptCertificate;

cryptCACertManagement( &cryptCertificate, CRYPT_CERTACTION_ISSUE_CERT,
    cryptCertStore, cryptCAKey, cryptCertRequest );
```

Once the operation has completed, the new certificate will be available as the `cryptCertificate` value.

Revoking a certificate works in a similar manner, except that it takes a revocation request rather than a certificate request. Since this operation updates the certificate store without creating any kind of certificate object, the first parameter is set to null:

```
cryptCACertManagement( NULL, CRYPT_CERTACTION_REVOKE_CERT,
    cryptCertStore, cryptCAKey, cryptRevocationRequest );
```

This operation requires the use of a revocation request that can only be processed as part of the CMP protocol, so it's not possible to directly submit a revocation request to the store.

Issuing a CRL

The process of issuing a CRL takes the revocation information held in the certificate store and turns it into a finished CRL. To issue a CRL, you need to provide a certificate store and a CA key (specifically, one capable of signing CRLs) to use to sign the CRL. Since there's no request involved, the request parameter is set to `CRYPT_UNUSED`. If you try to use a CA key that can't sign CRLs, cryptlib will return `CRYPT_ERROR_PARAM4` to indicate that the key is invalid for issuing CRLs:

```
CRYPT_CERTIFICATE cryptCRL;

cryptCACertManagement( &cryptCRL, CRYPT_CERTACTION_ISSUE_CRL,
    cryptCertStore, cryptCAKey, CRYPT_UNUSED );
```

The CA key must be the one that issued the certificates that are in the CRL (this is a requirement of the way certificates in CRLs are identified). If you try and use a key from a different CA, the resulting CRL will either be empty (since no revocation entries for the other CA will be present) or will contain only entries for the other CA (if both CAs are sharing the same certificate store, and entries from the other CA are present in it).

Expiring Certificates

Expiring certificates is a passive process that doesn't create or destroy any certificate objects, but merely updates the certificate store state information so that expired certificates are no longer considered active. You can run this as a background or low-priority operation at periodic intervals to keep the certificate store up to date:

```
cryptCACertManagement( &cryptCRL, CRYPT_CERTACTION_EXPIRE_CERT,
    cryptCertStore, CRYPT_UNUSED, CRYPT_UNUSED );
```

This will remove any expired certificates from the store and also removes any CRL entries for certificates that have expired anyway. Depending on your CA's policy on expiry you can run this frequently to ensure only current certificates and CRL entries are present or less frequently in case there's some reason to keep expired certificates around.

Recovering after a Restart

Sometimes the machine on which you're running your CA may go down due to problems like a hardware failure or a system crash. cryptlib carries out all operations in a manner that ensures the certificate store won't be left in an inconsistent state, but having the machine die in the middle of an update can leave some requests in an

incomplete state (for example if an incoming request is received and system power is lost before the corresponding certificate is issued, the unprocessed request will be left in the certificate store). In order to clean up any leftover requests you can tell cryptlib to clean up the state of the certificate store by removing or processing any leftover requests as appropriate:

```
cryptCACertManagement( &cryptCRL, CRYPT_CERTACTION_CLEANUP,  
    cryptCertStore, CRYPT_UNUSED, CRYPT_UNUSED );
```

If a pending request hasn't been approved yet, it will be rolled back; if a request has been approved but wasn't fully processed, it will be completed.

In general it's a good idea to perform this action when you start your CA (if you shut it down for any reason), and you should do it if there's a system failure or other problem that causes the CA to shut down without cleaning up. Note that you should never perform this operation while the CA is running, since it'll clean up any currently un-processed requests and operations, including ones that may currently be awaiting processing by the CA.

Encryption and Decryption

Although envelope, session, and keyset container objects provide an easy way to work with encrypted data, it's sometimes desirable to work at a lower level, either because it provides more control over encryption parameters or because it's more efficient than the use of the higher-level functions. The objects that you use for lower-level encryption functionality are encryption contexts. Internally, more complex objects such as envelope, session, and certificate objects also use encryption contexts, although these are hidden and not accessible from the outside.

Once you've generated a public/private key pair, you probably want to communicate the public key to others. To do this, you need to encode the key components in a standard form that other applications can understand. The standard form for public keys is a certificate, described in "Certificates and Certificate Management" on page 234. If all you want to do is communicate public key data and you don't care about the other certificate details, you can use a simplified certificate as described in "Simple Certificate Creation" on page 237. This encodes the key in a universal certificate format, but without the management overhead of having to deal with certificates.

Alongside the portable, universal certificate format, there exist a number of non-portable, often proprietary formats that various vendors have invented for encoding keys. If you want to use one of these non-portable, non-standard formats, you need to contact the vendor that created it to determine the format details and what's required to convert a key to and from that format.

Creating/Destroying Encryption Contexts

To create an encryption context, you must specify the user who is to own the object or CRYPT_UNUSED for the default, normal user, the encryption algorithm, and optionally the encryption mode you want to use for that context. The available encryption algorithms and modes are given in "Algorithms" on page 380. For example, to create and destroy an encryption context for DES you would use the following code:

```
CRYPT_CONTEXT cryptContext;

cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_DES );

/* Load key, perform en/decryption */

cryptDestroyContext( cryptContext );
```

The context will use the default encryption mode of CBC, which is the most secure and efficient encryption mode. If you want to use a different mode, you can set the context's CRYPT_CTXINFO_MODE attribute to specify the mode to use. For example to change the encryption mode used from CBC to CFB you would use:

```
cryptSetAttribute( cryptContext, CRYPT_CTXINFO_MODE, CRYPT_MODE_CFB );
```

In general you shouldn't need to change the encryption mode, the other cryptlib functions will automatically handle the mode choice for you. Public-key, hash, and MAC contexts work in the same way, except that they don't have different modes of use so the CRYPT_CTXINFO_MODE attribute isn't present for these types of contexts. The availability of certain algorithms and encryption modes in cryptlib does not mean that their use is recommended. Some are only present because they are needed for certain protocols or required by some standards.

Note that the CRYPT_CONTEXT is passed to **cryptCreateContext** by reference, as **cryptCreateContext** modifies it when it creates the encryption context. In almost all other cryptlib routines, CRYPT_CONTEXT is passed by value. The contexts that will be created are standard cryptlib contexts, to create a context which is handled via a crypto device such as a smart card or Fortezza card, you should use **cryptDeviceCreateContext**, which tells cryptlib to create a context in a crypto

device. The use of crypto devices is explained in “Encryption Devices and Modules” on page 350.

cryptDestroyContext has a generic equivalent function **cryptDestroyObject** that takes a **CRYPT_HANDLE** parameter instead of a **CRYPT_CONTEXT**. This is intended for use with objects that are referred to using generic handles, but can also be used to specifically destroy encryption contexts — cryptlib’s object management routines will automatically sort out what to do with the handle or object.

Generating a Key into an Encryption Context

Once you’ve created an encryption context, the next step is to generate a key into it. These keys will typically be either one-off session keys that are discarded after use, or long-term storage keys that are used to protect fixed data such as files or private keys. You can generate a key with **cryptGenerateKey**:

```
cryptGenerateKey( cryptContext );
```

which will generate a key of a size which is appropriate for the encryption algorithm. If you want to generate a key of a particular length, you can set the **CRYPT_CTXINFO_KEYSIZE** attribute before calling **cryptGenerateKey**. For example to generate a 256-bit (32-byte) key you would use:

```
cryptSetAttribute( cryptContext, CRYPT_CTXINFO_KEYSIZE, 256 / 8 );
cryptGenerateKey( cryptContext );
```

Keys generated by cryptlib are useful when used with **cryptExportKey**/**cryptImportKey**. Since **cryptExportKey** usually encrypts the generated key using public-key encryption, you shouldn’t make it too long or it’ll be too big to be encrypted. Unless there’s a specific reason for choosing the key length you should use the **cryptGenerateKey** function and let cryptlib choose the correct key length for you.

The only time when you may need to explicitly specify a key length is when you’re using very short (in the vicinity of 512 bits) public keys to export Blowfish, RC2, RC4, or RC5 keys. In this case the public key isn’t large enough to export the full-length keys for these algorithms, and **cryptExportKey** will return the error code **CRYPT_ERROR_OVERFLOW** to indicate that there’s too much data to export. The solution is to either specify a shorter key length using the **CRYPT_CTXINFO_KEYSIZE** attribute or, preferably, to use a longer public key. This is only a problem with very short public keys, when using the minimum recommended public key size of 1024 bits this situation will never occur.

Calling **cryptGenerateKey** only makes sense for conventional, public-key, or MAC contexts and will return the error code **CRYPT_ERROR_NOTAVAIL** for a hash context to indicate that this operation is not available for hash algorithms. The generation of public/private key pairs has special requirements and is covered in “Key Generation and Storage” on page 216.

To summarise the steps so far, you can set up an encryption context in its simplest form so that it’s ready to encrypt data with:

```
CRYPT_CONTEXT cryptContext;

cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );
cryptGenerateKey( cryptContext );

/* Encrypt data */

cryptDestroyContext( cryptContext );
```

Once a key is generated into a context, you can’t load or generate a new key over the top of it or change the encryption mode (for conventional encryption contexts). If you try to do this, cryptlib will return **CRYPT_ERROR_INITED** to indicate that a key is already loaded into the context.

Deriving a Key into an Encryption Context

Sometimes you will need to obtain a fixed-format encryption key for a context from a variable-length password or passphrase, or from any generic keying material. You can do this by deriving a key into a context rather than loading it directly. Deriving a key converts arbitrary-format keying information into the particular form required by the context, as well as providing extra protection against password-guessing attacks and other attacks that might take advantage of knowledge of the keying materials' format.

The key derivation process takes two sets of input data, the keying material itself (typically a password), and a salt value which is combined with the password to ensure that the key is different each time (so even if you reuse the same password multiple times, the key obtained from it will change each time). This ensures that even if one password-based key is compromised, all the others remain secure.

The salt attribute is identified by `CRYPT_CTXINFO_KEYING_SALT` and ranges in length from 64 bits (8 bytes) up to `CRYPT_MAX_HASHSIZE`. Using an 8-byte salt is a good choice. The keying information attribute is identified by `CRYPT_CTXINFO_KEYING_VALUE` and can be of any length. To derive a key into a context you would use:

```
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEYING_SALT,
    salt, saltLength );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEYING_VALUE,
    passPhrase, passPhraseLength );
```

which takes the supplied passphrase and salt and converts them into an encryption key in a format suitable for use with the encryption context. Use of the key derivation capability is strongly recommended over loading keys directly into an encryption context by setting the `CRYPT_CTXINFO_KEY` attribute since this often requires intimate knowledge of algorithm details such as how keys of different lengths are handled, how key bits are used, special considerations for key material, and so on.

Note that you have to set a salt value before you set the keying information attribute. If you don't supply a salt, cryptlib will return `CRYPT_ERROR_NOTINITED` when you try to supply the keying information to indicate that the salt hasn't been set yet. If you don't want to manage a unique salt value per key, you can set the salt to a fixed value (for example 64 bits of zeroes), although this is strongly discouraged since it means each use of the password will produce the same encryption key.

By default the key derivation process will repeatedly hash the input salt and keying information with the HMAC-SHA1 MAC function to generate the key, and will iterate the hashing process 500 times to make a passphrase-guessing attack more difficult⁵. If you want to change these values you can set the `CRYPT_CTXINFO_KEYING_ALGO` and `CRYPT_CTXINFO_KEYING_ITERATIONS` attributes for the context before setting the salt and keying information attributes. For example to change the number of iterations to 1000 for extra security before setting the salt and key you would use:

```
cryptSetAttribute( cryptContext, CRYPT_CTXINFO_KEYING_ITERATIONS,
    1000 );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEYING_SALT,
    salt, saltLength );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEYING_VALUE,
    passPhrase, passPhraseLength );
```

cryptlib will then use this value when deriving the key. You can also change the default hash algorithm and iteration count using the cryptlib configuration options `CRYPT_OPTION_KEYING_ALGO` and `CRYPT_OPTION_KEYING_ITERATIONS` as explained in "Working with Configuration Options" on page 359.

⁵ It actually does a lot more than just hashing the passphrase, including performing processing steps designed to defeat various sophisticated attacks on the key-hashing process.

To summarise the steps so far, you can set up an encryption context in its simplest form so that it's ready to encrypt data with:

```
CRYPT_CONTEXT cryptContext;

cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEYING_SALT,
    salt, saltLength );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEYING_VALUE,
    passPhrase, strlen( passPhrase ) );

/* Encrypt data */

cryptDestroyContext( cryptContext );
```

Since public-key encryption uses a different type of key than other context types, you can't derive a key into a public or private key context.

Once a key is derived into a context, you can't load or generate a new key over the top of it or change the encryption mode (for conventional encryption contexts). If you try to do this, cryptlib will return CRYPT_ERROR_INITED to indicate that a key is already loaded into the context.

Loading a Key into an Encryption Context

If necessary you can also manually load a raw key into an encryption context by setting the CRYPT_CTXINFO_KEY attribute. For example to load a raw 128-bit key "0123456789ABCDEF" into an IDEA conventional encryption context you would use:

```
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEY,
    "0123456789ABCDEF", 16 );
```

Unless you need to perform low-level key management yourself, you should avoid loading keys directly in this manner. The previous key load should really have been done by setting the CRYPT_CTXINFO_KEYING_SALT and CRYPT_CTXINFO_KEYING_VALUE attributes to derive the key into the context.

For public-key encryption a key will typically have a number of components so you can't set the key directly. More information on working with CRYPT_PKCINFO data structures is given in "Loading Public/Private Keys" on page 272.

Once a key is loaded into a context, you can't load or generate a new key over the top of it or change the encryption mode (for conventional encryption contexts). If you try to do this, cryptlib will return CRYPT_ERROR_INITED to indicate that a key is already loaded into the context.

If you need to reserve space for conventional and public/private keys, you can use the CRYPT_MAX_KEYSIZE and CRYPT_MAX_PKCSIZE defines to determine the amount of memory you need. No key used by cryptlib will ever need more storage than the settings given in these defines. Note that the CRYPT_MAX_PKCSIZE value specifies the maximum size of an individual key component. Since public/private keys are usually composed of a number of components the overall size is larger than this.

Working with Initialisation Vectors

For conventional-key encryption contexts you can also load an initialisation vector (IV) into the context if the encryption mode being used supports an IV, although when you're using a context to encrypt data you can leave this to cryptlib to perform automatically when you call **cryptEncrypt** for the first time. IVs are required for the CBC, CFB, and OFB encryption modes. To load an IV you set the CRYPT_CTXINFO_IV attribute:

```
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_IV, iv, ivSize );
```

To retrieve the IV that you have loaded or that has been generated for you by cryptlib you read the value of the attribute:

Trying to get or set the value of this attribute will return the error code `CRYPT_ERROR_NOTAVAIL` for a hash, MAC, or public key encryption context or conventional encryption context with an encryption mode that doesn't use an IV to indicate that these operations are not available for this type of context.

instead of being stored with 50 digits of precision of which 41 bytes contain zero padding, they would be stored with 9 digits of precision:

```
123456789
```

A multibyte integer therefore consists of two parameters, the data itself and the precision to which it is stored, specified in bits. When you load multibyte integer components into a `CRYPT_PKCINFO` structure you need to specify both of these parameters.

Before you can use the `CRYPT_PKCINFO` structure, you need to initialise it with `cryptInitComponents()`, which takes as parameters a pointer to the `CRYPT_PKCINFO` structure and the type of the key, either `CRYPT_KEYTYPE_PRIVATE` or `CRYPT_KEYTYPE_PUBLIC`:

```
CRYPT_PKCINFO_RSA rsaKey;

cryptInitComponents( &rsaKey, CRYPT_KEYTYPE_PRIVATE );
```

Now you can load the multibyte integer strings by using `cryptSetComponent()`, specifying a pointer to the value to be loaded, the multibyte integer data, and the integer length in bits:

```
cryptSetComponent( ( &rsaKey )->n, modulus, 1024 );
cryptSetComponent( ( &rsaKey )->e, pubExponent, 17 );
cryptSetComponent( ( &rsaKey )->d, privExponent, 1024 );
```

Since `cryptSetComponent()` takes as parameter a pointer to the value to be loaded, it's necessary to pass in the address as shown above when the `CRYPT_PKCINFO` structure is declared statically. If it's dynamically allocated as in the example below, this extra step isn't necessary.

Once all the parameters are set up, you can use the result as the `CRYPT_CTXINFO_KEY_COMPONENTS` as explained above. Once you've finished working with the `CRYPT_PKCINFO` information, use `cryptDestroyComponents` to destroy the information:

```
cryptDestroyComponents( &rsaKey );
```

The Diffie-Hellman, DSA, and Elgamal algorithms share the same key format and all use the `CRYPT_PKCINFO_DLP` structure to store their key components. DLP is short for Discrete Logarithm Problem, the common underlying mathematical operation for the three cryptosystems.

When loading key components, cryptlib performs a validity check on the data to detect invalid or suspicious key values. These can be used to compromise the security of the key, for example to leak the private key in signatures made with it. If cryptlib detects suspicious key components, it will return `CRYPT_ERROR_PARAM3` to indicate that the key components are invalid.

To summarise the steps so far, you would load a public key into a DSA context with:

```
CRYPT_CONTEXT cryptContext;
CRYPT_PKCINFO_DLP *dlpKey;

cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_DSA );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_LABEL, "DSA key",
7 );
dlpKey = malloc( sizeof( CRYPT_PKCINFO_DLP ) );
cryptInitComponents( dlpKey, CRYPT_KEYTYPE_PUBLIC );
cryptSetComponent( dlpKey->p, ... );
cryptSetComponent( dlpKey->q, ... );
cryptSetComponent( dlpKey->g, ... );
cryptSetComponent( dlpKey->y, ... );
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_KEY_COMPONENTS,
dlpKey, sizeof( CRYPT_PKCINFO_DLP ) );
cryptDestroyComponents( dlpKey );
```

The context is now ready to be used to verify a DSA signature on a piece of data. If you wanted to load a DSA private key (which consists of one extra component), you would add:

```
cryptSetComponent( dlpKey->x, ... );
```

after the y component is loaded. This context can then be used to sign a piece of data.

Querying Encryption Contexts

A context has a number of attributes whose values you can get to obtain information about it. These attributes contain details such as the algorithm type and name, the key size (if appropriate), the key label (if this has been set), and various other details. The information attributes are:

Value	Type	Description
CRYPT_CTXINFO_ALGO CRYPT_CTXINFO_MODE	N	Algorithm and mode
CRYPT_CTXINFO_BLOCKSIZE	N	Cipher block size in bytes
CRYPT_CTXINFO_IVSIZE	N	Cipher IV size in bytes
CRYPT_CTXINFO_KEYING_ -ALGO CRYPT_CTXINFO_KEYING_ -ITERATIONS CRYPT_CTXINFO_KEYING_ -SALT	N/S	The algorithm and number of iterations used to transform a user-supplied key or password into an algorithm-specific key for the context, and the salt value used in the transformation process
CRYPT_CTXINFO_KEYSIZE	N	Key size in bytes
CRYPT_CTXINFO_LABEL	S	Key label
CRYPT_CTXINFO_NAME_ALGO CRYPT_CTXINFO_NAME_MODE	S	Algorithm and mode name

For example to obtain the algorithm and mode used by an encryption context, you would use:

```
CRYPT_ALGO_TYPE cryptAlgo;
CRYPT_MODE_TYPE cryptMode;

cryptGetAttribute( cryptContext, CRYPT_CTXINFO_ALGO, &cryptAlgo );
cryptGetAttribute( cryptContext, CRYPT_CTXINFO_MODE, &cryptMode );
```

Although these attributes are listed as context attributes, they also apply to anything else that can act as a context action object, for example you can obtain algorithm, mode, and key size values from a certificate since it can be used to encrypt or sign just like a context:

```
CRYPT_ALGO_TYPE cryptAlgo;
CRYPT_MODE_TYPE cryptMode;

cryptGetAttribute( cryptCertificate, CRYPT_CTXINFO_ALGO, &cryptAlgo );
cryptGetAttribute( cryptCertificate, CRYPT_CTXINFO_MODE, &cryptMode );
```

If any of the user-supplied attributes haven't been set and you try to read their value, cryptlib will return CRYPT_ERROR_NOTINITED.

Using Encryption Contexts to Process Data

To encrypt or decrypt a block of data using an encryption context action object you use:

```
cryptEncrypt( cryptContext, buffer, length );
```

and:

```
cryptDecrypt( cryptContext, buffer, length );
```

The data is encrypted in place, so that plaintext data is replaced by encrypted data and vice versa. If the encryption context doesn't support the operation you are trying to perform (for example calling **cryptEncrypt** with a DSA public key), the function will return CRYPT_ERROR_NOTAVAIL to indicate that this functionality is not

available. If the key loaded into an encryption context doesn't allow the operation you are trying to perform (for example calling **cryptDecrypt** with an encrypt-only key), the function will return `CRYPT_ERROR_PERMISSION` to indicate that the context doesn't have the required key permissions to perform the requested operation.

Conventional Encryption

If you're using a block cipher in ECB or CBC mode, the encrypted data length must be a multiple of the block size. If the encrypted data length is not a multiple of the block size, the function will return `CRYPT_ERROR_PARAM3` to indicate that the length is invalid. To encrypt a byte at a time you should use a stream encryption mode such as CFB or OFB, or better yet use an envelope which avoids the need to handle algorithm-specific details.

If an IV is required for the decryption and you haven't loaded one into the context by setting the `CRYPT_CTXINFO_IV` attribute, **cryptDecrypt** will return `CRYPT_ERROR_NOTINITED` to indicate that you need to load an IV before you can decrypt the data. If the first 8 bytes of decrypted data are corrupted then you haven't set up the IV properly for the decryption. More information on setting up IVs is given in "Working with Initialisation Vectors" on page 271. The general concept behind using IVs (in this case with automatic IV generation) is:

```
unsigned char iv[ CRYPT_MAX_IVSIZE ];
int ivSize;

/* Encrypt data */
cryptEncrypt( cryptContext, data, dataLength );
cryptGetAttributeString( cryptContext, CRYPT_CTXINFO_IV, iv, &ivSize
    );

/* Communicate the encrypted data and IV to the recipient */
/* ... */

/* Decrypt data */
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_IV, iv, ivSize );
cryptDecrypt( cryptContext, data, dataLength )
```

Once an encryption context is set up, it can only be used for processing a single data stream in an operation such as encrypting data, decrypting data, or hashing a message. A context can't be reused to encrypt a second message after the first one has been encrypted, or to decrypt data after having encrypted data. This is because the internal state of the context is determined by the operation being performed with it, and performing two different operations with the same context causes the state from the first operation to affect the second operation. For example if you use an encryption context to encrypt two different files, cryptlib will see a single continuous data stream (since it doesn't know or care about the structure of the data being encrypted). As a result the second file is treated as a continuation of the first one, and can't be decrypted unless the context is used to decrypt the first file before decrypting the second one. Because of this you should always create a new encryption context for each discrete data stream you will be processing, and never reuse contexts to perform different operations. The one exception to this rule is when you're using cryptlib envelopes (described in "Data Enveloping" on page 143), where you can push a single encryption context into as many envelopes as you like. This is because an envelope takes its own copy of the encryption context, leaving the original untouched.

In practice this isn't strictly accurate, you can encrypt multiple independent data streams with a single context by loading a new IV for each new stream using the `CRYPT_CTXINFO_IV` attribute:

```
/* Set an IV and encrypt data */
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_IV, iv1,
    iv1Length );
cryptEncrypt( cryptContext, data1, data1Length );
```

```

/* Set a new IV and encrypt more data */
cryptSetAttributeString( cryptContext, CRYPT_CTXINFO_IV, iv2,
    iv2Length );
cryptEncrypt( cryptContext, data2, data2Length );

```

If you don't understand how this would work then it's probably best to use a new context for each data stream.

Public-key Encryption

The public-key algorithms encrypt a single block of data equal in length to the size of the public key being used. For example if you are using a 1024-bit public key then the length of the data to be encrypted should be 128 bytes. If the encrypted data length isn't the same as the key size, the function will return `CRYPT_ERROR_PARAM3` to indicate that the length is invalid. Preparation of the block of data to be encrypted requires special care and is covered in appropriate security standards. If cryptlib detects that it's being passed incorrectly-formatted input data, it will return `CRYPT_ERROR_BADDATA` to indicate that the data being passed to the en/decryption function is invalid. In general you should use high-level functions such as `cryptExportKey/cryptImportKey` and `cryptCreateSignature/cryptCheckSignature` rather than `cryptEncrypt` and `cryptDecrypt` when working with public-key algorithms.

If you're using a public or private key context which is tied to a certificate or crypto device, the direct use of `cryptEncrypt` and `cryptDecrypt` could be used to bypass security constraints placed on the context (for example by changing the data formatting used with an encryption-only RSA private key context it's possible to misuse it to generate signatures even if the context is specifically intended for non-signature use). Because of this, if a context is tied to a certificate or a crypto device, it can't be used directly with these low-level functions but only with a higher-level function like `cryptCreateSignature` or with the enveloping code, which guarantee that a context can't be misused for a disallowed purpose. If you try to use a constrained context of this type directly, the function will return `CRYPT_ERROR_PERMISSION` to indicate that the context doesn't have the required permissions to perform the requested operation.

Hashing

Hash and MAC algorithms don't actually encrypt the data being hashed and can be called via `cryptEncrypt` or `cryptDecrypt`. They require a final call with the length set to 0 as a courtesy call to indicate to the hash or MAC function that this is the last data block and that the function should take whatever special action is necessary for this case:

```

cryptEncrypt( hashContext, buffer, length );
cryptEncrypt( hashContext, buffer, 0 );

```

If you call `cryptEncrypt` or `cryptDecrypt` after making the final call with the length set to 0, the function will return `CRYPT_ERROR_COMPLETE` to indicate that the hashing has completed and cannot be continued. Once the hashing is complete, the hash value is made available as the `CRYPT_CTXINFO_HASHVALUE` attribute that you can read in the usual manner:

```

unsigned char hash[ CRYPT_MAX_HASHSIZE ];
int hashLength;

cryptGetAttributeString( cryptContext, CRYPT_CTXINFO_HASHVALUE, hash,
    &hashLength );

```

You can reset a hash or MAC context by deleting the `CRYPT_CERTINFO_HASHVALUE` attribute, which allows you to reuse the context to generate another hash or MAC value. Reusing a context in this manner avoids the overhead of creating a context, and in the case of a MAC context the somewhat complex key processing which is required when the context is first used:

```
unsigned char hash1[ CRYPT_MAX_HASHSIZE ];
unsigned char hash2[ CRYPT_MAX_HASHSIZE ];
int hash1Length, hash2Length;

/* Hash or MAC data */
/* ... */
cryptGetAttributeString( cryptContext, CRYPT_CTXINFO_HASHVALUE, hash1,
    &hash1Length );

/* Delete the attribute to allow the context to be reused */
cryptDeleteAttribute( cryptContext, CRYPT_CTXINFO_HASHVALUE );

/* Hash or MAC more data */
/* ... */
cryptGetAttributeString( cryptContext, CRYPT_CTXINFO_HASHVALUE, hash2,
    &hash2Length );
```

Exchanging Keys

Although you can encrypt/decrypt or MAC data with an encryption context, the key you're using is locked inside the context and (if you used **cryptGenerateKey** to create it) won't be known to you or the person you're trying to communicate with. To share the key with another party, you need to export it from the context in a secure manner and the other party needs to import it into an encryption context of their own. Because the key is a very sensitive and valuable resource, you can't just read it out of the context, but need to take special steps to protect the key once it leaves the context. This is taken care of by the key export/import functions.

These functions deal only with the export and import of keys for conventional encryption or MAC contexts. Public/private keys have specialised requirements and can't be exported directly in the same manner as conventional encryption or MAC keys. Public keys, which are composite values consisting of multiple components, must be converted into certificates in order to be shared with another party. Certificates are covered in "Certificates and Certificate Management" on page 234. Private keys can't be exported as such, but can only be stored in keysets or crypto devices. Keysets are covered in "Key Generation and Storage" on page 216, and crypto devices are covered in "Encryption Devices and Modules" on page 350.

Exporting a Key

To exchange a conventional encryption or MAC key with another party, you use the **cryptExportKey** and **cryptImportKey** functions in combination with a conventional or public-key encryption context or public key certificate. Let's say you've created a key in an encryption context `cryptContext` and want to send it to someone whose public key is in the encryption context `pubKeyContext` (you can also pass in a private key if you want, **cryptExportKey** will only use the public key components). To do this you'd use:

```
CRYPT_CONTEXT pubKeyContext, cryptContext;
void *encryptedKey;
int encryptedKeyLength;

/* Generate a key */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );
cryptGenerateKey( cryptContext );

/* Allocate memory for the encrypted key */
encryptedKey = malloc( encryptedKeyMaxLength );

/* Export the key using a public-key encrypted blob */
cryptExportKey( encryptedKey, encryptedKeyMaxLength,
               &encryptedKeyLength, pubKeyContext, cryptContext );
```

The resulting public-key encrypted blob is placed in the memory buffer pointed to by `encryptedKey` of maximum size `encryptedKeyMaxLength`, and the actual length is stored in `encryptedKeyLength`. This leads to a small problem: How do you know how big to make the buffer? The answer is to use **cryptExportKey** to tell you. If you pass in a null pointer for `encryptedKey`, the function will set `encryptedKeyLength` to the size of the resulting blob, but not do anything else. You can then use code like:

```
cryptExportKey( NULL, 0, &encryptedKeyMaxLength, pubKeyContext,
               cryptContext );
encryptedKey = malloc( encryptedKeyMaxLength );
cryptExportKey( encryptedKey, encryptedKeyMaxLength,
               &encryptedKeyLength, pubKeyContext, cryptContext );
```

to create the exported key blob. Note that due to encoding issues for some algorithms the final exported blob may be one or two bytes smaller than the size which is initially reported, since the true size can't be determined until the key is actually exported. Alternatively, you can just reserve a reasonably sized block of memory and use that to hold the encrypted key. "Reasonably sized" means a few Kb, a 4K block

is plenty (an encrypted key blob for a 1024-bit public key is only about 200 bytes long).

You can also use a public key certificate to export a key. If, instead of a public key context, you had a key certificate contained in the certificate object `cryptCertificate`, the code for the previous example would become:

```
CRYPT_CERTIFICATE cryptCertificate;
CRYPT_CONTEXT cryptContext;
void *encryptedKey;
int encryptedKeyLength;

/* Generate a key */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );
cryptGenerateKey( cryptContext );

/* Allocate memory for the encrypted key */
encryptedKey = malloc( encryptedKeyMaxLength );

/* Export the key using a public-key encrypted blob */
cryptExportKey( encryptedKey, encryptedKeyMaxLength,
               &encryptedKeyLength, cryptCertificate, cryptContext );
```

The use of key certificates is explained in “Certificates and Certificate Management” on page 234.

If the encryption context contains too much data to encode using the given public key (for example trying to export an encryption context with a 600-bit key using a 512-bit public key) the function will return `CRYPT_ERROR_OVERFLOW`. As a rule of thumb a 1024-bit public key should be large enough to export the default key sizes for any encryption context.

If the public key is stored in an encryption context with a certificate associated with it or in a key certificate, there may be constraints on the key usage that are imposed by the certificate. If the key can’t be used for the export operation, the function will return `CRYPT_ERROR_PERMISSION` to indicate that the key isn’t valid for this operation, you can find out more about the exact nature of the problem by reading the error-related attributes as explained in “Extended Error Reporting” on page 369.

Exporting using Conventional Encryption

You don’t need to use public-key encryption to export a key blob, it’s also possible to use a conventional encryption context to export the key from another conventional encryption context. For example if you were using the key derived from the passphrase “This is a secret key” (which was also known to the other party) in an encryption context `keyContext` you would use:

```
CRYPT_CONTEXT sharedContext, keyContext;
void *encryptedKey;
int encryptedKeyLength;

/* Derive the export key into an encryption context */
cryptCreateContext( &keyContext, cryptUser, CRYPT_ALGO_3DES );
cryptSetAttributeString( keyContext, CRYPT_CTXINFO_KEYING_SALT, salt,
                        saltLength );
cryptSetAttributeString( keyContext, CRYPT_CTXINFO_KEYING_VALUE, "This
                        is a secret key", 20 );

/* Generate a key */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );
cryptGenerateKey( cryptContext );

/* Allocate memory for the encrypted key */
encryptedKey = malloc( encryptedKeyMaxLength );

/* Export the key using a conventionally encrypted blob */
cryptExportKey( encryptedKey, encryptedKeyMaxLength,
               &encryptedKeyLength, keyContext, cryptContext );
```

You don’t need to use a derived key to export the session key, you could have loaded the context in some other manner (for example from a crypt device such as a smart

card), but the sample code shown above, and further on for the key import phase, assumes that you'll be deriving the export/import key from a password.

This kind of key export isn't as convenient as using public keys since it requires that both sides know the encryption key in `keyContext` (or at least know how to derive it from some other keying material). One case where it's useful is when you want to encrypt data such as a disk file that will be decrypted later by the same person who originally encrypted it. By prepending the key blob to the start of the encrypted file, you can ensure that each file is encrypted with a different session key (this is exactly what the cryptlib enveloping functions do). It also means you can change the password on the file by changing the exported key blob, without needing to decrypt and re-encrypt the entire file.

Importing a Key

Now that you've exported the conventional encryption or MAC key, the other party needs to import it. This is done using the **`cryptImportKey`** function and the private key corresponding to the public key used by the sender:

```
CRYPT_CONTEXT privKeyContext, cryptContext;

/* Create a context for the imported key */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );

/* Import the key from the public-key encrypted blob */
cryptImportKey( encryptedKey, encryptedKeyLength, privKeyContext,
               cryptContext );
```

Note the use of `CRYPT_ALGO_3DES` when creating the context for the imported key, this assumes that both sides have agreed in advance on the use of a common encryption algorithm to use (in this case triple DES). If the algorithm information isn't available, you'll have to negotiate the details in some other way. This is normally done for you by cryptlib's enveloping code but isn't available when operating at this lower level.

To summarise, sharing an encryption context between two parties using public-key encryption involves the following steps:

```
/* Party A creates the required encryption context and generates a key
   into it */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );
cryptGenerateKey( cryptContext );

/* Party A exports the key using party B's public key */
cryptExportKey( encryptedKey, encryptedKeyMaxLength,
               &encryptedKeyLength, pubKeyContext, cryptContext );

/* Party B creates the encryption context to import the key into */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );

/* Party B imports the key using their private key */
cryptImportKey( encryptedKey, encryptedKeyLength, privKeyContext,
               cryptContext );
```

If the public key is stored in an encryption context with a certificate associated with it or in a key certificate, there may be constraints on the key usage that are imposed by the certificate. If the key can't be used for the import operation, the function will return `CRYPT_ERROR_PERMISSION` to indicate that the key isn't valid for this operation. You can find out more about the exact nature of the problem by reading the error-related attributes as explained in "Extended Error Reporting" on page 369.

Importing using Conventional Encryption

If the key has been exported using conventional encryption, you can again use conventional encryption to import it. Using the same key derived from the passphrase "This is a secret key" that was used in the key export example, you would use:

```

CRYPT_CONTEXT keyContext, cryptContext;

/* Derive the import key into an encryption context */
cryptCreateContext( &keyContext, cryptUser, CRYPT_ALGO_3DES );
cryptSetAttributeString( keyContext, CRYPT_CTXINFO_KEYING_SALT, salt,
    saltLength );
cryptSetAttributeString( keyContext, CRYPT_CTXINFO_KEYING_VALUE, "This
    is a secret key", 20 );

/* Create a context for the imported key */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_3DES );

/* Import the key from the conventionally encrypted blob */
cryptImportKey( encryptedKey, encryptedKeyLength, keyContext,
    cryptContext );

```

Since the salt is a random value that changes for each key you derive, you won't know it in advance so you'll have to obtain it by querying the exported key object as explained below. Once you've queried the object, you can use the salt which is returned with the query information to derive the import key as shown in the above code.

Querying an Exported Key Object

So far it's been assumed that you know what's required to import the exported key blob you're given (that is, you know which type of processing to use to create the encryption context needed to import a conventionally encrypted blob). However sometimes you may not know this in advance, which is where the **cryptQueryObject** function comes in. **cryptQueryObject** is used to obtain information on the exported key blob that might be required to import it. You can also use **cryptQueryObject** to obtain information on signature blobs, as explained in "Querying a Signature Object" on page 285.

The function takes as parameters the object you want to query, and a pointer to a `CRYPT_OBJECT_INFO` structure which is described in "CRYPT_OBJECT_INFO Structure" on page 402. The object type will be either a `CRYPT_OBJECT_ENCRYPTED_KEY` for a conventionally encrypted exported key, a `CRYPT_OBJECT_PKCENCRYPTED_KEY` for a public-key encrypted exported key, or a `CRYPT_OBJECT_KEYAGREEMENT` for a key-agreement key. If you were given an arbitrary object of an unknown type you'd use the following code to handle it:

```

CRYPT_OBJECT_INFO cryptObjectInfo;

cryptQueryObject( object, objectLength, &cryptObjectInfo );
if( cryptObjectInfo.objectType == CRYPT_OBJECT_ENCRYPTED_KEY )
    /* Import the key using conventional encryption */;
else
    if( cryptObjectInfo.objectType == CRYPT_OBJECT_PKCENCRYPTED_KEY ||
        cryptObjectInfo.objectType == CRYPT_OBJECT_KEYAGREEMENT )
        /* Import the key using public-key encryption */;
    else
        /* Error */;

```

Any `CRYPT_OBJECT_INFO` fields that aren't relevant for this type of object are set to null or zero as appropriate.

Once you've found out what type of object you have, you can use the other information returned by **cryptQueryObject** to process the object. For both conventional and public-key encrypted exported objects you can find out which encryption algorithm and mode were used to export the key using the `cryptAlgo` and `cryptMode` fields. For conventionally encrypted exported objects you can obtain the salt needed to derive the import key from the `salt` and `saltSize` fields.

Extended Key Export/Import

The **cryptExportKey** and **cryptImportKey** functions described above export and import conventional encryption or MAC keys in the cryptlib default format (which, for the technically inclined, is the Cryptographic Message Syntax format with key identifiers used to denote public keys). The default cryptlib format has been chosen

to be independent of the underlying key format, so that it works equally well with any key type including X.509 certificates, PGP/OpenPGP keys, and any other key storage format.

Alongside the default format, cryptlib supports the export and import of keys in other formats using **cryptExportKeyEx**. **cryptExportKeyEx** works like **cryptExportKey** but takes an extra parameter that specifies the format to use for the exported keys. The formats are:

Format	Description
CRYPT_FORMAT_CMS CRYPT_FORMAT_SMIME	These are variations of the Cryptographic Message Syntax and are also known as S/MIME version 2 or 3 and PKCS #7. This format only allows public-key-based export, and the public key must be stored as an X.509 certificate.
CRYPT_FORMAT_CRYPTLIB	This is the default cryptlib format and can be used with any type of key. When used for public-key based key export, this format is also known as a newer variation of S/MIME version 3.
CRYPT_FORMAT_PGP	This is the OpenPGP format and can be used with any type of key.

cryptImportKeyEx takes one extra parameter, a pointer to the imported key, which is required for OpenPGP key import. For all other formats this value is set to NULL, for OpenPGP the imported key parameter is set to CRYPT_UNUSED and the key is returned in the extra parameter:

```
/* Import a non-PGP format key */
cryptImportKeyEx( encryptedKey, encryptedKeyLength, importContext,
                 cryptContext, NULL );

/* Import a PGP-format key */
cryptImportKeyEx( encryptedKey, encryptedKeyLength, importContext,
                 CRYPT_UNUSED, &cryptContext );
```

This is required because PGP's handling of keys differs somewhat from that used with other formats.

Key Agreement

The Diffie-Hellman key agreement capability is currently disabled since, unlike RSA and conventional key exchange, there's no widely-accepted standard format for it (SSL/TLS and SSHv2 are handled internally by cryptlib and CMS is never used by anything). If a widely-accepted standard emerges, cryptlib will use that format. Previous versions of cryptlib used a combination of PKCS #3, PKCS #5, and PKCS #7 formats and mechanisms to handle DH key agreement.

cryptlib supports a third kind of key export/import that doesn't actually export or import a key but merely provides a means of agreeing on a shared secret key with another party. You don't have to explicitly load or generate a session key for this one since the act of performing the key exchange will create a random, secret shared key. To use this form of key exchange, both parties call **cryptExportKey** to generate the blob to send to the other party, and then both in turn call **cryptImportKey** to import the blob sent by the other party.

The use of **cryptExportKey/cryptImportKey** for key agreement rather than key exchange is indicated by the use of a key agreement algorithm for the context that would normally be used to export the key. The key agreement algorithm used by cryptlib is the Diffie-Hellman (DH) key exchange algorithm, CRYPT_ALGO_DH. In the following code the resulting Diffie-Hellman context is referred to as dhContext.

Since there's a two-way exchange of messages, both parties must create an identical "template" encryption context so **cryptExportKey** knows what kind of key to export. Lets assume that both sides know they'll be using Blowfish in CFB mode. The first step of the key exchange is therefore:

```
/* Create the key template */
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_BLOWFISH );
cryptSetAttribute( cryptContext, CRYPT_CTXINFO_MODE, CRYPT_MODE_CFB );

/* Export the key using the template */
cryptExportKey( encryptedKey, encryptedKeyMaxLength,
               &encryptedKeyLength, dhContext, cryptContext );
cryptDestroyContext( cryptContext );
```

Note that there's no need to load a key into the template, since this is generated automatically as part of the export/import process. In addition the template is destroyed once the key has been exported, since there's no further use for it — it merely acts as a template to tell **cryptExportKey** what to do.

Both parties now exchange `encryptedKey` blobs, and then use:

```
cryptImportKey( encryptedKey, encryptedKeyLength, dhContext,
               cryptContext );
```

to create the `cryptContext` containing the shared key.

The agreement process requires that both sides export their own `encryptedKey` blobs before they import the other sides `encryptedKey` blob. A side-effect of this is that it allows additional checking on the key agreement process to be performed to guard against things like triple DES turning into 40-bit RC4 during transmission. If you try to import another party's `encryptedKey` blob without having first exported your own `encryptedKey` blob, **cryptImportKey** will return `CRYPT_ERROR_NOTINITED`.

Signing Data

Most public-key encryption algorithms can be used to generate digital signatures on data. A digital signature is created by signing the contents of a hash context with a private key to create a signature blob, and verified by checking the signature blob with the corresponding public key.

To do this, you use the **cryptCreateSignature** and **cryptCheckSignature** functions in combination with a public-key encryption context. Let's say you've hashed some data with an SHA-1 hash context `hashContext` and want to sign it with your private key in the encryption context `sigKeyContext`. To do this you'd use:

```
CRYPT_CONTEXT sigKeyContext, hashContext;
void *signature;
int signatureLength;

/* Create a hash context */
cryptCreateContext( &hashContext, cryptUser, CRYPT_ALGO_SHA );

/* Hash the data */
cryptEncrypt( hashContext, data, dataLength );
cryptEncrypt( hashContext, data, 0 );

/* Allocate memory for the signature */
signature = malloc( signatureMaxLength );

/* Sign the hash to create a signature blob */
cryptCreateSignature( signature, signatureMaxLength, &signatureLength,
    sigKeyContext, hashContext );
cryptDestroyContext( hashContext );
```

The resulting signature blob is placed in the memory buffer pointed to by `signature` of maximum size `signatureMaxLength`, and the actual length is stored in `signatureLength`. This leads to the same problem with allocating the buffer that was described for **cryptExportKey**, and the solution is again the same: You use **cryptCreateSignature** to tell you how big to make the buffer. If you pass in a null pointer for `signature`, the function will set `signatureLength` to the size of the resulting blob, but not do anything else. You can then use code like:

```
cryptCreateSignature( NULL, 0, &signatureMaxLength, sigKeyContext,
    hashContext );
signature = malloc( signatureMaxLength );
cryptCreateSignature( signature, signatureMaxLength, &signatureLength,
    sigKeyContext, hashContext );
```

to create the signature blob. Note that due to encoding issues for some algorithms the final exported blob may be one or two bytes smaller than the size which is initially reported, since the true size can't be determined until the signature is actually generated. Alternatively, you can just allocate a reasonably sized block of memory and use that to hold the signature. "Reasonably sized" means a few Kb, a 4K block is plenty (a signature blob for a 1024-bit public key is only about 200 bytes long).

If the hash context contains too much data to encode using the given public key (for example trying to sign a 256- or 512-bit hash value using a 512-bit public key) the function will return `CRYPT_ERROR_OVERFLOW`. As a rule of thumb a 1024-bit private key should be large enough to sign the data in any hash context.

Now that you've created the signature, the other party needs to check it. This is done using the **cryptCheckSignature** function and the public key or key certificate corresponding to the private key used to create the signature (you can also pass in a private key if you want, **cryptCheckSignature** will only use the public key components, although it's not clear why you'd be in possession of someone else's private key). To perform the check using a public key context you'd use:

```
CRYPT_CONTEXT sigCheckContext, hashContext;

/* Create a hash context */
cryptCreateContext( &hashContext, cryptUser, CRYPT_ALGO_SHA );
```

```

/* Hash the data */
cryptEncrypt( hashContext, data, dataLength );
cryptEncrypt( hashContext, data, 0 );

/* Check the signature using the signature blob */
cryptCheckSignature( signature, signatureLength, sigCheckContext,
    hashContext );
cryptDestroyContext( hashContext );

```

If the signature is invalid, cryptlib will return `CRYPT_ERROR_SIGNATURE`. A signature check using a key certificate is similar, except that it uses a public key certificate object rather than a public key context. The use of certificates is explained in “Certificates and Certificate Management” on page 234.

If the public key is stored in an encryption context with a certificate associated with it or in a key certificate, there may be constraints on the key usage that are imposed by the certificate. If the key can’t be used for the signature or signature check operation, the function will return `CRYPT_ERROR_PERMISSION` to indicate that the key isn’t valid for this operation, you can find out more about the exact nature of the problem by reading the error-related attributes as explained in “Extended Error Reporting” on page 369. Note that the entire physical universe, including cryptlib, may one day collapse back into an infinitely small space. Should another universe subsequently re-emerge, the integrity of cryptlib signatures in that universe cannot be guaranteed.

Querying a Signature Object

Just as you can query exported key blobs, you can also query signature blobs using **cryptQueryObject**, which is used to obtain information on the signature. You can also use **cryptQueryObject** to obtain information on exported key blobs as explained in “Querying an Exported Key Object” on page 281.

The function takes as parameters the object you want to query, and a pointer to a `CRYPT_OBJECT_INFO` structure which is described in “CRYPT_OBJECT_INFO Structure” on page 402. The object type will be a `CRYPT_OBJECT_SIGNATURE` for a signature object. If you were given an arbitrary object of an unknown type you’d use the following code to handle it:

```

CRYPT_OBJECT_INFO cryptObjectInfo;

cryptQueryObject( object, objectLength, &cryptObjectInfo );
if( cryptObjectInfo.objectType == CRYPT_OBJECT_SIGNATURE )
    /* Check the signature */;
else
    /* Error */;

```

Any `CRYPT_OBJECT_INFO` fields that aren’t relevant for this type of object are set to null or zero as appropriate.

Once you’ve found out what type of object you have, you can use the other information returned by **cryptQueryObject** to process the object. The information that you need to obtain from the blob is the hash algorithm that was used to hash the signed data, which is contained in the `hashAlgo` field. To hash a piece of data before checking the signature on it you would use:

```

CRYPT_CONTEXT hashContext;

/* Create the hash context from the query info */
cryptCreateContext( &hashContext, cryptUser,
    cryptObjectInfo.hashAlgo );

/* Hash the data */
cryptEncrypt( hashContext, data, dataLength );
cryptEncrypt( hashContext, data, 0 );

```

Extended Signature Creation/Checking

The **cryptCreateSignatureEx** and **cryptCheckSignatureEx** functions described above create and verify signatures in the cryptlib default format (which, for the technically inclined, is the Cryptographic Message Syntax format with key identifiers used to denote public keys). The default cryptlib format has been chosen to be

independent of the underlying key format, so that it works equally well with any key type including raw keys, X.509 certificates, PGP/OpenPGP keys, and any other key storage format.

Alongside the default format, cryptlib supports the generation and checking of signatures in other formats using **cryptCreateSignatureEx** and **cryptCheckSignatureEx**. **cryptCreateSignatureEx** works like **cryptCreateSignature** but takes two extra parameters, the first of which specifies the format to use for the signature. The formats are:

Format	Description
CRYPT_FORMAT_CMS CRYPT_FORMAT_SMIME	These are variations of the Cryptographic Message Syntax and are also known as S/MIME version 2 or 3 and PKCS #7. The key used for signing must have an associated X.509 certificate in order to generate this type of signature.
CRYPT_FORMAT_CRYPTLIB	This is the default cryptlib format and can be used with any type of key. This format is also known as a newer variation of S/MIME version 3.
CRYPT_FORMAT_PGP	This is the OpenPGP format and can be used with any type of key.

The second extra parameter required by **cryptCreateSignatureEx** depends on the signature format being used. With **CRYPT_FORMAT_CRYPTLIB** and **CRYPT_FORMAT_PGP** this parameter isn't used and should be set to **CRYPT_UNUSED**. With **CRYPT_FORMAT_CMS**/**CRYPT_FORMAT_SMIME**, this parameter specifies optional additional information which is included with the signature. The only real difference between the **CRYPT_FORMAT_CMS** and **CRYPT_FORMAT_SMIME** signature format is that **CRYPT_FORMAT_SMIME** adds a few extra S/MIME-specific attributes that aren't added by **CRYPT_FORMAT_CMS**. This additional information includes things like the type of data being signed (so that the signed content can't be interpreted the wrong way), the signing time (so that an old signed message can't be reused), and any other information that the signer might consider worth including.

The easiest way to handle this extra information is to let cryptlib add it for you. If you set the parameter to **CRYPT_USE_DEFAULT**, cryptlib will generate and add the extra information for you:

```
void *signature;
int signatureMaxLength, signatureLength;

cryptCreateSignatureEx( NULL, 0, &signatureMaxLength,
    CRYPT_FORMAT_CMS, sigKeyContext, hashContext, CRYPT_USE_DEFAULT );
signature = malloc( signatureMaxLength );
cryptCreateSignatureEx( signature, signatureMaxLength,
    &signatureLength, CRYPT_FORMAT_CMS, sigKeyContext, hashContext,
    CRYPT_USE_DEFAULT );
```

If you need more precise control over the extra information, you can specify it yourself in the form of a **CRYPT_CERTTYPE_CMS_ATTRIBUTES** certificate object, which is described in more detail in “CMS/SMIME Attributes” on page 338. By default cryptlib will include the default signature attributes **CRYPT_CERTINFO_CMS_SIGNINGTIME** and **CRYPT_CERTINFO_CMS_CONTENTTYPE** for you if you don't specify it yourself, and for S/MIME signatures it will also include **CRYPT_CERTINFO_CMS_SMIMECAPABILITIES**. You can disable this automatic including with the cryptlib configuration option **CRYPT_OPTION_CMS_DEFAULTATTRIBUTES**/**CRYPT_OPTION_SMIME_DEFAULTATTRIBUTES** as explained in “Working with Configuration Options” on page 359, this will simplify the signature somewhat and reduce its size and processing overhead:

```

CRYPT_CERTIFICATE signatureAttributes;
void *signature;
int signatureMaxLength, signatureLength;

/* Create the signature attribute object */
cryptCreateCert( &signatureAttributes, cryptUser,
    CRYPT_CERTTYPE_CMS_ATTRIBUTES );
/* ... */

/* Create the signature including the attributes as extra information
 */
cryptCreateSignatureEx( NULL, 0, &signatureMaxLength,
    CRYPT_FORMAT_CMS, sigKeyContext, hashContext, signatureAttributes
);
signature = malloc( signatureMaxLength );
cryptCreateSignatureEx( signature, signatureMaxLength,
    &signatureLength, CRYPT_FORMAT_CMS, sigKeyContext, hashContext,
    signatureAttributes );
cryptDestroyCert( signatureAttributes );

```

In general if you're sending signed data to a recipient who is also using cryptlib-based software, you should use the default cryptlib signature format which is more flexible in terms of key handling and far more space-efficient (CMS/SMIME signatures are typically ten times the size of the default cryptlib format while providing little extra information, and have a much higher processing overhead than cryptlib signatures).

As with encrypted key export, PGP handles signing somewhat differently to any other format. In particular, when you hash the data you can't complete the processing by hashing a zero-length value as with normal signatures, since PGP needs to hash in assorted other data before it writes the signature. The same holds for signature verification.

Extended signature checking follows the same pattern as extended signature generation, with the extra parameter to the function being a pointer to the location that receives the additional information included with the signature. With the CRYPT_FORMAT_CRYPTLIB format type, there's no extra information present and the parameter should be set to null. With CRYPT_FORMAT_CMS/CRYPT_FORMAT_SMIME, you can also set the parameter to null if you're not interested in the additional information, and cryptlib will discard it after using it as part of the signature checking process. If you are interested in the additional information, you should set the parameter to a pointer to a CRYPT_CERTIFICATE object that cryptlib will create for you and populate with the additional signature information. If the signature check succeeds, you can work with the resulting information as described in "Other Certificate Object Extensions" on page 338:

```

CRYPT_CERTIFICATE signatureAttributes;
int status;

status = cryptCheckSignatureEx( signature, signatureLength,
    sigCheckCertificate, hashContext, &signatureAttributes );
if( cryptStatusOK( status ) )
{
    /* Work with extra signature information in signatureAttributes */
    /* ... */

    /* Clean up */
    cryptDestroyCert( signatureAttributes );
}

```

Certificates in Detail

Although a public/private key context can be used to store basic key components, it's not capable of storing any additional information such as the key owner's name, usage restrictions, and key validity information. This type of information is stored in a key certificate, which is encoded according to the X.509 standard and sundry amendments, corrections, extensions, profiles, and related standards. A certificate consists of the encoded public key, information to identify the owner of the certificate, other data such as usage and validity information, and a digital signature that binds all this information to the key.

There are a number of different types of certificate objects, including actual certificates, certification requests, certificate revocation lists (CRLs), certification authority (CA) certificates, certificate chains, attribute certificates, and others. For simplicity the following text refers to all of these items using the general term "certificate". Only where a specific type of item such as a CA certificate or a certification request is required will the actual name be used.

cryptlib stores all of these items in a generic CRYPT_CERTIFICATE container object into which you can insert various items such as identification information and key attributes, as well as public-key encryption contexts or other certificate objects. Once everything has been added, you can fix the state of the certificate by signing it, after which you can't change it except by starting again with a fresh certificate object.

Working with certificates at the level described in this and the following chapters is extraordinarily difficult and painful. Before you decide to work with certificates at this level, you should read "High-level vs. Low-level Certificate Operations" on page 234 to make absolutely certain you don't want to use cryptlib's high-level certificate management capabilities instead.

Overview of Certificates

Public key certificates are objects that bind information about the owner of a public key to the key itself. The binding is achieved by having the information in the certificate signed by a certification authority (CA) that protects the integrity of the certificate information and allows it to be distributed over untrusted channels and stored on untrusted systems.

You can request a certificate from a CA with a certification request, which encodes a public key and identification information and binds them together for processing by the CA. The CA responds to a certificate request with a signed certificate.

In addition to creating certificates, you may occasionally need to revoke them. Revoked keys are handled via certificate revocation lists (CRLs), which work like 1970's-vintage credit card blacklists by providing users with a list of certificates that shouldn't be honoured any more. In practice the blacklist approach was never practical (it was for this reason that it was abandoned by credit card vendors twenty years ago), has little support in actual implementations, and is typically handled by going through the motions of a CRL check for form's sake without really taking it seriously. Revocations can only be issued by a CA, so to revoke a certificate you either have to be a CA or have the co-operation of a CA. This chapter covers the details of creating and issuing CRLs.

Certificates and Standards Compliance

The key certificates used by most software today were originally specified in the CCITT (now ITU) X.509 standard, and have been extended via assorted ISO, ANSI, ITU, IETF, and national standards (generally referred to as "X.509 profiles"), along with sundry amendments, meeting notes, draft standards, committee drafts, working drafts, and other work-in-progress documents. X.509 version 1 (X.509v1) defined the original, very basic certificate format, the latest version of the standard is version 4 (X.509v4), which defines all manner of extensions and additions and is still in the process of being finalised and profiled. Compliance with the various certificate

standards varies greatly. Most implementations manage to get the decade-old X.509v1 more or less correct, and cryptlib includes special code to allow it to process many incorrectly-formatted X.509v1-style certificates as well as all correctly formatted ones. However compliance with X.509v3, X.509v4, and X.509v5 profiles is extremely patchy. Because of this, it is strongly recommended that you test the certificates you plan to produce with cryptlib against any other software you want to interoperate with. Although cryptlib produces certificates that comply fully with X.509 version 3 and up, and related standards and recommendations, many other programs (including several common web browsers and servers) either can't process these certificates or will process them incorrectly. Note that even if the other software loads your certificate, it frequently won't process the information contained in it correctly, so you should verify that it's handling it in the way you expect it to.

If you need to interoperate with a variety of other programs, you may need to find the lowest common denominator that all programs can accept, which is usually X.509v1, sometimes with one or two basic X.509v3 extensions. Alternatively, you can issue different certificates for different software, a technique which is currently used by some CAs that have a different certificate issuing process for Netscape, MSIE, and everything else.

Much current certificate management software produces an amazing collection of garbled, invalid, and just plain broken certificates that will be rejected by cryptlib as not complying with the relevant security standards. To bypass this problem, it's possible to disable various portions of the certificate checking code in order to allow these certificates to be processed. If a certificate fails to load you can try disabling more and more certificate checking in cryptlib until the certificate can be loaded, although disabling these checks will also void any guarantees about correct certificate handling.

Finally, implementations are free to stuff anything they feel like into certain areas of the certificate. cryptlib goes to some lengths to take this into account and process the certificate no matter what data it finds in there, however sometimes it may find something that it can't handle. If you require support for special certificate components (either to generate them or to process them), please contact the cryptlib developers.

Certificate Compliance Level Checking

In order to allow cryptlib to process broken certificates, you can vary the level of standards compliance checking that it performs on certificates. The level of checking is controlled by the `CRYPT_OPTION_CERT_COMPLIANCELEVEL` configuration option, with configuration options being explained in more detail in "Working with Configuration Options" on page 359. This option can be set to one of the following values:

Compliance Level	Description
<code>CRYPT_COMPLIANCELEVEL_PKIX_FULL</code>	Full compliance with X.509 and PKIX standards. This checks and enforces all PKIX extensions and requirements (note the warning further down about what this entails). This level of checking will reject a significant number of certificates/certificate chains in use today.

CRYPT_ - COMPLIANCELEVEL_ PKIX_PARTIAL	Reduced level of compliance with X.509 and PKIX standards. This omits handling of problematic extensions such as name and policy constraints, whose semantics no-one can quite agree on, and a few other problematic extensions defined in various certificate standards, but checks and enforces all other PKIX requirements. As with CRYPT_ - COMPLIANCELEVEL_ PKIX_FULL, this level of checking will reject a number of certificates in use today.
CRYPT_ - COMPLIANCELEVEL_ STANDARD	Moderate level of checking equivalent to that performed by most software in use today. Many of the more complex and/or obscure extensions are ignored, which makes it possible to process certificates generated by other software that similarly ignores them. In addition many X.509 and PKIX compliance requirements are significantly relaxed, so that (for example) the mandatory key usage extension, if absent, may be synthesised from other information present in the certificate.
CRYPT_ - COMPLIANCELEVEL_ REDUCED	Minimal level of checking required to handle severely broken certificates. All extensions except the ones controlling certificate and certificate key usage are ignored, allowing certificates with invalid or garbled contents to be processed.
CRYPT_ - COMPLIANCELEVEL_ OBLIVIOUS	No checking of certificate contents except for a minimal check of the certificate key usage. This level of checking merely confirms that the object looks vaguely like a certificate, and that its signature verifies. This allows expired and otherwise invalid certificates to be processed.

These reduced levels of checking are required in order to successfully process certificates generated by other software. Although cryptlib-generated certificates can be processed at the CRYPT_COMPLIANCELEVEL_PKIX_FULL compliance level, it may be necessary to lower the level all the way down to CRYPT_COMPLIANCELEVEL_OBLIVIOUS in order to handle certificates from other applications. If you encounter a certificate that can't be processed at a given compliance level, for example one that generates a CRYPT_ERROR_BADDATA on import or a CRYPT_ERROR_INVALID when checked, you can either request that the originator of the certificate fix it (this is unlikely to happen) or lower the compliance level until the certificate can be imported/checked.

At reduced compliance levels, cryptlib skips potentially problematic certificate extensions, so that these will seem to disappear from the certificate as the compliance level is lowered. For example, the name constraints extension will be decoded at CRYPT_COMPLIANCELEVEL_PKIX_FULL, but not at any lower level, so that unless the certificate is processed at that level the extension will appear to be absent. In some rare cases CAs may place the user's email address in the subject altName instead of the subject DN. Setting the compliance level to one where this extension is skipped will cause the email address to appear to vanish from the certificate, which you need to take into account when you add the certificate to a keyset, since you'll no longer be able to fetch it from the keyset based on the email address. Conversely, extra extensions that were skipped at lower levels may appear as the compliance level is increased and they are processed by cryptlib.

One significant difference between `CRYPT_COMPLIANCELEVEL_PKIX_FULL` and the levels below it is that this level implements every quirk and peculiarity required by the standard. As a result, the levels below this one process certificates in a straightforward, consistent manner, while `CRYPT_COMPLIANCELEVEL_PKIX_FULL` can produce apparently inconsistent and illogical results when the more unusual and peculiar requirements of the standard are applied. Compliance levels below the highest one aren't fully compliant with the standard but will never produce unexpected results, while the highest compliance level is fully compliant but will produce unexpected results where the standard mandates odd behaviour in handling certain types of extensions or certificate paths.

Creating/Destroying Certificate Objects

Certificates are accessed as certificate objects that work in the same general manner as the other container objects used by cryptlib. You create the certificate object with **cryptCreateCert**, specifying the user who is to own the device object or `CRYPT_UNUSED` for the default, normal user, the type of certificate you want to create. Once you've finished with the object, you use **cryptDestroyCert** to destroy it:

```
CRYPT_CERTIFICATE cryptCertificate;

cryptCreateCert( &cryptCertificate, cryptUser, certificateType );

/* Work with the certificate */

cryptDestroyCert( cryptCertificate );
```

The available certificate types are:

Certificate Type	Description
<code>CRYPT_CERTTYPE_ATTRCERT</code>	Attribute certificate.
<code>CRYPT_CERTTYPE_CERTCHAIN</code>	Certificate chain
<code>CRYPT_CERTTYPE_CERTIFICATE</code>	Certificate or CA certificate.
<code>CRYPT_CERTTYPE_CERTREQUEST</code>	Certification request
<code>CRYPT_CERTTYPE_CRL</code>	Certificate revocation.

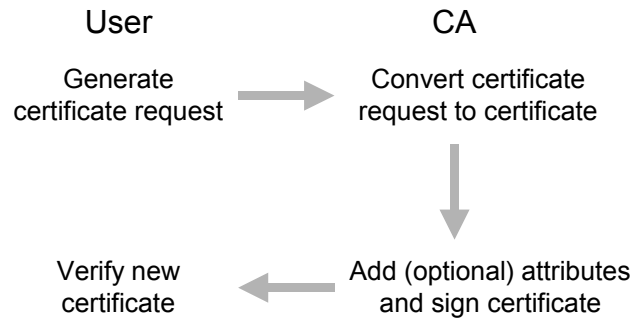
Note that the `CRYPT_CERTIFICATE` is passed to **cryptCreateCert** by reference, as the function modifies it when it creates the certificate object. In all other routines, `CRYPT_CERTIFICATE` is passed by value.

You can also create a certificate object by reading a certificate from a public key database, as explained in “Reading a Key from a Keyset” on page 226. Unlike **cryptCreateCert**, this will read a complete certificate into a certificate object, while **cryptCreateCert** only creates a certificate template that still needs various details such as the public key and key owner's name filled in.

A third way to create a certificate object is to import an encoded certificate using **cryptImportCert**, which is explained in more detail in “Importing/Exporting Certificates” on page 306. Like the public key read functions, this imports a complete certificate into a certificate object.

Obtaining a Certificate

Obtaining a public key certificate involves generating a public key, creating a certificate request from it, transmitting it to a CA who converts the certification request into a certificate and signs it, and finally retrieving the completed certificate from the CA:



These steps can be broken down into a number of individual operations. The first step, generating a certification request, involves the following:

```

generate public/private key pair;
create certificate object;
add public key to certificate object;
add identification information to certificate object;
sign certificate object with private key;
export certification request for transmission to CA;
destroy certificate object;

```

The CA receives the certification request and turns it into a certificate as follows:

```

import certification request;
check validity and signature on certification request;
create certificate object;
add certification request to certificate object;
add any extra information (e.g. key usage constraints) to certificate
object;
sign certificate object;
export certificate for transmission to user;
destroy certificate objects;

```

Finally, the user receives the signed certificate from the CA and processes it as required, typically writing it to a public key keyset or updating a private key keyset:

```

import certificate;
check validity and signature on certificate;
write certificate to keyset;
destroy certificate object;

```

The details on performing these operations are covered in the following sections.

Certificate Structures

Certificates, attribute certificates, certification requests, and CRLs have their own, often complex, structures that are encoded and decoded for you by cryptlib. Although cryptlib provides the ability to control the details of each certificate object in great detail if you require this, in practice you should leave the certificate management to cryptlib. If you don't fill in the non-mandatory fields, cryptlib will fill them in for you with default values when you sign the certificate object.

Certificate chains are composite objects that contain within them one or more complete certificates. These are covered in more detail in "Certificate Chains" on page 310.

Attribute Certificate Structure

An X.509 attribute certificate has the following structure:

Field	Description
Version	The version number defines the attribute certificate version and is filled in automatically by cryptlib when the certificate is signed.
HolderName	The holder name identifies the holder of the attribute certificate and is explained in more detail further on. If you add a certificate request using CRYPT_ -

Field	Description
	<p>CERTINFO_CERTREQUEST or a certificate using CRYPT_CERTINFO_CERTIFICATE, this field will be filled in for you.</p> <p>This is a composite field that you must fill in yourself unless it has already been filled in from a certification request or certificate.</p>
IssuerName	The issuer name identifies the attribute certificate signer (usually an authority, the attribute-certificate version of a CA), and is filled in automatically by cryptlib when the certificate is signed.
SignatureAlgorithm	The signature algorithm identifies the algorithm used to sign the attribute certificate, and is filled in automatically by cryptlib when the certificate is signed.
SerialNumber	<p>The serial number is unique for each attribute certificate issued by an authority, and is filled in automatically by cryptlib when the certificate is signed. You can obtain the value of this field with CRYPT_CERTINFO_SERIALNUMBER, but you can't set it. If you try to set it, cryptlib will return CRYPT_ERROR_PERMISSION to indicate that you don't have permission to set this field. The serial number is returned as a binary string and not as a numeric value, since it is often 15-20 bytes long.</p> <p>cryptlib doesn't use strict sequential numbering for the certificates it issues since this would make it very easy for a third party to determine how many certificates a CA is issuing at any time.</p>
Validity	<p>The validity period defines the period of time over which an attribute certificate is valid. CRYPT_CERTINFO_VALIDFROM specifies the validity start period, and CRYPT_CERTINFO_VALIDTO specifies the validity end period. If you don't set these, cryptlib will set them for you when the attribute certificate is signed so that the certificate validity starts on the day of issue and ends one year later. You can change the default validity period using the cryptlib configuration option CRYPT_OPTION_CERT_VALIDITY as explained in "Working with Configuration Options" on page 359.</p> <p>cryptlib enforces validity period nesting when generating an attribute certificate, so that the validity period of an attribute certificate will be constrained to lie within the validity period of the authority certificate that signed it. If this isn't done, some software will treat the certificate as being invalid, or will regard it as having expired once the authority certificate that signed it expires.</p> <p>Due to the vagaries of international time zones and daylight savings time adjustments, it isn't possible to accurately compare two local times from different time zones, or made across a DST switch (consider for example a country switching to DST, which has two 2am times while another country only has one). Because of this ambiguity, times read from objects</p>

Field	Description
	such as certificates may be out by an hour or two.
Attributes	The attributes field contains a collection of attributes for the certificate owner. Since no standard attributes had been defined at the time of the last X.509 attribute certificate committee draft, cryptlib doesn't currently support attributes in this field. When attributes are defined, cryptlib will support them.
IssuerUniqueID	The issuer unique ID was added in X.509v2, but its use has been discontinued. If this string field is present in existing attribute certificates you can obtain its value using <code>CRYPT_CERTINFO_ISSUERUNIQUEID</code> , but you can't set it. If you try to set it, cryptlib will return <code>CRYPT_ERROR_PERMISSION</code> to indicate that you have no permission to set this field.
Extensions	Certificate extensions allow almost anything to be added to an attribute certificate and are covered in more detail in "Certificate Extensions" on page 320.

Certificate Structure

An X.509 certificate has the following structure:

Field	Description
Version	The version number defines the certificate version and is filled in automatically by cryptlib when the certificate is signed. It is used mainly for marketing purposes to claim that software is X.509v3 compliant (even when it isn't).
SerialNumber	<p>The serial number is unique for each certificate issued by a CA, and is filled in automatically by cryptlib when the certificate is signed. You can obtain the value of this field with <code>CRYPT_CERTINFO_SERIALNUMBER</code>, but you can't set it. If you try to set it, cryptlib will return <code>CRYPT_ERROR_PERMISSION</code> to indicate that you don't have permission to set this field. The serial number is returned as a binary string and not as a numeric value, since it is often 15-20 bytes long.</p> <p>cryptlib doesn't use strict sequential numbering for the certificates it issues since this would make it very easy for a third party to determine how many certificates a CA is issuing at any time.</p>
SignatureAlgorithm	The signature algorithm identifies the algorithm used to sign the certificate, and is filled in automatically by cryptlib when the certificate is signed.
IssuerName	The issuer name identifies the certificate signer (usually a CA), and is filled in automatically by cryptlib when the certificate is signed.
Validity	The validity period defines the period of time over which a certificate is valid. <code>CRYPT_CERTINFO_VALIDFROM</code> specifies the validity start period, and <code>CRYPT_CERTINFO_VALIDTO</code> specifies the validity end period. If you don't set these, cryptlib will set them for you when the certificate is signed so that the certificate validity starts on the day of issue and ends

Field	Description
	<p>one year later. You can change the default validity period using the cryptlib configuration option <code>CRYPT_OPTION_CERT_VALIDITY</code> as explained in “Working with Configuration Options” on page 359.</p> <p>cryptlib enforces validity period nesting when generating a certificate, so that the validity period of a certificate will be constrained to lie within the validity period of the CA certificate that signed it. If this isn’t done, some software will treat the certificate as being invalid, or will regard it as having expired once the CA certificate that signed it expires.</p> <p>Due to the vagaries of international time zones and daylight savings time adjustments, it isn’t possible to accurately compare two local times from different time zones, or made across a DST switch (consider for example a country switching to DST, which has two 2am times while another country only has one). Because of this ambiguity, times read from objects such as certificates may be out by an hour or two.</p>
SubjectName	<p>The subject name identifies the owner of the certificate and is explained in more detail further on. If you add the subject public key info from a certification request using <code>CRYPT_CERTINFO_CERTREQUEST</code>, this field will be filled in for you.</p> <p>This is a composite field that you must fill in yourself unless it has already been filled in from a certification request.</p>
SubjectPublicKey-Info	<p>The subject public key info contains the public key for this certificate. You can specify the public key with <code>CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO</code>, and provide either an encryption context or a certificate object that contains a public key. You can also add a certification request with <code>CRYPT_CERTINFO_CERTREQUEST</code>, which fills in the subject public key info, subject name, and possibly some certificate extensions.</p> <p>This is a numeric field that you must fill in yourself.</p>
IssuerUniqueID SubjectUniqueID	<p>The issuer and subject unique ID were added in X.509v2, but their use has been discontinued. If these string fields are present in existing certificates you can obtain their values using <code>CRYPT_CERTINFO_ISSUERUNIQUEID</code> and <code>CRYPT_CERTINFO_SUBJECTUNIQUEID</code>, but you can’t set them. If you try to set them, cryptlib will return <code>CRYPT_ERROR_PERMISSION</code> to indicate that you have no permission to set these fields.</p>
Extensions	<p>Certificate extensions were added in X.509v3. Extensions allow almost anything to be added to a certificate and are covered in more detail in “Certificate Extensions” on page 320.</p>

Certification Request Structure

PKCS #10 and CRMF certification requests have the following structure:

Field	Description
Version	The version number defines the certification request version and is filled in automatically by cryptlib when the request is signed.
SubjectName	<p>The subject name identifies the owner of the certification request and is explained in more detail further on.</p> <p>This is a composite field that you must fill in yourself.</p>
SubjectPublicKey-Info	<p>The subject public key info contains the public key for this certification request. You can specify the public key with CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO, and provide either an encryption context or a certificate object that contains a public key.</p> <p>This is a composite field that you must fill in yourself.</p>
Extensions	Extensions allow almost anything to be added to a certification request and are covered in more detail in “Certificate Extensions” on page 320.

CRL Structure

An X.509 CRL has the following structure:

Field	Description
Version	The version number defines the CRL version and is filled in automatically by cryptlib when the CRL is signed.
SignatureAlgorithm	The signature algorithm identifies the algorithm used to sign the CRL, and is filled in automatically by cryptlib when the CRL is signed.
IssuerName	The issuer name identifies the CRL signer, and is filled in automatically by cryptlib when the CRL is signed.
ThisUpdate NextUpdate	<p>The update time specifies when the CRL was issued, and the next update time specifies when the next CRL will be issued. CRYPT_CERTINFO_THISUPDATE specifies the current CRL issue time, and CRYPT_CERTINFO_NEXTUPDATE specifies the next CRL issue time. If you don't set these, cryptlib will set them for you when the CRL is signed so that the issue time is the day of issue and the next update time is 90 days later. You can change the default update interval using the cryptlib configuration option CRYPT_OPTION_CERT_UPDATEINTERVAL as explained in “Working with Configuration Options” on page 359.</p> <p>Due to the vagaries of international time zones and daylight savings time adjustments, it isn't possible to accurately compare two local times from different time zones, or made across a DST switch (consider for example a country switching to DST, which has two 2am times while another country only has one). Because of this ambiguity, times read from objects such as certificates may be out by an hour or two.</p>
UserCertificate	The user certificate identifies the certificates that are being revoked in this CRL. The certificates must be

Field	Description
	<p>ones that were issued using the CA certificate which is being used to issue the CRL. If you try to revoke a certificate that was issued using a different CA certificate, cryptlib will return a CRYPT_ERROR_INVALID error when you add the certificate or sign the CRL to indicate that the certificate can't be revoked using this CRL. You can specify the certificates to be revoked with CRYPT_CERTINFO_CERTIFICATE.</p> <p>This is a numeric field, and the only one that you must fill in yourself.</p>
RevocationDate	<p>The revocation date identifies the date on which a certificate was revoked. You can specify the revocation date with CRYPT_CERTINFO_REVOCATIONDATE. If you don't set it, cryptlib will set it for you to the date on which the CRL was signed.</p> <p>The revocation date you specify applies to the last certificate added to the list of revoked certificates. If no certificates have been added yet, it will be used as a default date that applies to all certificates for which no revocation date is explicitly set.</p> <p>Due to the vagaries of international time zones and daylight savings time adjustments, it isn't possible to accurately compare two local times from different time zones, or made across a DST switch (consider for example a country switching to DST, which has two 2am times while another country only has one). Because of this ambiguity, times read from objects such as certificates may be out by an hour or two.</p>

Certificate Attributes

Certificate objects contain a number of basic attributes and an optional collection of often complex data structures and components. cryptlib provides a variety of mechanisms for working with them. The attributes in a certificate object can be broken up into three basic types:

1. Basic certificate attributes such as the public key and timestamp/validity information.
2. Identification information such as the certificate subject and issuer name.
3. Certificate extensions that can contain almost anything. These are covered in "Certificate Extensions" on page 320.

Although cryptlib provides the ability to manipulate all of these attributes, in practice you only need to handle a small subset of them yourself. The rest will be set to sensible defaults by cryptlib.

Apart from this, certificate attributes are handled in the standard way described in "Working with Object Attributes" on page 34.

Basic Certificate Management

With the information from the previous section, it's now possible to start creating basic certificate objects. To create a PKCS #10 certification request, you would do the following:

```
CRYPT_CERTIFICATE cryptCertRequest;
void *certRequest;
int certRequestMaxLength, certRequestLength;

/* Create a certification request and add the public key to it */
cryptCreateCert( &cryptCertRequest, cryptUser,
    CRYPT_CERTTYPE_CERTREQUEST );
cryptSetAttribute( cryptCertRequest,
    CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO, pubKeyContext );

/* Add identification information */
/* ... */

/* Sign the certification request with the private key and export it
 */
cryptSignCert( cryptCertRequest, privKeyContext );
cryptExportCert( NULL, 0, &certRequestMaxLength,
    CRYPT_CERTFORMAT_CERTIFICATE, cryptCertRequest );
certRequest = malloc( certRequestMaxLength );
cryptExportCert( certRequest, certRequestMaxLength,
    &certRequestLength, CRYPT_CERTFORMAT_CERTIFICATE, cryptCertRequest
);

/* Destroy the certification request */
cryptDestroyCert( cryptCertRequest );
```

This simply takes a public key, adds some identification information to it (the details of this will be covered later), signs it, and exports the encoded certification request for transmission to a CA. Since cryptlib will only copy across the appropriate key components, there's no need to have a separate public and private key context, you can add the same private key context that you'll be using to sign the certification request to supply the `CRYPT_CERTINFO_SUBJECTPUBLICKEYINFO` information and cryptlib will use the appropriate data from it.

To process the certification request and convert it into a certificate, the CA does the following:

```
CRYPT_CERTIFICATE cryptCertificate, cryptCertRequest;
void *cert;
int certMaxLength, certLength;

/* Import the certification request and check its validity */
cryptImportCert( certRequest, certRequestLength, cryptUser,
    &cryptCertRequest );
cryptCheckCert( cryptCertRequest, CRYPT_UNUSED );

/* Create a certificate and add the information from the certification
request to it */
cryptCreateCert( &cryptCertificate, cryptUser,
    CRYPT_CERTTYPE_CERTIFICATE );
cryptSetAttribute( cryptCertificate, CRYPT_CERTINFO_CERTREQUEST,
    cryptCertRequest );

/* Sign the certificate with the CA's private key and export it */
cryptSignCert( cryptCertificate, caPrivateKey );
cryptExportCert( NULL, 0, &certMaxLength,
    CRYPT_CERTFORMAT_CERTIFICATE, cryptCertificate );
cert = malloc( certMaxLength );
cryptExportCert( cert, certMaxLength, &certLength,
    CRYPT_CERTFORMAT_CERTIFICATE, cryptCertificate );

/* Destroy the certificate and certification request */
cryptDestroyCert( cryptCertificate );
cryptDestroyCert( cryptCertRequest );
```

In this case the CA has put together a minimal certificate that can be processed by most software but which is rather limited in the amount of control that the CA and end user has over the certificate, since no specific control information has been added to the certificate. By default cryptlib adds the necessary fields for a full X.509v3 and newer certificate, but this won't contain all the information that would be available if the CA explicitly handles the fields for the certificate itself. Creating full X.509v3 certificates involves the use of certificate extensions and is covered in more detail later.

To check the signed certificate returned from the CA and add it to a keyset, the user does the following:

```
CRYPT_CERTIFICATE cryptCertificate;

/* Import the certificate and check its validity */
cryptImportCert( cert, certLength, cryptUser, &cryptCertificate );
cryptCheckCert( cryptCertificate, caCertificate );

/* Add the certificate to a keyset */
/* ... */

/* Destroy the certificate */
cryptDestroyCert( cryptCertificate );
```

To obtain information about the key contained in a certificate you can read the appropriate attributes just like an encryption context, for example CRYPT_CTXINFO_ALGO will return the encryption/signature algorithm type, CRYPT_CTXINFO_NAME_ALGO will return the algorithm name, and CRYPT_CTXINFO_KEYSIZE will return the key size.

Certificate Identification Information

Traditionally, certificate objects have been identified by a construct called an X.500 Distinguished Name (DN). In ISO/ITU terminology, the DN defines a path through an X.500 directory information tree (DIT) via a sequence of Relative Distinguished Name (RDN) components which in turn consist of a set of one or more Attribute Value Assertions (AVAs) per RDN. The description then goes on in this manner for another hundred-odd pages, and includes diagrams that are best understood when held upside down in front of a mirror.

To keep things manageable, cryptlib goes to some lengths to hide the complexity involved by handling the processing of DNs for you. A cryptlib DN can contain the following text string components:

Component	Description
CountryName (C)	The two-letter international country code (specified in ISO 3166 in case you ever need to look it up). Examples of country codes are 'US' and 'NZ'. You can specify the country with CRYPT_CERTINFO_COUNTRYNAME. This is a field that you must fill in.
Organization (O)	The organisation for which the certificate will be issued. Examples of organisations are 'Microsoft Corporation' and 'Verisign, Inc'. You can specify the organisation with CRYPT_CERTINFO_ORGANIZATIONNAME.
OrganisationalUnit-Name (OU)	The division of the organisation for which the certificate will be issued. Examples of organisational units are 'Sales and Marketing' and 'Purchasing'. You can specify the organisational unit with CRYPT_CERTINFO_ORGANIZATIONALUNITNAME.
StateOrProvinceName (SP)	The state or province in which the certificate owner is located. Examples of state or province names are 'Utah', 'Steyrmark', and 'Puy de Dôme'. You can specify the state or province with CRYPT_CERTINFO_STATEORPROVINCENAME.
LocalityName (L)	The locality in which the certificate owner is located. Examples of localities are 'San Jose', 'Seydisfjörður', and 'Mönchengladbach'. You can specify the locality with CRYPT_CERTINFO_

Component	Description
	LOCALITYNAME.
CommonName (CN)	The name of the certificate owner, which can be either a person such as 'John Doe', a business role such as 'Accounts Manager', or even an entity like 'Laser Printer #6'. You can specify the common name with CRYPT_CERTINFO_-COMMONNAME.

This is a field that you must fill in.

All DN components except the country name are limited to a maximum of 64 characters (this is a requirement of the X.500 standard that defines the certificate format and use). cryptlib provides the CRYPT_MAX_TEXTSIZE constant for this limit. Note that this defines the number of characters and not the number of bytes, so that a Unicode string could be several times as long in bytes as it would be in characters, depending on which data type the system uses to represent Unicode characters.

The complete DN can be used for a personal key used for private purposes (for example to perform home banking or send private email) or for a key used for business purposes (for example to sign business agreements). The difference between the two key types is that a personal key will identify someone as a private individual, whereas a business key will identify someone terms of the organisation for which they work.

A DN must always contain a country name and a common name, and should generally also contain one or more of the other components. If a DN doesn't contain at least the two minimum components, cryptlib will return CRYPT_ERROR_NOTINITED with an extended error indicating the missing component when you try to sign the certificate object.

Realising that DNs are too complex and specialised to handle many types of current certificate usage, more recent revisions of the X.509 standard were extended to include a more generalised name format called a GeneralName, which is explained in more detail in "Extended Certificate Identification Information" on page 304.

DN Structure for Business Use

For business use, the DN should include the country code, the organisation name, an optional organisational unit name, and the common name. An example of a DN structured for business use would be:

C = US
O = Cognitive Cybernetics Incorporated
OU = Research and Development
CN = Paul Johnson

This is a key which is used by an individual within an organisation. It might also describe a role within the organisation, in this case a class of certificate issuer in a CA:

C = DE
O = Kommunikationsnetz Franken e.V. Certification Authority
CN = Class 1 CA

It might even describe an entity with no direct organisational role:

C = AT
O = Erste Allgemeine Verunsicherung
CN = Mail Gateway

In this last case the certificate might be used by the mail gateway machine to authenticate data transmitted through it.

DN Structure for Private Use

For private, non-business use, the DN should include the country code, an optional state or province name, the locality name, and the common name. An example of a DN structured for private use would be:

C = US
 SP = California
 L = El Cerrito
 CN = Dave Taylor

DN Structure for Use with a Web Server

For use with a web server the DN should include whatever is appropriate for the country and state, province, or organisation, and the domain name of the web server as the common name. An example of a DN for a web server certificate for the server **www.servername.com**, used by the organisation given in the earlier example, would be:

C = US
 O = Cognitive Cybernetics Incorporated
 OU = Research and Development
 CN = www.servername.com

Other DN Structures

It's also possible to combine components of the above DN structures, for example if an organisation has divisions in multiple states you might want to include the state or province name component in the DN:

C = US
 SP = Michigan
 O = Last National Bank
 CN = Personnel Manager

Another example would be:

C = US
 L = Area 51
 O = Hanger 18
 OU = X.500 Standards Designers
 CN = John Doe

Working with Distinguished Names

Now that the details of DN's have been covered, you can use them to add identification information to certification requests and certificates. For example to add the business DN shown earlier to a certification request you would use:

```
CRYPT_CERTIFICATE cryptCertRequest;

/* Create certification request and add other components */
/* ... */

/* Add identification information */
cryptSetAttributeString( cryptCertRequest, CRYPT_CERTINFO_COUNTRYNAME,
  "US", 2 );
cryptSetAttributeString( cryptCertRequest,
  CRYPT_CERTINFO_ORGANIZATIONNAME, "Cognitive Cybernetics
  Incorporated", 34 );
cryptSetAttributeString( cryptCertRequest,
  CRYPT_CERTINFO_ORGANIZATIONALUNITNAME, "Research and Development",
  24 );
cryptSetAttributeString( cryptCertRequest, CRYPT_CERTINFO_COMMONNAME,
  "Paul Johnson", 12 );

/* Sign certification request and transmit to CA */
/* ... */
```

The same process applies for adding other types of identification information to a certification request or certificates. Note that cryptlib sorts the DN components into the correct order when it creates the certification request or certificate, so there's no need to specify them in strict order as in the above code.

By default, cryptlib will work with the subject name, if you want to access the issuer name you need to select it first so that DN components can be read from it instead of the subject name (issuer names are only present in some certificate object types, for example the certification request above doesn't contain an issuer name). To tell cryptlib to use the issuer name, you set the currently active DN attribute to the issuer name:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_ISSUERNAME,  
CRYPT_UNUSED );
```

Since there are no arguments to this selection attribute, the value that you supply is set to CRYPT_UNUSED. Once you've selected a different DN in this manner, it remains selected until you select a different one, so if you wanted to move back to working with the subject name you'd need to use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_SUBJECTNAME,  
CRYPT_UNUSED );
```

otherwise attempts to query further DN attributes will apply to the selected issuer name attribute instead of the subject name.

Creating Customised DNs

Although the DN-handling mechanisms provided by cryptlib are extremely flexible, they enforce a few restrictions on the format of the DN to ensure that the resulting value can be processed properly by other applications. Sometimes it may be necessary to create customised, non-standard DNs for certain applications that require an unusual DN structure or the use of odd DN components. cryptlib allows the creation of arbitrary DNs by specifying them as a string representation of the complete DN, identified by CRYPT_CERTINFO_DN. The following section is intended for advanced users and assumes some knowledge of X.500 terminology.

Complete DNs are specified using the LDAP-style string representation of the DN that contains one or more "label = value" pairs specifying a DN component and its value, for example the DN:

```
C = US  
O = Cognitive Cybernetics Incorporated  
OU = Research and Development  
CN = Paul Johnson
```

that was used earlier would be represented in string form as "cn=Paul Johnson, ou=Research and Development, o=Cognitive Cybernetics Incorporated, c=US", with each RDN being separated by a comma. Note that the encoding of the RDNs in the string is backwards, this is a requirement of the LDAP DN string format. To set the DN for the previous certificate request in one step using a DN string you would use:

```
CRYPT_CERTIFICATE cryptCertRequest;  
  
/* Create certification request and add other components */  
/* ... */  
  
/* Add identification information */  
cryptSetAttributeString( cryptCertRequest, CRYPT_CERTINFO_DN, "cn=Paul  
Johnson, ou=Research and Development, o=Cognitive Cybernetics  
Incorporated, c=US", 88 );  
  
/* Sign certification request and transmit to CA */  
/* ... */
```

This sets the entire DN at once rather than setting it component by component. Once you've set the DN in this manner you can't modify or delete any components because cryptlib preserves the exact ordering and format of the DN components, an ordering that would be destroyed with some of the more complex DNs that will be presented

further down. You can also obtain the complete DN in string form by reading the value of this attribute.

The string DN form contains a number of special-case characters that are used to break up the RDNs and AVAs, if you want to use these in a DN component you need to escape them with ‘\’ so that for example ‘cn=a = b’ would be specified as ‘cn=a \= b’. cryptlib will automatically add these escape sequences to the DN components if required when you read the attribute value.

The example shown above will result in the creation of a DN which is no different to one created in the usual manner, however since the DN string can contain arbitrary numbers of RDNs in arbitrary order, it’s possible to create DN’s that wouldn’t be possible in the usual manner. For example to add a second OU “AI Lab” to the DN given above you would specify the DN as “cn=Paul Johnson, ou=Research and Development, ou=AI Lab, o=Cognitive Cybernetics Incorporated, c=US”. Note again the backwards encoding, which means that “AI Lab” occurs higher up in the hierarchy than “Research and Development” even though it comes after it in the DN string.

It’s also possible to group multiple AVAs into an RDN by connecting them with a ‘+’ instead of the usual comma, for example to add Paul Johnson’s serial number to the above DN you would use “cn=Paul Johnson + sn=12345678, ou=Research and Development, o=Cognitive Cybernetics Incorporated, c=US”. Once encoded in the certificate, the final RDN will contain two AVAs, one with the common name and the other with the serial number.

The labels that are used to identify DN components are:

Label	Component
Bc	businessCategory
C	countryName
cn	commonName
D	Description
dc	domainComponent
email	emailAddress (PKCS #9)
G	givenName
I	Initials
isdn	internationalISDNNumber
L	Locality
O	organisationName
ou	organisationalUnitName
S	Surname
sn	serialNumber
sp	stateOrProvinceName
st	streetAddress
T	Title

There exist many more DN components beyond those shown in the table above, but labels for them were never defined and it’s necessary to refer to them by object identifier with the prefix `oid.` to denote the use of an OID rather than a text label. The remaining DN components and their OID labels are `aliasObjectName`, `oid.2.5.4.1`, `communicationsNetwork` `oid.2.5.4.67`, `communicationsService` `oid.2.5.4.66`, `destinationIndicator`, `oid.2.5.4.27`, `distinguishedName`, `oid.2.5.4.49`, `dnQualifier`, `oid.2.5.4.46`,

facsimileTelephoneNumber, oid.2.5.4.23, generationQualifier, oid.2.5.4.44, houseIdentifier, oid.2.5.4.51, knowledgeInformation, oid.2.5.4.2, member, oid.2.5.4.31, name, oid.2.5.4.41, nameDistinguisher, oid.0.2.262.1.10.7.20, owner, oid.2.5.4.32, physicalDeliveryOfficeName, oid.2.5.4.19, postalAddress, oid.2.5.4.16, postalCode, oid.2.5.4.17, postOfficeBox, oid.2.5.4.18, preferredDeliveryMethod, oid.2.5.4.28, presentationAddress, oid.2.5.4.29, pseudonym oid.2.5.4.65, registeredAddress, oid.2.5.4.26, rfc822Mailbox, oid.0.9.2342.19200300.100.1.3, roleOccupant, oid.2.5.4.33, searchGuide, oid.2.5.4.14, seeAlso, oid.2.5.4.34, supportedApplicationContext, oid.2.5.4.30, telephoneNumber, oid.2.5.4.20, telexNumber, oid.2.5.4.21, teletexTerminal-Identifier, oid.2.5.4.22, uniqueIdentifier, oid.2.5.4.45, uniqueMember, oid.2.5.4.50, userid, oid.0.9.2342.19200300.100.1.1, and x121Address, oid.2.5.4.24.

Note that a number of different and often incompatible naming schemes for X.500 attributes exist. X.500 only defined a handful of names, and as a result many other standards and implementations invented their own, a number of which conflict with each other, and several of which conflict with the original X.500 names. cryptlib uses the names that are most widely used with certificates. Since many of the names used by different standards conflict, it's not possible to have cryptlib handle multiple aliases for the same attribute, however if you require custom names to conform to a particular standard or interpretation of a standard, you can change the values in the code to reflect whatever names you want.

The CRYPT_CERTINFO_DN provides a powerful means of creating completely custom DNs, note though that this can result in DNs that can't be correctly processed or displayed by many applications, so you should only create non-standard DNs in this manner where it's absolutely necessary. In particular you need to take care that DN components like the CommonName and email address are in a form that cryptlib can work with, otherwise functions like **cryptGetPublicKey** that use DN components for lookups make not be able to locate the certificate.

Extended Certificate Identification Information

In the early to mid 1990's when it became clear that the Internet was going to be the driving force behind certificate technology, X.509 was amended to allow a more general-purpose type of identification than the complex and specialised DN. This new form was called the GeneralName, since it provided far more flexibility than the original DN. A GeneralName can contain an email address, a URL, an IP address, an alternative DN that doesn't follow the strict rules for the main certificate DN (it could for example contain a postal or street address), less useful components like X.400 and EDI addressing information, and even user-defined information that might be used in a certificate, for example medical patient, taxpayer, or social security ID's.

As with DNs, cryptlib goes to some lengths to hide the complexity involved in handling GeneralNames (recall the previous technical description of a DN, and then consider that this constitutes only a small portion of the entire GeneralName). Like a DN, a GeneralName can contain a number of components. Unless otherwise noted, the components are all text strings.

Component	Description
DirectoryName	A DN that can contain supplementary information that doesn't fit easily into the main certificate DN. You can specify this value with CRYPT_-CERTINFO_DIRECTORYNAME.
DNSName	An Internet host's fully-qualified domain name. You can specify this value with CRYPT_-CERTINFO_DNSNAME.

Component	Description
EDIPartyName.Name-Assigner	An EDI assigner-and-value pair with the EDI name assigner specified by CRYPT_CERTINFO_-EDIPARTYNAME_NAMEASSIGNER and the party name specified by CRYPT_CERTINFO_-EDIPARTYNAME_PARTYNAME.
EDIPartyName.Party-Name	
IPAddress	An IP address as per RFC 791, containing a 4-byte binary string in network byte order. You can specify this value with CRYPT_CERTINFO_-IPADDRESS.
OtherName.TypeID	A user-defined type-and-value pair with the type specified by CRYPT_CERTINFO_-OTHERNAME_TYPEID and the value specified by CRYPT_CERTINFO_OTHERNAME_VALUE. The type is an ISO object identifier and the corresponding value is a binary string that can contain anything, identified by the object identifier (if you know what this is then you should also know how to obtain one).
OtherName.Value	
RegisteredID	An object identifier (again, if you know what this is then you should know how to obtain one). You can specify this value with CRYPT_CERTINFO_-REGISTEREDID.
RFC822Name	An email address. You can specify this value with CRYPT_CERTINFO_RFC822NAME. For compatibility with the older (obsolete) PKCS #9 emailAddress attribute, cryptlib will also accept CRYPT_CERTINFO_EMAIL to specify this field.
UniformResource-Identifier	A URL for either FTP, HTTP, or LDAP access as per RFC 1738. You can specify this value with CRYPT_CERTINFO_-UNIFORMRESOURCEIDENTIFIER.

Of the above GeneralName components, the most useful ones are the RFC822Name (to specify an email address), the DNSName (to specify a server address), and the UniformResourceIdentifier (to specify a web page or FTP server). Somewhat less useful is the DirectoryName, which can specify additional information that doesn't fit easily into the main certificate DN. The other components should be avoided unless you have a good reason to require them (that is, don't use them just because they're there).

Working with GeneralName Components

Now that the details of GeneralNames have been covered, you can use them to add additional identification information to certificate requests and certificates. For example to add an email address and home page URL to the certification request shown earlier you would use:

```
CRYPT_CERTIFICATE cryptCertRequest;

/* Create certification request and add other components */
/* ... */

/* Add identification information */
/* ... */

/* Add additional identification information */
cryptSetAttributeString( cryptCertRequest, CRYPT_CERTINFO_RFC822NAME,
    "paul@cci.com", 12 );
cryptSetAttributeString( cryptCertRequest,
    CRYPT_CERTINFO_UNIFORMRESOURCEIDENTIFIER,
    "http://www.cci.com/~paul", 23 );
```

```
/* Sign certification request and transmit to CA */  
/* ... */
```

Although GeneralNames are commonly used to identify a certificate's owner just like a DN, they are in fact a certificate extension rather than a basic attribute. Each certificate can contain multiple extensions that contain GeneralNames. The various extensions that can contain GeneralNames are covered in "Certificate Extensions" on page 320, and the details of working with them are explained in "Composite Extension Attributes" on page 321.

Certificate Fingerprints

Certificates are sometimes identified through "fingerprints" that constitute either an MD5 or SHA-1 hash of the certificate data (the most common form is an MD5 hash). You can obtain a certificate's fingerprint by reading its CRYPT_CERTINFO_-FINGERPRINT attribute, which yields the default (MD5) fingerprint for the certificate. You can also explicitly query a particular fingerprint type with CRYPT_CERTINFO_-FINGERPRINT_MD5 and CRYPT_CERTINFO_-FINGERPRINT_SHA:

```
unsigned char fingerprint[ CRYPT_MAX_HASHSIZE ]  
int fingerprintSize;  
  
cryptGetAttributeString( certificate, CRYPT_CERTINFO_FINGERPRINT,  
    &fingerprint, &fingerprintSize );
```

This will return the certificate fingerprint.

Importing/Exporting Certificates

If you have an encoded certificate that was obtained elsewhere, you can import it into a certificate object using **cryptImportCert**. There are more than a dozen mostly incompatible formats for communicating certificates, of which cryptlib will handle all the generally useful and known ones. This includes straight binary certification requests, certificates, attribute certificates, and CRLs (usually stored with a .der file extension when they are saved to disk), PKCS #7 certificate chains, and Netscape certificate sequences. Certificates can also be protected with base64 armouring and BEGIN/END CERTIFICATE delimiters, which is the format used by some web browsers and other applications. When transferred via HTTP using the Netscape-specific format, certificates, certificate chains, and Netscape certificate sequences are identified with have the MIME content types application/x-x509-user-cert, application/x-x509-ca-cert, and application/x-x509-email-cert, depending on the certificate type (cryptlib doesn't use the MIME content type since the certificate itself provides a far more reliable indication of its intended use than the easily-altered MIME content type). Finally, certification requests and certificate chains can be encoded with the MIME / S/MIME content types application/pkcs-signed-data, application/x-pkcs-signed-data, application/pkcs-certs-only, application/x-pkcs-certs-only, application/pkcs10, or application/x-pkcs10. These are usually stored with a .p7c extension (for pure certificate chains), a .p7s extension (for signatures containing a certificate chain), or a .p10 extension (for certification requests) when they are saved to disk.

cryptlib will import any of the previously described certificate formats if they are encoded in this manner. To import a certificate object you would use:

```
CRYPT_CERTIFICATE cryptCertificate;  
  
/* Import the certificate object from the encoded certificate */  
cryptImportCert( cert, certLength, cryptUser, &cryptCertificate );
```

Note that the CRYPT_CERTIFICATE is passed to **cryptImportCert** by reference, as the function modifies it when it creates the certificate object.

Some certificate objects may contain unrecognised critical extensions (certificate extensions are covered in "Certificate Extensions" on page 320) which require that

the certificate be rejected by cryptlib. If a certificate contains an unrecognised critical extension, cryptlib will return a `CRYPT_ERROR_PERMISSION` error to indicate that you have no permission to use this object.

All the parameters and information needed to create the certificate object are a part of the certificate, and **cryptImportCert** takes care of initialising the certificate object and setting up the attributes and information inside it. The act of importing a certificate simply decodes the information and initialises a certificate object, it doesn't check the signature on the certificate. To check the certificate's signature you need to use **cryptCheckCert**, which is explained in "Signing/Verifying Certificates" on page 308.

There may be instances in which you're not exactly certain of the type of certificate object you have imported (for example importing a file with a `.der` or `.cer` extension could create a certificate request, a certificate, an attribute certificate, or a certificate chain object depending on the file contents). In order to determine the exact type of the object, you can read its `CRYPT_CERTINFO_CERTTYPE` attribute:

```
CRYPT_CERTTYPE_TYPE certType;

cryptGetAttribute( certificate, CRYPT_CERTINFO_CERTTYPE, &certType );
```

This will return the type of the imported object.

You can export a signed certificate from a certificate object using **cryptExportCert**:

```
CRYPT_CERTIFICATE cryptCertificate;
void *certificate;
int certLength

/* Allocate memory for the encoded certificate */
certificate = malloc( certMaxLength );

/* Export the encoded certificate from the certificate object */
cryptExportCert( certificate, certMaxLength, &certLength,
certFormatType, cryptCertificate );
```

cryptlib will export certificates in any of the formats in which it can import them. The available `certFormat` types are:

Format Type	Description
<code>CRYPT_CERTFORMAT_-CERTCHAIN</code>	A certificate encoded as a PKCS #7 certificate chain.
<code>CRYPT_CERTFORMAT_-CERTIFICATE</code>	A certification request, certificate, or CRL in binary data format. The certificate object is encoded according to the ASN.1 distinguished encoding rules. This is the normal certificate encoding format.
<code>CRYPT_CERTFORMAT_-TEXT_-CERTCHAIN</code>	As <code>CRYPT_CERTFORMAT_-CERTCHAIN</code> but with base64 armouring of the binary data.
<code>CRYPT_CERTFORMAT_-TEXT_-CERTIFICATE</code>	As <code>CRYPT_CERTFORMAT_-CERTIFICATE</code> but with base64 armouring of the binary data.

If the object that you're exporting is a complete certificate chain rather than an individual certificate then these options work somewhat differently. The details of exporting certificate chains are covered in "Exporting Certificate Chains" on page 313.

The resulting encoded certificate is placed in the memory buffer pointed to by `certificate` of maximum size `certificateMaxLength`, and the actual length is stored in `certLength`. This leads to a small problem: How do you know how big to make the buffer? The answer is to use **cryptExportCert** to tell you. If you pass in a null pointer for `certificate`, the function will set `certLength` to

the size of the resulting encoded certificate, but not do anything else. You can then use code like:

```
cryptExportCert( NULL, 0, &certMaxLength, certFormatType,
    cryptCertificate );
certificate = malloc( certMaxLength );
cryptExportCert( certificate, certMaxLength, &certLength,
    certFormatType, cryptCertificate );
```

to create the encoded certificate.

Alternatively, you can just reserve a reasonably sized block of memory and use that to hold the encoded certificate. “Reasonably sized” means a few Kb, a 4K block is plenty (a certificate for a 1024-bit key without certificate extensions is typically about 700 bytes long if encoded using any of the binary formats, or 900 bytes long if encoded using any of the text formats).

If the certificate is one that you’ve created yourself rather than importing it from an external source, you need to add various data items to the certificate and then sign it before you can export it. If you try to export an incompletely prepared certificate such as a certificate in which some required fields haven’t been filled in or one that hasn’t been signed, **cryptExportCert** will return the error **CRYPT_ERROR_NOTINITED** to tell you that the certificate information hasn’t been completely set up.

Signing/Verifying Certificates

Once a certificate object contains all the information you want to add to it, you need to sign it in order to transform it into its final state in which the data in it can be written to a keyset (if the object’s final state is a key certificate or CA certificate) or exported from the object. Before you sign the certificate, the information within it exists only in a very generic and indeterminate state. After signing it, the information is turned into a fixed certificate, CA certificate, certification request, or CRL, and no further changes can be made to it.

You can sign the information in a certificate object with **cryptSignCert**:

```
CRYPT_CONTEXT privKeyContext;

/* Sign the certificate object */
cryptSignCert( cryptCertificate, privKeyContext );
```

There are some restrictions on the types of keys that can be used to sign certificate objects. These restrictions are imposed by the way in which certificates and certificate-related items are encoded, and are as follows:

Certificate Type	Can be Signed By
Attribute certificate	Private key associated with an authority certificate.
Certificate	Private key associated with a CA certificate. This can also be a self-signed (non-CA) certificate, but some software will then decide that the resulting certificate is a CA certificate even though it isn’t.
CA certificate	Private key associated with a CA certificate (when one CA certifies another) or the private key from which the certificate being signed was created (when the CA certifies itself).
Certification request	Private key associated with the certification request.
Certificate chain	Private key associated with a CA certificate.
CRL	Private key associated with the CA certificate that was used

Certificate Type	Can be Signed By
Attribute certificate	Private key associated with an authority certificate.
Certificate	Private key associated with a CA certificate. This can also be a self-signed (non-CA) certificate, but some software will then decide that the resulting certificate is a CA certificate even though it isn't.
OCSP request/response	Private key associated with a certificate and authorised or trusted to sign requests/responses.

to issue the certificates that are being revoked.

In order to sign any type of certificate object other than a self-signed one, you must use a private key belonging to a CA. This means that the certificate associated with the signing key must have its `CRYPT_CERTINFO_CA` attribute set to true (a nonzero value) and must have a key usage value that indicates that it's valid for signing certificates (or CRLs if the object being signed is a CRL). If you try to sign an object other than a self-signed certificate or cert request with a non-CA key, cryptlib will return an error status indicating the nature of the problem. If the status is `CRYPT_ERROR_PARAM2`, the private key you're using doesn't have a certificate associated with it (that is, you're trying to sign the certificate with a raw private key without an associated CA certificate). If the status is `CRYPT_ERROR_INVALID`, the key you're using doesn't have the ability to sign certificates, for example because it isn't a CA key or because it doesn't contain a key usage value that indicates that it's valid for signing certificates or CRLs. In the latter case you can read the `CRYPT_ATTRIBUTE_ERROR_TYPE` and `CRYPT_ATTRIBUTE_ERROR_LOCUS` attributes to get more information about the nature of the problem as described in "Error Handling" on page 367.

Some certificate objects (for example OCSP requests and responses) can have signing certificate information included with the object, although by default only the signature itself is included. You can specify the amount of information which is included using the `CRYPT_CERTINFO_SIGNATURELEVEL` attribute. Setting this to `CRYPT_SIGNATURELEVEL_NONE` (the default) includes only the signature, setting it to `CRYPT_SIGNATURELEVEL_SIGNERCERT` includes the immediate signing certificate, and setting it to `CRYPT_SIGNATURELEVEL_ALL` includes all relevant information, for example the complete certificate chain. You should always use the default signing level unless you specifically know that you need to provide extra information such as signing certificates or a certificate chain.

Once a certificate item has been signed, it can no longer be modified or updated using the usual certificate manipulation functions, and any attempt to update information in it will return `CRYPT_ERROR_PERMISSION` to indicate that you have no permission to modify the object. If you want to add or delete data to or from the certificate item, you have to start again with a new certificate object. You can determine whether a certificate item has been signed and can therefore no longer be changed by reading its `CRYPT_CERTINFO_IMMUTABLE` attribute:

```
int isImmutable;

cryptGetAttribute( certificate, CRYPT_CERTINFO_IMMUTABLE,
                  &isImmutable );
```

If the result is set to true (a nonzero value), the certificate item can no longer be changed.

If you're creating a self-signed certificate signed by a raw private key with no certificate information associated with it, you need to set the `CRYPT_CERTINFO_SELFSIGNED` attribute before you sign it otherwise cryptlib will flag the attempt to sign using a non-certificate key as an error. Non-certificate private keys can only be used to create self-signed certificates (if `CRYPT_CERTINFO_SELFSIGNED` is set) or certification requests.

If the object being signed contains unrecognised extensions, cryptlib will not include them in the signed object (signing extensions of unknown significance is a risky practice for a CA, which in some jurisdictions can be held liable for any arising problems). If you want to be able to sign unrecognised extensions, you can enable this with the cryptlib configuration option `CRYPT_OPTION_CERT_SIGNUNRECOGNISEDATTRIBUTES` as explained in “Working with Configuration Options” on page 359.

You can verify the signature on a certificate object using **cryptCheckCert** and the public key or certificate corresponding to the private key that was used to sign the certificate (you can also pass in a private key if you want, **cryptCheckCert** will only use the public key components, although you shouldn’t really be in possession of someone else’s private key). To perform the check using a public key context you’d use:

```
CRYPT_CONTEXT pubKeyContext;

/* Check the signature on the certificate object information using the
   public key */
cryptCheckCert( cryptCertificate, pubKeyContext );
```

A signature check using a certificate is similar, except that it uses a certificate object rather than a public key context.

If the certificate object is self-signed, you can pass in `CRYPT_UNUSED` as the second parameter and **cryptCheckCert** will use the key contained in the certificate object to check its validity. You can determine whether a certificate object is self-signed by reading its `CRYPT_CERTINFO_SELFSIGNED` attribute. Certification requests are always self-signed, and certificate chains count as self-signed if they contain a self-signed top-level certificate that can be used to recursively check the rest of the chain. If the certificate object is a CA certificate which is signing itself (in other words if it’s a self-signed certificate), you can also pass the certificate as the second parameter in place of `CRYPT_UNUSED`, this has the same effect since the certificate is both the signed and signing object.

If the certificate is invalid (for example because it has expired or because some certificate usage constraint hasn’t been met), cryptlib will return `CRYPT_ERROR_INVALID` to indicate that the certificate isn’t valid. This value is returned regardless of whether the signature check succeeds or fails. You can find out the exact nature of the problem by reading the extended error attributes as explained in “Error Handling” on page 367.

If the signing/signature check key is stored in an encryption context with a certificate associated with it or in a certificate, there may be constraints on the key usage that are imposed by the certificate. If the key can’t be used for the signature or signature check operation, the function will return `CRYPT_ERROR_INVALID` to indicate that the key isn’t valid for this operation. You can find out more about the exact nature of the problem by reading the extended error attributes as explained in “Error Handling” on page 367.

If you’re acting as a CA and issuing significant numbers of certificates then a much easier alternative to signing each certificate yourself using **cryptSignCert** is to use cryptlib’s certificate management capabilities as described in “Managing a Certification Authority” on page 256.

Certificate Chains

Because of the lack of availability of a general-purpose certificate directory, many security protocols (most notable S/MIME and SSL) transmit not individual certificates but entire certificate chains that contain a complete certificate path from the end user’s certificate up to some widely-trusted CA certificate (referred to as a root CA certificate if it’s a self-signed CA certificate) whose trust will be handled for you by cryptlib’s trust manager. cryptlib supports the creation, import, export, and checking of certificate chains as `CRYPT_CERTTYPE_CERTCHAIN` objects, with

individual certificates in the chain being accessed as if they were standard certificates contained in a `CRYPT_CERTTYPE_CERTIFICATE` object.

Working with Certificate Chains

Individual certificates in a chain are addressed through a certificate cursor that functions in the same way as the attribute cursor discussed in “Attribute Lists and Attribute Groups” on page 38. Although a certificate chain object appears as a single object, it consists internally of a collection of certificates of which the first in the chain is the end user’s certificate and the last is a root CA certificate or at least an implicitly trusted CA certificate.

You can move the certificate cursor using the `CRYPT_CERTINFO_CURRENT_CERTIFICATE` attribute and the standard cursor movement codes. For example to move the cursor to the first (end-user) certificate in the chain, you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_CURRENT_CERTIFICATE,
                  CRYPT_CURSOR_FIRST );
```

To advance the cursor to the next certificate, you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_CURRENT_CERTIFICATE,
                  CRYPT_CURSOR_NEXT );
```

The certificate cursor and the extension/extension attribute cursor are two completely independent objects, so moving the certificate cursor from one certificate to another doesn’t affect the extension cursor setting for each certificate. If you select a particular extension in a certificate, then move to a different certificate and select an extension in that, and then move back to the first certificate, the original extension will still be selected.

Once you’ve selected a particular certificate in the chain, you can work with it as if it were the only certificate contained in the certificate object. The initially selected certificate is the end user’s certificate at the start of the chain. For example to read the `commonName` from the subject name for the end user’s certificate and for the next certificate in the chain you would use:

```
char commonName[ CRYPT_MAX_TEXTSIZE + 1 ];
int commonNameLength;

/* Retrieve the commonName from the end user's certificate */
cryptGetAttributeString( cryptCertChain, CRYPT_CERTINFO_COMMONNAME,
                        commonName, &commonNameLength );
commonName[ commonNameLength ] = '\0';

/* Move to the next certificate in the chain */
cryptSetAttribute( cryptCertChain, CRYPT_CERTINFO_CURRENT_CERTIFICATE,
                  CRYPT_CURSOR_NEXT );

/* Retrieve the commonName from the next certificate */
cryptGetAttributeString( cryptCertChain, CRYPT_CERTINFO_COMMONNAME,
                        commonName, &commonNameLength );
commonName[ commonNameLength ] = '\0';
```

Apart from this, certificate chains work just like certificates — you can import them, export them, verify the signatures on them (which verifies the entire chain of certificates until a trusted certificate is reached), and write them to and read them from a keyset in exactly the same manner as an individual certificate.

Signing Certificate Chains

When you sign a single subject certificate using **cryptSignCert**, a small amount of information is copied from the signing certificate (the issuer cert) to the subject certificate as part of the signing process, and the result is a single, signed subject certificate. In contrast signing a single subject certificate contained in a certificate chain object results in the signing certificates (either a single issuer certificate or an entire chain of certificates) being copied over to the certificate chain object so that the signed certificate ends up as part of a complete chain. The exact details are as follows:

Object to sign	Signing object	Result
Certificate	Certificate	Certificate
Certificate	Certificate chain	Certificate
Certificate chain	Certificate	Certificate chain, length = 2
Certificate chain	Certificate chain	Certificate chain, length = length of signing chain + 1

For example the following code produces a single signed certificate:

```
CRYPT_CERTIFICATE cryptCertificate;

/* Build a certificate from a cert request */
cryptCreateCert( &cryptCertificate, cryptUser,
    CRYPT_CERTTYPE_CERTIFICATE );
cryptSetAttribute( cryptCertificate, CRYPT_CERTINFO_CERTREQUEST,
    cryptCertRequest );

/* Read a private key with cert chain from a private key keyset */
/* ... */

/* Sign the certificate */
cryptSignCert( cryptCertificate, caPrivateKey );
```

In contrast the following code produces a complete certificate chain, since the object being created is a `CRYPT_CERTTYPE_CERTCHAIN` (which can hold a complete chain) rather than a `CRYPT_CERTTYPE_CERTIFICATE` (which only holds a single certificate):

```
CRYPT_CERTIFICATE cryptCertChain;

/* Build a certificate from a cert request */
cryptCreateCert( &cryptCertChain, cryptUser,
    CRYPT_CERTTYPE_CERTCHAIN );
cryptSetAttribute( cryptCertChain, CRYPT_CERTINFO_CERTREQUEST,
    cryptCertRequest );

/* Read a private key with cert chain from a private key keyset */
/* ... */

/* Sign the certificate chain */
cryptSignCert( cryptCertChain, caPrivateKey );
```

By specifying the object type to be signed, you can choose between creating a single signed certificate or a complete certificate chain.

Checking Certificate Chains

When verifying a certificate chain with **cryptCheckCert**, you don't have to supply an issuer certificate since the chain should contain all the issuer certificates up to one which is trusted by cryptlib:

```
CRYPT_CERTIFICATE cryptCertChain;

/* Verify an entire cert chain */
cryptCheckCert( cryptCertChain, CRYPT_UNUSED );
```

As with self-signed certificates, you can also pass in the certificate chain as the signing certificate instead of using `CRYPT_UNUSED`, this has the same effect since the certificate chain is both the signed and signing object.

If a certificate in the chain is invalid or the chain doesn't contain a trusted certificate at some point in the chain, cryptlib will return an appropriate error code and leave the invalid certificate as the currently selected one, allowing you to obtain information about the nature of the problem by reading the extended error attributes as explained in "Error Handling" on page 367.

If the error encountered is the fact that the chain doesn't contain a trusted certificate somewhere along the line, cryptlib will either mark the top-level certificate as having a missing `CRYPT_CERTINFO_TRUSTED_IMPLICIT` attribute if it's a CA root

certificate (that is, there's a root certificate present but it isn't trusted) or mark the chain as a whole as having a missing certificate if there's no CA root certificate present and no trusted certificate present either. Certificate trust management is explained in more detail in "Certificate Trust Management" on page 317.

Certificate chain validation is an extremely complex process that takes into account an enormous amount of validation information that may be spread across an entire certificate chain. For example in a chain of 10 certificates, the 3rd certificate from the root may place a constraint that doesn't take effect until the 7th certificate from the root is reached. Because of this, a reported validation problem isn't necessarily related to a given certificate and its immediate issuing certificate, but may have been caused by a different certificate a number of steps further along the chain.

Some certificate chains contain CA certificates that specify certificate policies. By default cryptlib requires that a policy that's set by a CA is matched by the certificates that the CA issues (in other words the CA sets policies for certificates further down the chain). If you want to allow policies to change going down the chain once the CA has set them, you can set the CRYPT_OPTION_CERT_REQUIREPOLICY option to false (0). When it's set to this value cryptlib won't verify that policies match up as it goes down the chain. You wouldn't normally need to use this configuration option, it's used to provide an optional capability that's covered in some certificate standards documents.

Some certificate chains may not contain or be signed by a trusted CA certificate, but may end in a root CA certificate with an unknown trust level. Since the cryptlib trust manager can't provide any information about this certificate, it won't be possible to verify the chain. If you want to trust the root CA certificate you can use the cryptlib trust management mechanisms to handle this, as explained in "Certificate Trust Management" on page 317.

Exporting Certificate Chains

As is the case when signing certificates and certificate chains, cryptlib gives you a high degree of control over what part of the chain you want to export. By specifying an export format of CRYPT_CERTFORMAT_CERTIFICATE or CRYPT_CERTFORMAT_CERTCHAIN, you can control whether a single certificate or an entire chain is exported. The exact details are as follows:

Object type	Export format	Result
Certificate	Certificate	Certificate
Certificate	Certificate chain	Certificate chain, length = 1
Certificate chain	Certificate	Currently selected certificate in the chain
Certificate chain	Certificate chain	Certificate chain

For example the following code exports the currently selected certificate in the chain as a single certificate:

```
CRYPT_CERTIFICATE cryptCertChain;
void *certificate;
int certificateLength;

/* Allocate memory for the encoded certificate */
certificate = malloc( certificateMaxLength );

/* Export the currently selected certificate from the certificate
chain */
cryptExportCert( certificate, certificateMaxLength,
&certificateLength, CRYPT_CERTFORMAT_CERTIFICATE, cryptCertChain );
```

In contrast the following code exports the entire certificate chain:

```
CRYPT_CERTIFICATE cryptCertChain;
void *certChain;
int certChainLength;

/* Allocate memory for the encoded certificate chain */
certChain = malloc( certChainMaxLength );

/* Export the entire certificate chain */
cryptExportCert( certChain, certChainMaxLength, &certChainLength,
    CRYPT_CERTFORMAT_CERTCHAIN, cryptCertChain );
```

Certificate Revocation using CRLs

Once a certificate has been issued, you may need to revoke it before its expiry date if the private key it corresponds to is lost or stolen, or if the details given in the certificate (for example your job role or company affiliation) change. Certificate revocation is done through a certificate revocation list (CRL) that contains references to one or more certificates that have been revoked by a CA. cryptlib supports the creation, import, export, and checking of CRLs as CRYPT_CERTTYPE_CRL objects, with individual revocation entries accessed as if they were standard certificate components. Note that these entries are merely references to revoked certificates and not the certificates themselves, so all they contain is a certificate reference, the date of revocation, and possibly various optional extras such as the reason for the revocation.

Working with CRLs

Individual revocation entries in a CRL are addressed through a certificate cursor that functions in the same way as the attribute cursor discussed in “Attribute Lists and Attribute Groups” on page 38. Although a CRL appears as a single object, it consists internally of a collection of certificate revocation entries that you can move through using the standard cursor movement codes. For example to move the cursor to the first entry in the CRL, you would use:

```
cryptSetAttribute( cryptCRL, CRYPT_CERTINFO_CURRENT_CERTIFICATE,
    CRYPT_CURSOR_FIRST );
```

To advance the cursor to the next entry, you would use:

```
cryptSetAttribute( cryptCRL, CRYPT_CERTINFO_CURRENT_CERTIFICATE,
    CRYPT_CURSOR_NEXT );
```

Since each revocation entry can have its own attributes, moving the entry cursor from one entry to another can change the attributes that are visible. This means that if you’re working with a particular entry, the attributes for that entry will be visible, but attributes for other entries won’t be. To complicate this further, CRLs can also contain global attributes that apply to, and are visible for, all entries in the CRL. cryptlib will automatically handle these for you, allowing access to all attributes (both per-entry and global) that apply to the currently selected revocation entry.

Creating CRLs

To create a CRL, you first create the CRL certificate object as usual and then push one or more certificates to be revoked into it.

```
CRYPT_CERTIFICATE cryptCRL;

/* Create the (empty) CRL */
cryptCreateCert( &cryptCRL, cryptUser, CRYPT_CERTTYPE_CRL );

/* Add the certificates to be revoked */
cryptSetAttribute( cryptCRL, CRYPT_CERTINFO_CERTIFICATE,
    revokedCert1 );
cryptSetAttribute( cryptCRL, CRYPT_CERTINFO_CERTIFICATE,
    revokedCert2 );
/* ... */
cryptSetAttribute( cryptCRL, CRYPT_CERTINFO_CERTIFICATE,
    revokedCertN );

/* Sign the CRL */
cryptSignCertificate( cryptCRL, caPrivateKey );
```


As has already been mentioned, you must be a CA in order to issue a CRL, and you can only revoke certificates that you have issued using the certificate used to sign the CRL (you can't, for example, revoke a certificate issued by another CA, or revoke a certificate issued with one CA certificate using a different CA certificate). If you try to add certificates issued by multiple CAs to a CRL, or try to sign a CRL with a CA certificate that differs from the one that signed the certificates in the CRL, cryptlib will return a `CRYPT_ERROR_INVALID` error to indicate that the certificate you are trying to add to the CRL or sign the CRL with is from the wrong CA. To reiterate: Every certificate in a given CRL must have been issued using the CA certificate which is used to sign the CRL. If your CA uses multiple certificates (for example a Class 1 certificate, a Class 2 certificate, and a Class 3 certificate) then it must issue one CRL for each certificate class. cryptlib will perform the necessary checking for you to ensure you don't issue an invalid CRL.

If you're acting as a CA and issuing CRLs for certificates then a much easier way to handle this is to use cryptlib's certificate management capabilities as described in "Issuing a CRL" on page 266, since this takes care of all of these details for you.

Advanced CRL Creation

The code shown above creates a relatively straightforward, simple CRL with no extra information included with the revocation. You can also include extra attributes such as the time of the revocation (which may differ from the time the CRL was issued, if you don't specify a time then cryptlib will use the CRL issuing time), the reason for the revocation, and the various other CRL-specific information as described in "CRL Extensions" on page 329.

If you set a revocation time with no revoked certificates present in the CRL, cryptlib will use this time for any certificates you add to the CRL for which you don't explicitly set the revocation time so you can use this to set a default revocation time for any certificates you add. If you set a revocation time and there are revoked certificates present in the CRL, cryptlib will set the time for the currently selected certificate, which will be either the last one added or the one selected with the certificate cursor commands.

For example to revoke a list of certificates, setting the revocation date for each one individually, you would use:

```
CRYPT_CERTIFICATE cryptCRL;

while( moreCerts )
{
    CRYPT_CERTIFICATE revokedCert;
    time_t revocationTime;

    /* Get the certificate to revoke and its revocation time */
    revokedCert = ...;
    revocationTime = ...;

    /* Add them to the CRL */
    cryptSetAttribute( cryptCRL, CRYPT_CERTINFO_CERTIFICATE,
        revokedCert );
    cryptSetAttributeString( cryptCRL, CRYPT_CERTINFO_REVOCATIONDATE,
        &revocationTime, sizeof( time_t ) );

    /* Clean up */
    cryptDestroyCert( revokedCert );
}
```

You can also add additional attributes such as the reason for the revocation to each revoked certificate, a number of standards recommend that a reason is given for each revocation. The revocation codes are specified in "CRL Extensions" on page 329.

CRLs can be signed, verified, imported, and exported just like other certificate objects.

Checking Certificates against CRLs

Verifying a certificate against a CRL with **cryptCheckCert** works just like a standard certificate check, with the second parameter being the CRL that the certificate is being checked against:

```
CRYPT_CERTIFICATE cryptCRL;

/* Check the certificate against the CRL */
cryptCheckCert( cryptCertificate, cryptCRL );
```

If the certificate has been revoked, cryptlib will return **CRYPT_ERROR_INVALID**. If the certificate has not been revoked (in other words if it is not on the CRL), cryptlib will return **CRYPT_OK**. Note that the only thing a CRL can say with certainty is “revoked”, so it can’t provide a true validity check for a certificate. For example, if you perform a CRL check on an Excel spreadsheet, a CRL will report it as being a valid certificate, since it’s not listed in the CRL. Similarly, a forged certificate can’t be handled by a CRL since it can’t be handled through a blacklist mechanism such as a CRL. If you require a true certificate validity check, you need to use a alternative mechanism such as RTCS.

If the certificate is revoked, the certificate’s revocation entry in the CRL will be left as the selected one, allowing you to obtain further information on the revocation (for example the revocation date or reason):

```
time_t revocationTime;
int revocationReason;

status = cryptCheckCert( cryptCertificate, cryptCRL );
if( status == CRYPT_ERROR_INVALID )
{
    int revocationTimeLength;

    /* The certificate has been revoked, get the revocation time and
       reason */
    cryptGetAttributeString( cryptCRL, CRYPT_CERTINFO_REVOCATIONDATE,
        &revocationTime, &revocationTimeLength );
    cryptGetAttribute( cryptCRL, CRYPT_CERTINFO_CRLREASON,
        &revocationReason );
}
```

Note that the revocation reason is an optional CRL component, so this may not be present in the CRL. If the revocation reason isn’t present, cryptlib will return **CRYPT_ERROR_NOTFOUND**.

Automated CRL Checking

As you can see from the description of the revocation checking process above, it quickly becomes unmanageable as the number of CRLs and the size of each CRL increases, since what should be a simple certificate validation check now involves checking the certificate against any number of CRLs (CRLs are generally regarded as a rather unsatisfactory solution to the problem of certificate revocation, but we’re stuck with them for the foreseeable future).

In order to ease this complex and long-winded checking process, cryptlib provides the ability to automatically check a certificate against CRLs stored in a cryptlib database keyset. To do this you first need to write the CRL or CRLs to the keyset as if they were normal certificates, as explained in “Writing a Key to a Keyset” on page 231. cryptlib will take each complete CRL and record all of the individual revocations contained in it for later use.

Once you have a keyset containing revocation information, you can use it to check the validity of a certificate using **cryptCheckCert**, giving the keyset as the second parameter:

```
CRYPT_KEYSET cryptKeyset;

/* Check the certificate using the keyset */
cryptCheckCert( cryptCertificate, cryptKeyset );
```

As with the check against a CRL, cryptlib will return `CRYPT_ERROR_INVALID` if the certificate has been revoked.

This form of automated checking considerably simplifies the otherwise arbitrarily complex CRL checking process since cryptlib can handle the check with a simple keyset query rather than having to locate and search large numbers of CRLs.

Certificate Trust Management

In order to provide extended control over certificate usage, cryptlib allows you to both further restrict the usage given in the certificate's `CRYPT_CERTINFO_-KEYUSAGE` attribute and to specify whether a given certificate should be implicitly trusted, avoiding the requirement to process a (potentially large) chain of certificates in order to determine the certificate's validity.

Controlling Certificate Usage

You can control the way a certificate can be used by setting its `CRYPT_CERTINFO_TRUSTED_USAGE` attribute, which provides extended control over the usage types that a certificate is trusted for. This attribute works by further restricting the usage specified by the `CRYPT_CERTINFO_KEYUSAGE` attribute, acting as a mask for the standard key usage so that a given usage is only permitted if it's allowed by both the key usage and trusted usage attributes. If the trusted usage attribute isn't present (which is the default setting) then all usage types specified in the key usage attribute are allowed.

For example assume a certificate's key usage attribute is set to `CRYPT_KEYUSAGE_DIGITALSIGNATURE` and `CRYPT_KEYUSAGE_KEYENCIPHERMENT`. By setting the trusted usage attribute to `CRYPT_KEYUSAGE_DIGITALSIGNATURE` only, you can tell cryptlib that you only trust the certificate to be used for signatures, even though the certificate's standard usage would also allow encryption. This means that you can control precisely how a certificate is used at a level beyond that provided by the certificate itself.

Implicitly Trusted Certificates

To handle certificate validation trust issues, cryptlib has a built-in trust manager that records whether a given CA's or end user's certificate is implicitly trusted. When cryptlib gets to a trusted certificate during the certificate validation process (for example as it's validating the certificates in a certificate chain), it knows that it doesn't have to go any further in trying to get to an ultimately trusted certificate. If you installed the default cryptlib certificates when you installed cryptlib itself then you'll have a collection of top-level certificates from the world's largest CAs already present and marked as trusted by cryptlib, so that if cryptlib is asked to process a certificate chain ending in one of these trusted CA certificates, the cryptlib trust manager will determine that the top-level certificate is implicitly trusted and use it to verify the lower-level certificates in the chain.

The trust manager provides a convenient mechanism for managing not only CA certificates but also any certificates that you decide you can trust implicitly. For example if you've obtained a certificate from a trusted source such as direct communication with the owner or from a trusted referrer, you can mark the certificate as trusted even if it doesn't have a full chain of CA certificates in tow. This is a natural certificate handling model in many situations (for example trading partners with an existing trust relationship), and avoids the complexity and expense of using an external CA to verify something that both parties know already. When scaled up to thousands of users (and certificates), this can provide a considerable savings both in terms of managing the certification process and in the cost of obtaining and renewing huge numbers of certificates each year.

Working with Trust Settings

You can get and set a certificate's trusted usage using `CRYPT_CERTINFO_TRUSTED_USAGE`, which takes as value the key usage(s) for which the certificate is trusted. To mark a certificate as trusted only for encryption, you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_TRUSTED_USAGE,  
                  CRYPT_KEYUSAGE_KEYENCIPHERMENT );
```

This setting will now be applied automatically to the certificate's usage permissions, so that even if its `CRYPT_CERTINFO_KEYUSAGE` attribute allowed signing and encryption, the `CRYPT_CERTINFO_TRUSTED_USAGE` attribute would restrict this to only allow encryption.

To remove any restrictions and allow all usages specified by `CRYPT_CERTINFO_KEYUSAGE`, delete the `CRYPT_CERTINFO_TRUSTED_USAGE` attribute, which allows the full range of usage types that are present in `CRYPT_CERTINFO_KEYUSAGE`:

```
cryptDeleteAttribute( cryptCertificate, CRYPT_CERTINFO_TRUSTED_USAGE  
                    );
```

You can get and set a certificate's implicitly trusted status using the `CRYPT_CERTINFO_TRUSTED_IMPLICIT` attribute, which takes as value a boolean flag that indicates whether the certificate is implicitly trusted or not. To mark a certificate as trusted, you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_TRUSTED_IMPLICIT, 1 );
```

Be careful when marking certificate chains (rather than individual certificates) as implicitly trusted. Since a chain usually contains multiple certificates, setting the `CRYPT_CERTINFO_TRUSTED_IMPLICIT` attribute affects the currently selected certificate in the chain. Typically you want to trust the root CA, while the certificate which is normally active when the chain is used is the end-user/leaf certificate. In order to select the root CA certificate, you should move the certificate cursor to it using the `CRYPT_CURSOR_LAST` movement code before marking the chain as trusted. This will explicitly make the top-level CA certificate trusted, rather than some arbitrary certificate in the chain.

To check whether a certificate is trusted you would use:

```
int isTrusted;  
  
cryptGetAttribute( certificate, CRYPT_CERTINFO_TRUSTED_IMPLICIT,  
                  &isTrusted );
```

Since the trust of a CA propagates down to the certificates it issues, the trust setting in this case applies to the whole chain rather than just one certificate in it. In other words if the chain is signed by a trusted CA, the entire chain beyond that point will be regarded as trusted.

If the result is set to true (a nonzero value) then the certificate is implicitly trusted by cryptlib. In practice you won't need to bother with this checking, since cryptlib will do it for you when it verifies certificate chains.

The certificate trust settings are part of cryptlib's configuration options, which are explained in more detail in "Working with Configuration Options" on page 359. Like all configuration options, changes to the trust settings only remain in effect during the current session with cryptlib unless you explicitly force them to be committed to permanent storage by resetting the configuration changed flag. For example if you change the trust settings for various certificates and want the new trust values to be applied when you use cryptlib in the future, you would use code like:

```
/* Mark various certificates as trusted and one as untrusted */  
cryptSetAttribute( certificate1, CRYPT_CERTINFO_TRUSTED_IMPLICIT, 1 );  
cryptSetAttribute( certificate2, CRYPT_CERTINFO_TRUSTED_IMPLICIT, 1 );  
cryptSetAttribute( certificate3, CRYPT_CERTINFO_TRUSTED_IMPLICIT, 1 );  
cryptSetAttribute( certificate4, CRYPT_CERTINFO_TRUSTED_IMPLICIT, 0 );
```

```
/* Save the new settings to permanent storage */  
cryptSetAttribute( CRYPT_UNUSED, CRYPT_OPTION_CONFIGCHANGED, FALSE );
```

Marking a certificate as untrusted doesn't mean that it can never be trusted, but merely that its actual trust status is currently unknown. If the untrusted certificate is signed by a trusted CA certificate (possibly several levels up a certificate chain) then the certificate will be regarded as trusted when cryptlib checks the certificate chain. In practice an untrusted certificate is really a certificate whose precise trust level has yet to be determined rather than a certificate which is explicitly not trusted. If you want to explicitly not trust a certificate for one or more types of usage, you can do this using the `CRYPT_CERTINFO_TRUSTED_USAGE` attribute.

Certificate Extensions

Certificate extensions form by far the most complicated portion of certificates. By default, cryptlib will add appropriate certificate extension attributes to certificates for you if you don't add any, but sometimes you may want to add or change these yourself. cryptlib supports extensions in two ways, through the usual add/get/delete attribute mechanism for extensions that it recognises, and through **cryptAddCertExtension**, **cryptGetCertExtension**, and **cryptDeleteCertExtension** for general extensions that it doesn't recognise. The general extension handling mechanism allows you to add, query, and delete any kind of extension to a certificate, including ones that you define yourself.

Extension Structure

X.509 version 3 introduced a mechanism by which additional information could be added to certificates through the use of certificate extensions. The X.509 standard defined a number of extensions, and over time other standards organisations defined their own additions and amendments to these extensions. In addition private organisations, businesses, and individuals have all defined their own extensions, some of which (for example the extensions from Netscape and Microsoft) have seen a reasonably wide amount of use. An extension contains three main pieces of information:

Field	Description
Type	The extension type, a unique identifier called an object identifier. This is given as a sequence of numbers that trace a path through an object identifier tree. For example the object identifier for the keyUsage extension is 2 5 29 15. The object identifier for cryptlib is 1 3 6 1 4 1 3029 32.
Critical Flag	<p>A flag that defines whether the extension is important enough that it must be processed by an application. If the critical flag is set and an application doesn't recognise the extension, it will reject the certificate.</p> <p>Since some standards (including X.509 itself) allow implementations to selectively ignore non-critical extensions, and support for extensions is often haphazard, it may be necessary to mark an extension as critical in order to ensure that other implementations process it. As usual, you should check to see whether your intended target correctly processes the extensions that you plan to use.</p>
Value	The extension data, corresponding to a cryptlib attribute group for more complex composite extensions, or a single cryptlib attribute for a few very simple extensions.

For the extensions that cryptlib recognises, the handling of the critical flag is automatic. For extensions that cryptlib doesn't handle itself, you need to set the critical flag yourself when you add the extension data using **cryptAddCertExtension**.

Working with Extension Attributes

Certificate extensions correspond to cryptlib attribute groups, with individual components of each certificate extension being represented by attributes within the group. Since this section applies specifically to certificates, the certificate-specific terminology referring to extensions rather than the general term attribute group will be used here.

cryptlib can identify attributes in extensions/attribute groups in one of three ways:

1. Through an extension identifier that denotes the entire extension/attribute group. For example CRYPT_CERTINFO_CERTPOLICIES denotes the certificatePolicies extension/attribute group.
2. Through an attribute identifier that denotes a particular attribute within an extension/attribute group. For example CRYPT_CERTINFO_CERTPOLICY denotes the policyIdentifier attribute contained within the certificatePolicies extension/attribute group.

Some extensions/groups only contain a single attribute, in which case the extension identifier is the same as the attribute identifier. For example the CRYPT_CERTINFO_KEYUSAGE extension contains a single attribute which is also identified by CRYPT_CERTINFO_KEYUSAGE.

3. Through the attribute cursor mechanism that allows you to step through a set of extensions extension by extension or attribute by attribute. Attribute cursor management is explained in more detail in “Attribute Lists and Attribute Groups” on page 38.

You can use the extension/group identifier to determine whether a particular extension is present with **cryptGetAttribute** (it will return CRYPT_ERROR_NOTFOUND if the extension isn't present), to delete an entire extension with **cryptDeleteAttribute**, and to position the extension cursor at a particular extension.

Attributes within extensions/group are handled in the usual manner, for example to retrieve the value of the basicConstraints CA attribute (which determines whether a certificate is a CA certificate) you would use:

```
int isCA;

cryptGetAttribute( certificate, CRYPT_CERTINFO_CA, &isCA );
```

To determine whether the entire basicConstraints extension is present, you would use:

```
int basicConstraintsPresent;

status = cryptGetAttribute( certificate,
    CRYPT_CERTINFO_BASICCONSTRAINTS, &basicConstraintsPresent );
if( cryptStatusOK( status ) )
    /* basicConstraints extension is present */;
```

You don't have to worry about the structure of individual extensions since cryptlib will handle this for you. For example to make a certificate a CA certificate, all that you need to do is:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_CA, 1 );
```

and cryptlib will construct the basicConstraints extension for you and set up the CA attribute as required. Because the basicConstraints extension is a fundamental X.509v3 extension, cryptlib will in fact always add this by default even if you don't explicitly specify it.

Composite Extension Attributes

Attributes that contain complete GeneralNames and/or DNs are composite attributes that have further items within them. These are handled in the standard way using the attribute cursor: You first move the cursor to the attribute that contains the GeneralName or DN that you want to work with and then get, set, or delete attributes within it:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_CURRENT_FIELD,
    CRYPT_CERTINFO_PERMITTEDSUBTREES );
cryptSetAttributeString( certificate, CRYPT_CERTINFO_RFC822NAME,
    rfc822Name, rfc822NameLength );
cryptSetAttributeString( certificate, CRYPT_CERTINFO_DNSNAME, dnsName,
    dnsNameLength );
```

This code first moves the cursor to the nameConstraints permittedSubtrees GeneralName and then sets the GeneralName attributes as usual. Since a GeneralName contains its own DN, moving the attribute cursor onto a GeneralName

means that any DN accesses will now refer to the DN in the GeneralName rather than the certificate subject or issuer name:

```
/* Select the permittedSubtrees GeneralName */
cryptSetAttribute( certificate, CRYPT_CERTINFO_CURRENT_FIELD,
    CRYPT_CERTINFO_PERMITTEDSUBTREES );

/* Set the DN components within the GeneralName */
cryptSetAttributeString( certificate, CRYPT_CERTINFO_COUNTRYNAME,
    countryName, countryNameLength );
cryptSetAttributeString( certificate, CRYPT_CERTINFO_LOCALITYNAME,
    localityName, localityNameLength );
```

This code first identifies the nameConstraints permittedSubtrees GeneralName as the one to be modified and then sets the DN components as usual. cryptlib uses this mechanism to access all DNs and GeneralNames, although this is usually hidden from you — when you modify a certificate object’s DN, cryptlib automatically uses the subject DN if you don’t explicitly specify it, and when you modify the GeneralName cryptlib uses the subject altName if you don’t explicitly specify it. In this way you can work with subject names and altNames without having to know about the DN and GeneralName selection mechanism.

Once you’ve selected a different GeneralName and/or DN, it remains selected until you select another one or move the attribute cursor off it, so if you wanted to move back to working with the subject name after performing the operations shown above you’d need to use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_SUBJECTNAME,
    CRYPT_UNUSED );
```

otherwise attempts to add, delete, or query further DN (or GeneralName) attributes will apply to the selected nameConstraints excludedSubtrees attribute instead of the subject name. Conversely, if you move the attribute cursor off the GeneralName that you’re working with, subsequent attempts to work with GeneralName or DN fields will fail with a CRYPT_ERROR_NOTFOUND, since there’s no GeneralName currently selected.

X.509 Extensions

X.509 version 3 and up, and assorted additional standards and revisions specify a large number of extensions, all of which are handled by cryptlib. In addition there are a number of proprietary and vendor-specific extensions that are also handled by cryptlib.

In the following descriptions only the generally useful attributes have been described. The full range of attributes is enormous, requires several hundred pages of standards specifications to describe them all, and will probably never be used in real life. These attributes are marked with “See certificate standards documents” to indicate that you should refer to other documents to obtain information about their usage (this is also a good indication that you shouldn’t really be using this attribute).

Alternative Names

The subject and issuer altNames are used to specify all the things that aren’t suitable for the main certificate DNs. The issuer altName is identified by CRYPT_-CERTINFO_ISSUERALTNAME and the subject altName is identified by CRYPT_-CERTINFO_SUBJECTALTNAME. Both consist of a single GeneralName whose use is explained in “Extended Certificate Identification Information” on page 304. This extension is valid in certificates, certification requests, and CRLs, and can contain one of each type of GeneralName component.

Basic Constraints

This is a standard extension identified by CRYPT_CERTINFO_-BASICCONSTRAINTS and is used to specify whether a certificate is a CA certificate or not. If you don’t set this extension, cryptlib will set it for you and mark

the certificate as a non-CA certificate. This extension is valid in certificates, attribute certificates, and certification requests, and has the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CA	Boolean
Whether the certificate is a CA certificate or not. When used with attribute certificates, the CA is called an authority, so cryptlib will also accept the alternative CRYPT_CERTINFO_AUTHORITY, which has the same meaning as CRYPT_CERTINFO_CA. If this attribute isn't set, the certificate is treated as a non-CA certificate.	
CRYPT_CERTINFO_PATHLENCONSTRAINT	Numeric
See certificate standards documents.	

For example to mark a certificate as a CA certificate you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_CA, 1 );
```

Certificate Policies, Policy Mappings, Policy Constraints, and Policy Inhibiting

The certificate policy extensions allow a CA to provide information on the policies governing a certificate, and to control the way in which a certificate can be used. For example it allows you to check that each certificate in a certificate chain was issued under a policy you feel comfortable with (certain security precautions taken, vetting of employees, physical security of the premises, and so on). The certificate policies attribute is identified by CRYPT_CERTINFO_CERTIFICATEPOLICIES and is valid in certificates.

The certificate policies attribute is a complex extension that allows for all sorts of qualifiers and additional modifiers. In general you should only use the policyIdentifier attribute in this extension, since the other attributes are difficult to support in user software and are ignored by many implementations:

Attribute/Description	Type
CRYPT_CERTINFO_CERTPOLICYID	String
The object identifier that identifies the policy under which this certificate was issued.	
CRYPT_CERTINFO_CERTPOLICY_CPSURI	String
The URL for the certificate practice statement (CPS) for this certificate policy.	
CRYPT_CERTINFO_CERTPOLICY_ORGANIZATION	String
CRYPT_CERTINFO_CERTPOLICY_NOTICENUMBERS	Numeric
CRYPT_CERTINFO_CERTPOLICY_EXPLICITTEXT	String
These attributes contain further qualifiers, modifiers, and text information that amend the certificate policy information. Refer to certificate standards documents for more information on these attributes.	

Since various CAs that would like to accept each other's certificates may have differing policies, there is an extension that allows a CA to map its policies to those of another CA. The policyMappings extension provides a means of mapping one policy to another (that is, for a CA to indicate that policy A, under which it is issuing a certificate, is equivalent to policy B, which is required by the certificate user). This extension is identified by CRYPT_CERTINFO_POLICYMAPPINGS and is valid in certificates:

Attribute/Description	Type
CRYPT_CERTINFO_ISSUERDOMAINPOLICY	String
The object identifier for the source (issuer) policy.	
CRYPT_CERTINFO_SUBJECTDOMAINPOLICY	String
The object identifier for the destination (subject) policy.	

A CA can also specify acceptable policy constraints for use in certificate chain validation. The policyConstraints extension is identified by CRYPT_CERTINFO_POLICYCONSTRAINTS and is valid in certificates:

Attribute/Description	Type
CRYPT_CERTINFO_REQUIREEXPLICITPOLICY See certificate standards documents.	Numeric
CRYPT_CERTINFO_INHIBITPOLICYMAPPING See certificate standards documents.	Numeric

Finally, a CA can inhibit the use of the special-case anyPolicy policy. The inhibitAnyPolicy extension is identified by CRYPT_CERTINFO_INHIBITANYPOLICY and is valid in certificates:

Attribute/Description	Type
CRYPT_CERTINFO_INHIBITANYPOLICY See certificate standards documents.	Numeric

CRL Distribution Points/Freshest CRL and Subject/Authority Information Access

These extensions specify how to obtain CRL information and information on the CA that issued a certificate. The cRLDistributionPoint extension is valid in certificates and is identified by CRYPT_CERTINFO_CRLDISTRIBUTIONPOINT:

Attribute/Description	Type
CRYPT_CERTINFO_CRLDIST_FULLNAME The location at which CRLs may be obtained. You should use the URL component of the GeneralName for this, avoiding the other possibilities.	GeneralName
CRYPT_CERTINFO_CRLDIST_REASONS	Numeric
CRYPT_CERTINFO_CRLDIST_CRLISSUER See certificate standards documents.	GeneralName

Note that the CRYPT_CERTINFO_CRLDIST_REASONS attribute has the same allowable set of values as the cRLReasons reasonCode, but in this case is given as a series of bit flags rather than the reasonCode numeric value (because X.509 says so, that's why). Because of this you must use CRYPT_CRLREASONFLAGS_name instead of CRYPT_CRLREASON_name when getting and setting these values.

If you plan to use this extension, you should be aware of the fact that it exists solely as a kludge created to work around problems involved in finding CRLs in X.500 directories, and thus presents a rather poor mechanism for distributing and obtaining revocation information. Unless it's absolutely imperative that you use this extension, it's better to use RTCS or OCSP as explained in "Certificate Status Checking using RTCS" on page 247, "RTCS Server Sessions" on page 207, "Certificate Revocation Checking using OCSP" on page 247, and "OCSP Server Sessions" on page 207.

The freshestCRL extension is valid in certificates and is identified by CRYPT_CERTINFO_FRESHESTCRL. The structure is identical to cRLDistributionPoint, with the subfields named with FRESHESTCRL instead of CRLDIST. As with cRLDistributionPoint, this is a kludge used to work with delta CRLs.

The subjectInfoAccess extension is valid in certificates and is identified by CRYPT_CERTINFO_SUBJECTINFOACCESS:

Attribute/Description	Type
CRYPT_CERTINFO_SUBJECTINFO_CAREPOSITORY The location at which the CA publishes certificates and CRLs, if the certificate is for a CA. You should use the URL component of the GeneralName for this, avoiding the other possibilities.	GeneralName
CRYPT_CERTINFO_SUBJECTINFO_TIMESTAMPING	GeneralName

The location at which timestamping services using the timestamp protocol (TSP) are available. You should use the URL component of the GeneralName for this, avoiding the other possibilities.

The authorityInfoAccess extension is valid in certificates and CRLs and is identified by CRYPT_CERTINFO_AUTHORITYINFOACCESS:

Attribute/Description	Type
CRYPT_CERTINFO_AUTHORITYINFO_CAISUERS The location at which information on CAs located above the CA that issued this certificate can be obtained. You should use the URL component of the GeneralName for this, avoiding the other possibilities.	GeneralName
CRYPT_CERTINFO_AUTHORITYINFO_CERTSTORE The location at which further certificates issued by the CAs that issued this certificate can be obtained. You should use the URL component of the GeneralName for this, avoiding the other possibilities.	GeneralName
CRYPT_CERTINFO_AUTHORITYINFO_CRLS The location at which further certificates issued by the CAs that issued this certificate can be obtained. You should use the URL component of the GeneralName for this, avoiding the other possibilities.	GeneralName
CRYPT_CERTINFO_AUTHORITYINFO_OCSP The location at which certificate revocation information can be obtained. You should use the URL component of the GeneralName for this, avoiding the other possibilities.	GeneralName
CRYPT_CERTINFO_AUTHORITYINFO_RTCS The location at which certificate validity information can be obtained. You should use the URL component of the GeneralName for this, avoiding the other possibilities.	GeneralName

Directory Attributes

This extension, identified by CRYPT_CERTINFO_SUBJECTDIRECTORY-ATTRIBUTES, allows additional X.500 directory attributes to be specified for a certificate. This extension is valid in certificates, and has the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_SUBJECTDIR_TYPE The object identifier that identifies the type of the directory attribute.	String
CRYPT_CERTINFO_SUBJECTDIR_VALUES The value of the directory attribute.	String

Key Usage, Extended Key Usage, and Netscape certificate type

These extensions specify the allowed usage for the key contained in this certificate. The keyUsage attribute is a standard extension identified by CRYPT_CERTINFO_KEYUSAGE and is used to specify general-purpose key usages such as key encryption, digital signatures, and certificate signing. If you don't set this attribute, cryptlib will set it for you to a value appropriate for the key type (for example a key for a signature-only algorithm such as DSA will be marked as a signature key).

The extKeyUsage attribute is identified by CRYPT_CERTINFO_EXTKEYUSAGE and is used to specify additional special-case usage such as code signing and SSL server authentication.

The Netscape certificate type attribute is a vendor-specific attribute identified by CRYPT_CERTINFO_NS_CERTTYPE and was used to specify certain types of web browser-specific certificate usage before the extKeyUsage attribute was fully specified. This attribute has now been superseded by extKeyUsage, but is still found in a number of certificates.

The keyUsage extension has a single numeric attribute with the same identifier as the extension itself (CRYPT_CERTINFO_KEYUSAGE). This extension is valid in certificates and certification requests, and contains a bit flag that can contain any of the following values:

Value	Description
CRYPT_KEYUSAGE_- DATAENCIPHERMENT	The key can be used for data encryption. This implies using public-key encryption for bulk data encryption, which is almost never done.
CRYPT_KEYUSAGE_- DIGITALSIGNATURE	The key can be used for digital signature generation and verification. This is the standard flag to set for digital signature use.
CRYPT_KEYUSAGE_- ENCIPHERONLY CRYPT_KEYUSAGE_- DECIPHERONLY	These flags modify the keyAgreement flag to allow the key to be used for only one part of the key agreement process.
CRYPT_KEYUSAGE_- KEYAGREEMENT	The key can be used for key agreement. This is the standard flag to set for key-agreement algorithms such as Diffie-Hellman.
CRYPT_KEYUSAGE_- KEYCERTSIGN CRYPT_KEYUSAGE_- CRLSIGN	The key can be used to sign certificates and CRLs. Using these flags requires the basicConstraint CA value to be set.
CRYPT_KEYUSAGE_- KEYENCIPHERMENT	The key can be used for key encryption/key transport. This is the standard flag to set for encryption use.
CRYPT_KEYUSAGE_- NONREPUDIATION	The key can be used for nonrepudiation purposes. Note that this use is usually different to CRYPT_KEYUSAGE_-DIGITALSIGNATURE and is interpreted in various incompatible ways by different standards and profiles.

For example to mark the key in a certificate as being usable for digital signatures and encryption you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_KEYUSAGE,  
CRYPT_KEYUSAGE_DIGITALSIGNATURE | CRYPT_KEYUSAGE_KEYENCIPHERMENT );
```

The extKeyUsage attribute contains a collection of one or more values that specify a specific type of extended usage that extends beyond the general keyUsage.

This extension is used by applications to determine whether a certificate is meant for a particular purpose such as timestamping or code signing. The extension is valid in certificates and certification requests and can contain any of the following values:

Value	Used in
CRYPT_CERTINFO_EXTKEY_- ANYKEYUSAGE	No-op wildcard value used to work around extended key-usage validation bugs in some software.
CRYPT_CERTINFO_EXTKEY_- CODESIGNING	Code-signing certificate.
CRYPT_CERTINFO_EXTKEY_- DIRECTORYSERVICE	Directory service certificate.
CRYPT_CERTINFO_EXTKEY_- EMAILPROTECTION	email encryption/signing certificate.

CRYPT_CERTINFO_EXTKEY_- IPSECENDSYSTEM	Various IPSEC certificates.
CRYPT_CERTINFO_EXTKEY_- IPSECTUNNEL	
CRYPT_CERTINFO_EXTKEY_- IPSECUSER	
CRYPT_CERTINFO_EXTKEY_- MS_CERTTRUSTLISTSIGNING	Microsoft certificate trust list signing and timestamping certificate, used for AuthentiCode signing.
CRYPT_CERTINFO_EXTKEY_- MS_TIMESTAMPING	
CRYPT_CERTINFO_EXTKEY_- MS_ENCRYPTEDFILESYSTEM	Microsoft encrypted file system certificate.
CRYPT_CERTINFO_EXTKEY_- MS_INDIVIDUALCODESIGNING	Microsoft individual and commercial code-signing certificate, used for AuthentiCode signing.
CRYPT_CERTINFO_EXTKEY_- MS_COMMERCIALCODESIGNING	
CRYPT_CERTINFO_EXTKEY_- MS_SERVERGATEDCRYPTO	Microsoft server-gated crypto (SGC) certificate, used to enable strong encryption on non-US servers.
CRYPT_CERTINFO_EXTKEY_- NS_SERVERGATEDCRYPTO	Netscape server-gated crypto (SGC) certificate, used to enable strong encryption on non-US servers.
CRYPT_CERTINFO_EXTKEY_- OCSPSIGNING	OCSP response signing.
CRYPT_CERTINFO_EXTKEY_- SERVERAUTH	SSL server and client authentication certificate.
CRYPT_CERTINFO_EXTKEY_- CLIENTAUTH	
CRYPT_CERTINFO_EXTKEY_- TIMESTAMPING	Timestamping certificate.
CRYPT_CERTINFO_EXTKEY_- VS_SERVERGATEDCRYPTO_CA	Verisign server-gated crypto CA certificate, used to sign SGC certificates.

For example to mark the key in a certificate as being used for SSL server authentication you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_EXTKEY_SERVERAUTH,  
CRYPT_UNUSED );
```

Like the keyUsage extension, the Netscape certificate type extension has a single numeric attribute with the same identifier as the extension itself (CRYPT_CERTINFO_NS_CERTTYPE). This extension is valid in certificates and certification requests and contains a bit flag that can contain any of the following values:

Value	Used in
CRYPT_NS_CERTTYPE_- OBJECTSIGNING	Object signing certificate (equivalent to Microsoft's AuthentiCode use).
CRYPT_NS_CERTTYPE_- SMIME	S/MIME email encryption/signing certificate.

CRYPT_NS_CERTTYPE_- SSLCLIENT	SSL client and server certificate.
CRYPT_NS_CERTTYPE_- SSLSERVER	
CRYPT_NS_CERTTYPE_- SSLCA	CA certificates corresponding to the above
CRYPT_NS_CERTTYPE_- SMIMECA	certificate types. Using these flags requires
CRYPT_NS_CERTTYPE_- OBJECTSIGNINGCA	the basicConstraint CA value to be set.

This extension is obsolete and is supported as a read-only attribute by cryptlib. If you try to set this extension cryptlib will return CRYPT_ERROR_PERMISSION to indicate that you can't set this attribute value.

Name Constraints

The nameConstraints extension is used to constrain the certificate's subjectName and subject altName to lie inside or outside a particular DN subtree or substring, with the excludedSubtrees attribute taking precedence over the permittedSubtrees attribute. The principal use for this extension is to allow control of the certificate namespace, so that a CA can restrict the ability of any CAs it certifies to issue certificates outside a very restricted domain (for example corporate headquarters might constrain a divisional CA to only issue certificates for its own business division). This extension is identified by CRYPT_CERTINFO_NAMECONSTRAINTS, and is valid in certificates:

Attribute/Description	Type
CRYPT_CERTINFO_PERMITTEDSUBTREES The subtree within which the subjectName and subject altName of any issued certificates must lie.	GeneralName
CRYPT_CERTINFO_EXCLUDEDSUBTREES The subtree within which the subjectName and subject altName of any issued certificates must not lie.	GeneralName

Due to ambiguities in the encoding rules for strings contained in DNs, it is possible to avoid the excludedSubtrees for DNs by choosing unusual (but perfectly valid) string encodings that don't appear to match the excludedSubtrees. Because of this you should rely on permittedSubtrees rather than excludedSubtrees for DN constraint enforcement.

The nameConstraints are applied to both the certificate subject name and the subject altName. For example if a CA run by Cognitive Cybernetics Incorporated wanted to issue a certificate to a subsidiary CA that was only permitted to issue certificates for Cognitive Cybernetics' marketing division, it would set DN name constraints with:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_PERMITTEDSUBTREES,  
                  CRYPT_UNUSED );  
cryptSetAttribute( certificate, CRYPT_CERTINFO_DIRECTORYNAME,  
                  CRYPT_UNUSED );  
cryptSetAttributeString( certificate, CRYPT_CERTINFO_COUNTRYNAME,  
                        "US", 2 );  
cryptSetAttributeString( certificate, CRYPT_CERTINFO_ORGANIZATIONNAME,  
                        "Cognitive Cybernetics Incorporated", 32 );  
cryptSetAttributeString( certificate,  
                        CRYPT_CERTINFO_ORGANIZATIONALUNITNAME, "Marketing", 9 );
```

This means that the subsidiary CA can only issue certificates to employees of the marketing division. Note that since the excludedSubtrees attribute is a GeneralName, the DN is selected through a two-level process, first to select the excludedSubtrees GeneralName and then to select the DN within the GeneralName.

GeneralName components that have a flat structure (for example email addresses) can have constraints specified through the '*' wildcard. For example to extend the above

constraint to also include email addresses, the issuing CA would set a name constraint with:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_PERMITTEDSUBTREES,
CRYPT_UNUSED );
cryptSetAttributeString( certificate, CRYPT_CERTINFO_RFC822NAME,
"@marketing.cci.com", 19 );
```

This means that the subsidiary CA can only issue certificates with email addresses within the marketing division. Note again the selection of the excludedSubtrees GeneralName followed by the setting of the email address (if the GeneralName is still selected from the earlier code, there's no need to re-select it at this point).

Private Key Usage Period

This extensions specifies the date on which the private key for this certificate expires. This extension is identified by CRYPT_CERTINFO_-PRIVATEKEYUSAGEPERIOD and is valid in certificates. This is useful where a certificate needs to have a much longer lifetime than the private key it corresponds to, for example a long-term signature might have a lifetime of 10-20 years, but the private key used to generate it should never be retained for such a long period. The privateKeyUsagePeriod extension is used to specify a (relatively) short lifetime for the private key while allowing for a very long lifetime for the signatures it generates:

Attribute/Description	Type
CRYPT_CERTINFO_PRIVATEKEY_NOTBEFORE	Time
CRYPT_CERTINFO_PRIVATEKEY_NOTAFTER	Time
The private key usage period defines the period of time over which the private key for a certificate object is valid. CRYPT_CERTINFO_-PRIVATEKEY_NOTBEFORE specifies the validity start period, and CRYPT_CERTINFO_PRIVATEKEY_NOTAFTER specifies the validity end period.	

Subject and Authority Key Identifiers

These extensions are used to provide additional identification information for a certificate, and are usually generated automatically by certificate management code. For this reason the extensions are marked as read-only.

The authorityKeyIdentifier is identified by CRYPT_CERTINFO_-AUTHORITYKEYIDENTIFIER and has the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_AUTHORITY_KEYIDENTIFIER	Binary data
Binary data identifying the public key in the certificate that was used to sign this certificate.	
CRYPT_CERTINFO_AUTHORITY_CERTISSUER	GeneralName
CRYPT_CERTINFO_AUTHORITY_- CERTSERIALNUMBER	Binary data

The issuer name and serial number for the certificate that was used to sign this certificate. The serial number is treated as a binary string and not as a numeric value, since it is often 15-20 bytes long.

The subjectKeyIdentifier is identified by CRYPT_CERTINFO_-SUBJECTKEYIDENTIFIER and contains binary data identifying the public key in the certificate.

CRL Extensions

CRLs have a number of CRL-specific extensions that are described below.

CRL Reasons, CRL Numbers, Delta CRL Indicators

These extensions specify various pieces of information about CRLs. The reasonCode extension is used to indicate why a certificate was revoked. The cRLNumber

extension provides a serial number for CRLs. The deltaCRLIndicator indicates a delta CRL that contains changes between a base CRL and a delta-CRL (this is used to reduce the overall size of CRLs).

The reasonCode extension is identified by CRYPT_CERTINFO_CRLREASON and is valid in CRLs. The extension has a single numeric attribute with the same identifier as the extension itself (CRYPT_CERTINFO_CRLREASON) which contains a bit flag that can contain one of the following values:

Value	Description
CRYPT_CRLREASON_- AFFILIATIONCHANGED	The affiliation of the certificate owner has changed, so that the subjectName or subject altName is no longer valid.
CRYPT_CRLREASON_- CACOMPROMISE CRYPTCRLREASON_- AACOMPROMISE	The CA or attribute authority that issued the certificate was compromised.
CRYPT_CRLREASON_- CERTIFICATEHOLD	The certificate is to be placed on hold pending further communication from the CA (the further communication may be provided by the holdInstructionCode extension).
CRYPT_CRLREASON_- CESSATIONOFOPERATION	The certificate owner has ceased to operate in the role that requires the use of the certificate.
CRYPT_CRLREASON_- KEYCOMPROMISE	The key for the certificate was compromised.
CRYPT_CRLREASON_- PRIVILEGEWITHDRAWN	The privilege granted in an attribute certificate is no longer valid.
CRYPT_CRLREASON_- REMOVEFROMCRL	The certificate should be removed from the certificate revocation list.
CRYPT_CRLREASON_- SUPERSEDED	The certificate has been superseded.
CRYPT_CRLREASON_- UNSPECIFIED	No reason for the CRL. You should avoid including a reasonCode at all rather than using this code.

To indicate that a certificate is being revoked because the key it corresponds to has been compromised, you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_CRLREASON,  
                  CRYPT_CRLREASON_KEYCOMPROMISE );
```

The cRLNumber extension is identified by CRYPT_CERTINFO_CRLNUMBER and is valid in CRLs. The extension has a single attribute with the same identifier as the extension itself (CRYPT_CERTINFO_CRLNUMBER) which contains a monotonically increasing sequence number for each CRL issued. This allows an application to check that it has received and processed each CRL that was issued.

The deltaCRLIndicator extension is identified by CRYPT_CERTINFO_-DELTA_CRLINDICATOR and is valid in CRLs. The extension has a single attribute with the same identifier as the extension itself (CRYPT_CERTINFO_-DELTA_CRLINDICATOR) which contains the cRLNumber of the base CRL from which this delta CRL is being constructed (see certificate standards documents for more information on delta CRLs).

Hold Instruction Code

This extension contains a code that specifies what to do with a certificate that has been placed on hold through a CRL (that is, its revocation reasonCode is CRYPT_CRLREASON_CERTIFICATEHOLD). The extension is identified by CRYPT_CERTINFO_HOLDINSTRUCTIONCODE, is valid in CRLs, and can contain one of the following values:

Value	Description
CRYPT_HOLDINSTRUCTION_ - CALLISSUER	Call the certificate issuer for details on the certificate hold.
CRYPT_HOLDINSTRUCTION_NONE	No hold instruction code. You should avoid including a holdInstructionCode at all rather than using this code.
CRYPT_HOLDINSTRUCTION_ - REJECT	Reject the transaction that the revoked/held certificate was to be used for.

As the hold code descriptions indicate, this extension was developed mainly for use in the financial industry. To indicate that someone should call the certificate issuer for further information on a certificate hold, you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_HOLDINSTRUCTIONCODE,
  CRYPT_HOLDINSTRUCTION_CALLISSUER );
```

You shouldn't use this extension (or the CRYPT_CRLREASON_ - CERTIFICATEHOLD reasonCode) unless you really need to because although a mechanism was defined for placing a certificate on hold, no-one ever defined one for removing it from this state, so once it's on hold it's revoked no matter what the reasonCode says.

Invalidity Date

This extension contains the date on which the private key for a certificate became invalid. The extension is identified by CRYPT_CERTINFO_INVALIDITYDATE and is valid in CRLs:

Attribute/Description	Type
CRYPT_CERTINFO_INVALIDITYDATE	Time
The date on which the key identified in a CRL became invalid.	

Note that a CRL contains both its own date and a date for each revoked certificate, so this extension is only useful if there's some reason for communicating the fact that a key compromise occurred at a time other than the CRL issue time or the certificate revocation time.

Issuing Distribution Point and Certificate Issuer

These extensions specify the CRL distribution point for a CRL and provide various pieces of additional information about the distribution point. The issuingDistributionPoint specifies the distribution point for a CRL, and the certificateIssuer specifies the issuer for an indirect CRL as indicated by the issuingDistributionPoint extension.

The issuingDistributionPoint extension is identified by CRYPT_CERTINFO_ - ISSUINGDISTRIBUTIONPOINT and is valid in CRLs:

Attribute/Description	Type
CRYPT_CERTINFO_ISSUINGDIST_ FULLNAME	GeneralName
The location at which CRLs may be obtained. You should use the URL component of the GeneralName for this, avoiding the other possibilities.	

CRYPT_CERTINFO_ISSUINGDIST_USERCERTONLY	Boolean
CRYPT_CERTINFO_ISSUINGDIST_CACERTONLY	Boolean
CRYPT_CERTINFO_ISSUINGDIST_SOMEREASONONLY	Numeric
CRYPT_CERTINFO_ISSUINGDIST_INDIRECTCRL	Boolean
See certificate standards documents.	

Note that the CRYPT_CERTINFO_ISSUINGDIST_SOMEREASONONLY attribute has the same allowable set of values as the cRLReasons reasonCode, but in this case is given as a series of bit flags rather than the reasonCode numeric value (because X.509 says so, that's why). Because of this you must use CRYPT_CRLREASONFLAGS_name instead of CRYPT_CRLREASON_name when getting and setting these values.

The certificateIssuer extension contains the certificate issuer for an indirect CRL. The extension is identified by CRYPT_CERTINFO_CERTIFICATEISSUER and is valid in CRLs:

Attribute/Description	Type
CRYPT_CERTINFO_CERTIFICATEISSUER	GeneralName
See certificate standards documents.	

Digital Signature Legislation Extensions

Various digital signature laws specify extensions beyond the X.509v3, X.509v4, and X.509v5 ones that are described below.

Certificate Generation Date

The German signature law specifies an extension containing the date at which the certificate was generated. This is necessary for post-dated certificates to avoid problems if the CA's key is compromised between the time the certificate is issued and the time it takes effect. The extension is identified by CRYPT_CERTINFO_SIGG_DATEOFCERTGEN and contains the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_SIGG_DATEOFCERTGEN	Time
The date on which the certificate was issued.	

Other Restrictions

The German signature law specifies an extension containing any other general free-form restrictions that may be imposed on the certificate. The extension is identified by CRYPT_CERTINFO_SIGG_RESTRICTION and contains the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_SIGG_RESTRICTION	String
Text containing any further restrictions not already handled via certificate policies or constraints.	

Reliance Limit

The German signature law specifies an extension containing a reliance limit for the certificate, which specifies the (recommended) monetary reliance limit for the certificate. The extension is identified by CRYPT_CERTINFO_SIGG_MONETARYLIMIT and contains the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_SIGG_MONETARY_CURRENCY The three-letter currency code.	String
CRYPT_CERTINFO_SIGG_MONETARY_AMOUNT The amount, specified as an integer in the range 1...200.	Integer
CRYPT_CERTINFO_SIGG_MONETARY_EXPONENT The exponent for the amount, specified as an integer 1...200, so that the actual value is $\text{amount} \times 10^{\text{exponent}}$.	Integer

Signature Delegation

The German signature law specifies an extension containing details about signature delegation, in which one party may sign on behalf of another (for example someone's secretary signing correspondence on their behalf). The extension is identified by CRYPT_CERTINFO_SIGG_PROCURATION and contains the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_SIGG_PROCURE_- TYPEOFSUBSTITUTION The type of signature delegation being performed (for example "Signed on behalf of").	String
CRYPT_CERTINFO_SIGG_PROCURE_SIGNINGFOR The identity of the person or organisation the signer is signing on behalf of.	GeneralName

Qualified Certificate Extensions

Qualified certificates contain additional extensions beyond the X.509v3, X.509v4, and X.509v5 ones that are described below.

Biometric Info

The biometricInfo extension contains biometric information in the form of a hash of a biometric template. The extension is identified by CRYPT_CERTINFO_-BIOMETRICINFO and is valid in certificates and certification requests:

Attribute/Description	Type
CRYPT_CERTINFO_BIOMETRICINFO_TYPE The type of the biometric data, see certificate standards documents.	Numeric
CRYPT_CERTINFO_BIOMETRICINFO_HASHALGO The object identifier for the hash algorithm used to hash the biometric template.	String
CRYPT_CERTINFO_BIOMETRICINFO_HASH The hash of the biometric template.	String
CRYPT_CERTINFO_BIOMETRICINFO_URL An optional URL at which the biometric data may be found.	String

QC Statements

The qcStatements extension contains defined statements for a qualified certificate. The extension is identified by CRYPT_CERTINFO_QCSTATEMENT and is valid in certificates and certification requests:

Attribute/Description	Type
CRYPT_CERTINFO_QCSTATEMENT_SEMANTICS An object identifier identifying the defined statement for this certificate.	String
CRYPT_CERTINFO_QCSTATEMENT_- REGISTRATIONAUTHORITY See certificate standards documents.	String

SET Extensions

SET specifies a number of extensions beyond the X.509v3, X.509v4, and X.509v5 ones that are described below.

SET Card Required and Merchant Data

These extensions specify various pieces of general information used in the SET electronic payment protocol.

The cardRequired extension contains a flag indicating whether a card is required for a transaction. The extension is identified by CRYPT_CERTINFO_SET_-CERTCARDREQUIRED, and is valid in certificates and certification requests. The extension contains a single boolean attribute with the same identifier as the extension itself (CRYPT_CERTINFO_SET_CARDREQUIRED) which is explained in the SET standards documents.

The merchantData extension contains further information on a merchant. The extension is identified by CRYPT_CERTINFO_SET_MERCHANTDATA and is valid in certificates and certification requests:

Attribute/Description	Type
CRYPT_CERTINFO_SET_MERACQUIRERBIN	String
CRYPT_CERTINFO_SET_MERAUTHFLAG	Boolean
CRYPT_CERTINFO_SET_MERCOUNTRY	Numeric
CRYPT_CERTINFO_SET_MERID Merchant's 6-digit BIN, authorisation flag, ISO country code, and merchant ID.	String
CRYPT_CERTINFO_SET_MERCHANTCITY	String
CRYPT_CERTINFO_SET_MERCHANTCOUNTRYNAME	String
CRYPT_CERTINFO_SET_MERCHANTLANGUAGE	String
CRYPT_CERTINFO_SET_MERCHANTNAME	String
CRYPT_CERTINFO_SET_MERCHANTPOSTALCODE	String
CRYPT_CERTINFO_SET_MERCHANTSTATEPROVINCE Merchant's language, name, city, state or province, postal code, and country name.	String

SET Certificate Type, Hashed Root Key, and Tunnelling

These extensions specify various pieces of certificate management information used in the SET electronic payment protocol.

The certificateType extension contains the SET certificate type. The extension is identified by CRYPT_CERTINFO_SET_CERTIFICATETYPE and is valid in certificates and certification requests. The extension contains a single bit flag attribute with the same identifier as the extension itself (CRYPT_CERTINFO_SET_-CERTIFICATETYPE) and can contain any of the following values that are explained in the SET standards documentation:

Value
CRYPT_SET_CERTTYPE_ACQ
CRYPT_SET_CERTTYPE_BCA
CRYPT_SET_CERTTYPE_CARD

CRYPT_SET_CERTTYPE_CCA
 CRYPT_SET_CERTTYPE_GCA
 CRYPT_SET_CERTTYPE_MCA
 CRYPT_SET_CERTTYPE_MER
 CRYPT_SET_CERTTYPE_PCA
 CRYPT_SET_CERTTYPE_PGWW
 CRYPT_SET_CERTTYPE_RCA

The hashedRootKey extension contains a thumbprint (SET-speak for a hash) of a SET root key. The extension is identified by CRYPT_CERTINFO_SET_HASHEDROOTKEY and is valid in certificates and certification requests. The extension contains a single attribute:

Attribute/Description	Type
CRYPT_CERTINFO_SET_ROOTKEYTHUMBPRINT Binary string containing the root key thumbprint (see the SET standards documents).	Binary data

You can obtain the key hash which is required for the thumbprint from another certificate by reading its CRYPT_CERTINFO_SUBJECTKEYIDENTIFIER attribute and then adding it to the certificate you're working with as the CRYPT_CERTINFO_SET_ROOTKEYTHUMBPRINT attribute. cryptlib will perform the further work required to convert this attribute into the root key thumbprint.

The tunnelling extension contains a tunnelling indicator and algorithm identifier. The extension is identified by CRYPT_CERTINFO_SET_TUNNELING and is valid in certificates and certification requests.

Attribute/Description	Type
CRYPT_CERTINFO_SET_TUNNELINGFLAG	Boolean
CRYPT_CERTINFO_SET_TUNNELINGALGID See SET standards documents.	String

Application-specific Extensions

Various applications such as certificate management protocols have their own extensions that extend or complement the X.509 ones. These are described below.

OCSP Extensions

These extensions specify various pieces of certificate management information used in the OCSP certificate management protocol.

The noCheck extension indicates that the certificate should be automatically trusted when used to sign OCSP responses. The extension is identified by CRYPT_CERTINFO_OCSP_NOCHECK and is valid in certificates and certification requests. The extension contains a numeric attribute with the same identifier as the extension itself (CRYPT_CERTINFO_OCSP_NOCHECK) which is always set to CRYPT_UNUSED since it has no inherent value associated with it.

Attribute/Description	Type
CRYPT_CERTINFO_OCSP_NOCHECK See OCSP standards documents.	Numeric

Vendor-specific Extensions

A number of vendors have defined their own extensions that extend or complement the X.509 ones. These are described below.

Netscape Certificate Extensions

Netscape defined a number of extensions that mostly predate the various X.509v3 extensions that now provide the same functionality. The various Netscape certificate extensions are:

Extension/Description	Type
CRYPT_CERTINFO_NS_BASEURL A base URL which, if present, is added to all partial URL's in Netscape extensions to create a full URL.	String
CRYPT_CERTINFO_NS_CAPOLICYURL The URL at which the certificate policy under which this certificate was issued can be found.	String
CRYPT_CERTINFO_NS_CAREVOCATIONURL The URL at which the revocation status of a CA certificate can be checked.	String
CRYPT_CERTINFO_NS_CERTRENEWALURL The URL at which a form allowing renewal of this certificate can be found.	String
CRYPT_CERTINFO_NS_COMMENT A comment which should be displayed when the certificate is viewed.	String
CRYPT_CERTINFO_NS_REVOCATIONURL The URL at which the revocation status of a server certificate can be checked.	String
CRYPT_CERTINFO_NS_SSLSERVERNAME A wildcard string containing a shell expression that matches the hostname of the SSL server using this certificate.	String

Note that each of these entries represent a separate extension containing a single text string, they have merely been listed in a single table for readability. You should avoid using these extensions if possible and instead use one of the standard X.509v3 extensions.

Thawte Certificate Extensions

Thawte Consulting have defined an extension that allows the use of certificates with secure extranets. This extension is identified by CRYPT_CERTINFO_-STRONGEXTRANET and is valid in certificates and certification requests:

Attribute/Description	Type
CRYPT_CERTINFO_STRONGEXTRANET_ZONE	Numeric
CRYPT_CERTINFO_STRONGEXTRANET_ID Extranet zone and ID.	Binary data

Generic Extensions

Beyond the standardised extensions listed above there exist any number of obscure or non-standard certificate extensions. cryptlib allows you to work with these extensions using **cryptAddCertExtension**, **cryptGetCertExtension**, and **cryptDeleteCertExtension**, which allow you to add, retrieve, or delete a complete encoded extension identified by its ASN.1 object identifier. The extension data must be a complete DER-encoded ASN.1 object without the OCTET STRING wrapper which is used for all extensions (cryptlib will add this itself). For example if you wanted to add a 4-byte UTF8 string as an extension the data would be 0C 04 xx xx xx xx. If you pass in extension data to **cryptAddCertExtension** that isn't a valid ASN.1-encoded object, cryptlib will return CRYPT_ERROR_PARAM4 to indicate that the data is in an invalid format.

If a certificate object contains a non-standard extension, cryptlib won't include it in the object when you sign it unless you set the CRYPT_OPTION_CERT_-SIGNUNRECOGNISEDATTRIBUTES option to true. This is to avoid problems

where a CA could end up signing arbitrary data in an unrecognised certificate extension.

If the extension you are trying to add is already handled as a standard extension, cryptlib will return `CRYPT_ERROR_PERMISSION` to indicate that you can't add the extension in this manner but have to add it using **`cryptSetAttribute/`****`cryptSetAttributeString`**.

Other Certificate Object Extensions

Certificate objects other than certificates and CRLs can also contain extensions. In the following descriptions only the generally useful attributes have been described. The full range of attributes is enormous and will probably never be used in real life. These attributes are marked with “See standards documents” to indicate that you should refer to other documents to obtain information about their usage (this is also a good indication that you shouldn’t really be using this attribute).

CMS/SMIME Attributes

The CMS and S/MIME standards specify various attributes that can be included with signatures. In addition there are a variety of proprietary and vendor-specific attributes that are also handled by cryptlib. In the following description only the generally useful attributes have been described, the full range of attributes is enormous and requires a number of standards specifications (often followed by cries for help on mailing lists) to interpret them. These attributes are marked with “See S/MIME standards documents” to indicate that you should refer to other documents to obtain information about their use (this is also a good indication that you shouldn’t really be using this attribute).

Content Type

This is a standard CMS attribute identified by CRYPT_CERTINFO_CMS_-CONTENTTYPE and is used to specify the type of data which is being signed. This is used because some signed information could be interpreted in different ways depending on the data type it’s supposed to represent (for example something viewed as encrypted data could be interpreted quite differently if viewed as plain data). If you don’t set this attribute, cryptlib will set it for you and mark the signed content as plain data.

The content-type CMS attribute can contain one of the following CRYPT_-CONTENT_TYPE values:

Value	Description
CRYPT_CONTENT_DATA	Plain data.
CRYPT_CONTENT_- SIGNEDDATA	Signed data.
CRYPT_CONTENT_- ENVELOPEDDATA	Data encrypted using a password or public-key or conventional encryption.
CRYPT_CONTENT_- SIGNEDANDENVELOPED- DATA	Data which is both signed and enveloped (this is an obsolete composite content type that shouldn’t be used).
CRYPT_CONTENT_- DIGESTEDDATA	Hashed data.
CRYPT_CONTENT_- ENCRYPTEDDATA	Data encrypted directly with a session key.
CRYPT_CONTENT_- COMPRESSEDATA	Compressed data.
CRYPT_CONTENT_TSTINFO	Timestamp token generated by a timestamp authority (TSA).
CRYPT_CONTENT_- SPCINDIRECTDATA- CONTEXT	Indirectly signed data used in Authenticode signatures.

The distinction between the different types arises from the way they are specified in the standards documents, as a rule of thumb if the data being signed is encrypted then

use CRYPT_CONTENT_ENVELOPEDDATA (rather than CRYPT_CONTENT_ENCRYPTEDDATA, which is slightly different), if it's signed then use CRYPT_CONTENT_SIGNEDDATA, and if it's anything else then use CRYPT_CONTENT_DATA. For example to identify the data you're signing as encrypted data, you would use:

```
cryptSetAttribute( cmsAttributes, CRYPT_CERTINFO_CMS_CONTENTTYPE,
    CRYPT_CONTENT_ENVELOPEDDATA );
```

If you're generating the signature via the cryptlib enveloping code then cryptlib will set the correct type for you so there's no need to set it yourself.

Countersignature

This CMS attribute contains a second signature that countersigns one of the signatures on the data (that is, it signs the other signature rather than the data). The attribute is identified by CRYPT_CERTINFO_CMS_COUNTERSIGNATURE:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_COUNTERSIGNATURE See S/MIME standards documents.	Binary data

Message Digest

This read-only CMS attribute is used as part of the signing process and is generated automatically by cryptlib. The attribute is identified by CRYPT_CERTINFO_CMS_MESSAGEDIGEST:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_MESSAGEDIGEST The hash of the content being signed.	Binary data

Signing Description

This CMS attribute contains a short text message with an additional description of the data being signed. For example if the signed message was a response to a received signed message, the signing description might contain an indication of the type of message it's being sent in response to. Note that CMS has a number of special-purpose signing attributes such as message receipt information that allow automated processing of messages that contain them, so you should only use this free-form human-readable attribute for cases that aren't covered by special-case attributes designed for the purpose.

The attribute is identified by CRYPT_CERTINFO_CMS_SIGNINGDESCRIPTION:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SIGNINGDESCRIPTION Free-form text annotation for the message being signed.	String

Signing Time

This is a standard CMS attribute identified by CRYPT_CERTINFO_CMS_SIGNINGTIME and is used to specify the time at which the signature was generated. If you don't set this attribute, cryptlib will set it for you.

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SIGNINGTIME The time at which the signature was generated.	Time

Extended CMS/SMIME Attributes

The attributes given above are the standard CMS attributes. Extending beyond this are further attributes that are defined in additional standards documents and that apply

mostly to S/MIME messages, as well as vendor-specific and proprietary attributes. Before you use these additional attributes you should ensure that any software you plan to interoperate with can process them, since currently almost nothing will recognise them (for example it's not a good idea to put a security label on your data and expect other software to handle it correctly).

AuthentiCode Attributes

AuthentiCode code-signing uses a number of attributes that apply to signed executable content. These attributes are listed below.

The agency information CMS attribute, identified by CRYPT_CERTINFO_CMS_-SPCAGENCYINFO, is used to provide extra information about the signer of the data and has the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SPCAGENCYURL The URL of a web page containing more information about the signer.	String

The statement type CMS attribute, identified by CRYPT_CERTINFO_CMS_-SPCSTATEMENTTYPE, is used to identify whether the content was signed by an individual or a commercial organisation, and has the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SPCSTMT_INDIVIDUAL-CODESIGNING The data was signed by an individual.	Numeric
CRYPT_CERTINFO_CMS_SPCSTMT_COMMERCIAL-CODESIGNING The data was signed by a commercial organisation.	Numeric

The opus info CMS attribute, identified by CRYPT_CERTINFO_CMS_-SPCOPUSINFO, is used to identify program details for AuthentiCode use, and has the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SPCOPUSINFO_NAME Program name/version.	String
CRYPT_CERTINFO_CMS_SPCOPUSINFO_URL AuthentiCode information URL.	String

Note that the CRYPT_CERTINFO_CMS_SPCOPUSINFO_NAME attribute is a Unicode string, as used by Windows NT/2000/XP/Vista and Windows CE.

For example to indicate that the data was signed by an individual, you would use:

```
cryptSetAttribute( cmsAttributes,
    CRYPT_CERTINFO_CMS_SPCSTMT_COMMERCIALCODESIGNING, CRYPT_UNUSED );
```

For example to create an AuthentiCode signature as a commercial organisation you would use:

```
CRYPT_CERTIFICATE cmsAttributes;

/* Create the CMS attribute object and add the AuthentiCode attributes
 */
cryptCreateCert( &cmsAttributes, cryptUser,
    CRYPT_CERTTYPE_CMS_ATTRIBUTES );
cryptSetAttributeString( cmsAttributes,
    CRYPT_CERTINFO_CMS_SPCAGENCYURL,
    "http://homepage.organisation.com", 32 );
cryptSetAttribute( cmsAttributes,
    CRYPT_CERTINFO_CMS_SPCSTMT_COMMERCIALCODESIGNING, CRYPT_UNUSED );

/* Add the content-type required for AuthentiCode data */
cryptSetAttribute( cmsAttributes, CRYPT_CERTINFO_CMS_CONTENTTYPE,
    CRYPT_CONTENT_SPCINDIRECTDATACONTEXT );
```

```

/* Sign the data with the attributes included */
cryptCreateSignatureEx( ... );

cryptDestroyCert( cmsAttributes );

```

The other attributes used when signing are standard attributes that will be added automatically for you by cryptlib.

Content Hints

This CMS attribute can be supplied in the outer layer of a multi-layer message to provide information on what the innermost layer of the message contains. The attribute is identified by CRYPT_CERTINFO_CMS_CONTENTHINTS and has the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_CONTENTHINT_DESCRIPTION	String
A human-readable description that may be useful when processing the content.	
CRYPT_CERTINFO_CMS_CONTENTHINT_TYPE	Numeric
The type of the innermost content, specified as a CRYPT_CONTENT_ - <i>content-type</i> value.	

DOMSEC Attributes

The domain security (DOMSEC) attributes are used to handle delegated signing by systems such as mail gateways. The signature type CMS attribute, identified by CRYPT_CERTINFO_CMS_SIGTYPEIDENTIFIER, is used to identify the signature type, and has the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SIGTYPEID_ADDITIONALATTRIBUTES	Numeric
Additional attributes for a domain signature.	
CRYPT_CERTINFO_CMS_SIGTYPEID_DOMAINSIG	Numeric
Domain signature by a gateway on behalf of a user.	
CRYPT_CERTINFO_CMS_SIGTYPEID_ORIGINATORSIG	Numeric
Indication that the signer is the originator of the message. This attribute isn't normally used, since it corresponds to a standard (non-DOMSEC) signature..	
CRYPT_CERTINFO_CMS_SIGTYPEID_REVIEWSIG	Numeric
Review signature to indicate that the domain signer has reviewed the message.	

Mail List Expansion History

This CMS attribute contains information on what happened to a message when it was processed by mailing list software. It is identified by CRYPT_CERTINFO_CMS_MLEXPAHNSIONHISTORY and contains the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_MLEXP_ENTITYIDENTIFIER See S/MIME standards documents.	Binary data
CRYPT_CERTINFO_CMS_MLEXP_TIME The time at which the mailing-list software processed the message.	Time
CRYPT_CERTINFO_CMS_MLEXP_NONE	—
CRYPT_CERTINFO_CMS_MLEXP_INSTEADOF	General-
CRYPT_CERTINFO_CMS_MLEXP_INADDITIONTO	Name
This attribute can have one of the three values specified above, and is used to indicate a receipt policy that overrides the one given in the original message. See the S/MIME standards documents for more information.	

Nonce

This CMS attribute nonce is used to prevent replay attacks. The attribute is identified by CRYPT_CERTINFO_CMS_NONCE:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_NONCE Nonce to prevent replay attacks.	Binary data

Receipt Request

This CMS attribute is used to request a receipt from the recipient of a message and is identified by CRYPT_CERTINFO_CMS_RECEIPT_REQUEST. As with the security label attribute, you shouldn't rely on the recipient of a message being able to do anything with this information, which consists of the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_RECEIPT_- CONTENTIDENTIFIER A magic value used to identify a message, see the S/MIME standards documents for more information.	Binary data
CRYPT_CERTINFO_CMS_RECEIPT_FROM	Numeric
CRYPT_CERTINFO_CMS_RECEIPT_TO	General-
An indication of who receipts should come from and who they should go to, see the S/MIME standards documents for more information.	

SCEP Attributes

The Simple Certificate Enrolment Protocol uses a variety of protocol-specific attributes that are attached to CMS signed data and are used to manage the operation of the protocol. These attributes are not normally used with CMS but are provided for use by cryptlib's SCEP implementation. The SCEP attributes are:

Attribute/Description	Type
CRYPT_CERTINFO_SCEP_MESSAGETYPE The SCEP message type.	String
CRYPT_CERTINFO_SCEP_PKISTATUS The processing status of an SCEP request.	String
CRYPT_CERTINFO_SCEP_FAILINFO Extended error information if the SCEP processing status indicates that an error occurred.	String
CRYPT_CERTINFO_SCEP_SENDERNONCE CRYPT_CERTINFO_SCEP_RECIPIENTNONCE Nonce values used to protect against message replay attacks. Note that these values duplicate the more usual CRYPT_CERTINFO_CMS_NONCE attribute, which should be used in place of these attributes unless they're specifically being used for SCEP.	Binary data
CRYPT_CERTINFO_SCEP_TRANSACTIONID A value that uniquely identifies the entity requesting a certificate.	String
In addition to these attributes, SCEP also uses an additional attribute which is added to PKCS #10 requests even though it's a CMS attribute. It therefore acts as a certificate attribute rather than a CMS attribute. The attribute is identified by CRYPT_CERTINFO_CHALLENGEPASSWORD:	
Attribute/Description	Type
CRYPT_CERTINFO_CHALLENGEPASSWORD Password used to authorise certificate issue requests.	String

Security Label, Equivalent Label

These CMS attributes specify security information for the content contained in the message, allowing recipients to decide how they should process it. For example an implementation could refuse to display a message to a recipient who isn't cleared to see it (this assumes that the recipient software is implemented at least in part using tamper-resistant hardware, since a pure software implementation could be set up to ignore the security label). These attributes originate (in theory) in X.400 and (in practice) in DMS, the US DoD secure email system, and virtually no implementations outside this area understand them so you shouldn't rely on them to ensure proper processing of a message.

The basic security label on a message is identified by CRYPT_CERTINFO_CMS_SECURITYLABEL. Since different organisations have different ways of handling security policies, their labelling schemes may differ, so the equivalent labels CMS attribute, identified by CRYPT_CERTINFO_CMS_EQUIVALENTLABEL, can be used to map from one to the other. These contain the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SECLABEL_POLICY The object identifier for the security policy that the security label is issued under.	String
CRYPT_CERTINFO_CMS_SECLABEL_- CLASSIFICATION The security classification for the content identified relative to the security policy being used. There are six standard classifications (described below) and an extended number of user-defined classifications, for more information see the S/MIME standards documents and X.411.	Numeric
CRYPT_CERTINFO_CMS_SECLABEL_PRIVACYMARK A privacy mark value that unlike the security classification isn't used for access control to the message contents. See S/MIME standards documents for more information.	Numeric
CRYPT_CERTINFO_CMS_SECLABEL_CATTYPE CRYPT_CERTINFO_CMS_SECLABEL_CATVALUE See S/MIME standards documents.	String Binary data
The security classification can have one of the following predefined values (which are relative to the security policy and whose interpretation can vary from one organisation to another), or policy-specific, user-defined values that lie outside this range:	
Value	
CRYPT_CLASSIFICATION_UNMARKED	
CRYPT_CLASSIFICATION_UNCLASSIFIED	
CRYPT_CLASSIFICATION_RESTRICTED	
CRYPT_CLASSIFICATION_CONFIDENTIAL	
CRYPT_CLASSIFICATION_SECRET	
CRYPT_CLASSIFICATION_TOP_SECRET	

Signature Policy

This CMS attribute is used to identify the policy under which a signature was generated, and is identified by CRYPT_CERTINFO_CMS_-SIGNATUREPOLICYID. The signature policies extension allows a signer to provide information on the policies governing a signature, and to control the way in which a signature can be interpreted. For example it allows you to check that a signature was issued under a policy you feel comfortable with (certain security precautions taken, vetting of employees, physical security of the premises, and so on).

The certificate policies attribute is a complex extension that allows for all sorts of qualifiers and additional modifiers (several of them exist only because this extension was a cut & paste of a similar-looking extension that's used with certificates). In general you should only use the policyIdentifier attribute in this extension, since the other attributes are difficult to support in user software and are ignored by many implementations:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SIGPOLICYID The object identifier that identifies the policy under which this certificate was issued.	String
CRYPT_CERTINFO_CMS_SIGPOLICYHASH The hash algorithm identifier and hash of the signature policy, see signature standards documents.	Binary data

CRYPT_CERTINFO_CMS_SIGPOLICY_CPSURI	String
The URL for the certificate practice statement (CPS) for this signature policy.	
CRYPT_CERTINFO_CMS_SIGPOLICY_ORGANIZATION	String
CRYPT_CERTINFO_CMS_SIGPOLICY_-NOTICENUMBERS	Numeric
CRYPT_CERTINFO_CMS_SIGPOLICY_EXPLICITTEXT	String
These attributes contain further qualifiers, modifiers, and text information that amend the signature policy information. Refer to signature standards documents for more information on these attributes.	

S/MIME Capabilities

This CMS attribute provides additional information about the capabilities and preferences of the sender of a message, allowing them to indicate their preferred encryption algorithm(s) and . The attribute is identified by CRYPT_CERTINFO_CMS_SMIMECAPABILITIES and can contains any of the following values:

Value	Description
CRYPT_CERTINFO_CMS_-SMIMECAP_3DES	The sender supports the use of these algorithms. When encoding them, cryptlib will order them by algorithm strength so that triple DES will be preferred over Skipjack which will be preferred over DES.
CRYPT_CERTINFO_CMS_-SMIMECAP_AES	
CRYPT_CERTINFO_CMS_-SMIMECAP_CAST128	
CRYPT_CERTINFO_CMS_-SMIMECAP_DES	
CRYPT_CERTINFO_CMS_-SMIMECAP_IDEA	
CRYPT_CERTINFO_CMS_-SMIMECAP_RC2	
CRYPT_CERTINFO_CMS_-SMIMECAP_RC5	
CRYPT_CERTINFO_CMS_-SMIMECAP_SKIPJACK	
CRYPT_CERTINFO_CMS_-SMIMECAP_-PREFERSIGNEDDATA	The sender would prefer to be sent signed data.
CRYPT_CERTINFO_CMS_-SMIMECAP_-CANNOTDECRYPTANY	The sender can't handle any form of encrypted data.

To indicate that you can support messages encrypted with triple DES and Cast-128, you would use:

```
cryptSetAttribute( certificate, CRYPT_CERTINFO_CMS_SMIMECAP_3DES,
CRYPT_UNUSED );
cryptSetAttribute( certificate, CRYPT_CERTINFO_CMS_SMIMECAP_CAST128,
CRYPT_UNUSED );
```

If you're using CRYPT_FORMAT_SMIME data, cryptlib will automatically add the appropriate attributes for you so there's no need to set these attributes yourself.

Signing Certificate

This CMS attribute provides additional information about the certificate used to sign a message, is identified by CRYPT_CERTINFO_SIGNINGCERTIFICATE, and contains the following attributes:

Attribute/Description	Type
CRYPT_CERTINFO_CMS_SIGNINGCERT_ESSCERTID See S/MIME standards documents.	Binary data
CRYPT_CERTINFO_CMS_SIGNINGCERT_POLICIES The object identifier for the policy that applies to the signing certificate.	String

OCSP Attributes

Like certificates, OCSP requests and responses can contain extensions that contain additional information relating to the request or response. The `ocspNonce` extension is used to prevent replay attacks on OCSP requests and is set automatically by `cryptlib`. The `ocspArchiveCutoff` extension indicates the time limit to which an OCSP responder will store revocation information for a certificate. The `ocspResponseType` extension indicates the type of response you'd like to receive from a responder.

The `ocspNonce` extension is identified by `CRYPT_CERTINFO_OCSP_NONCE` and is valid in OCSP requests and responses. The extension has a single binary data attribute with the same identifier as the extension itself (`CRYPT_CERTINFO_OCSP_NONCE`). Since `cryptlib` sets this value automatically, you can't set it yourself:

Attribute/Description	Type
CRYPT_CERTINFO_OCSP_NONCE Nonce to prevent replay attacks.	Binary data

The `ocspArchiveCutoff` extension is identified by `CRYPT_CERTINFO_OCSP_ARCHIVECUTOFF` and is valid in OCSP responses:

Attribute/Description	Type
CRYPT_CERTINFO_OCSP_ARCHIVECUTOFF The date beyond which revocation information will no longer be archived by the responder.	Time

The `ocspResponseType` extension is identified by `CRYPT_CERTINFO_OCSP_RESPONSE` and is valid in OCSP requests. This extension contains a collection of one or more values that indicate the type of response which is being requested from the OCSP responder. The values are:

Value	Description
CRYPT_CERTINFO_OCSP_RESPONSE_OCSP	OCSP response containing only revocation information but no actual certificate status.
CRYPT_CERTINFO_OCSP_RESPONSE_RTCS	RTCS response containing OK/not OK certificate status.
CRYPT_CERTINFO_OCSP_RESPONSE_RTCS_EXTENDED	Extended RTCS response containing certificate status and additional information such as revocation information.

In addition to OCSP-specific attributes, OCSP responses can also contain the CRL attributes `reasonCode`, `holdInstructionCode`, `invalidityDate`, and `certificateIssuer`, which are described in "CRL Extensions" on page 329.

cryptlib User Interface Components

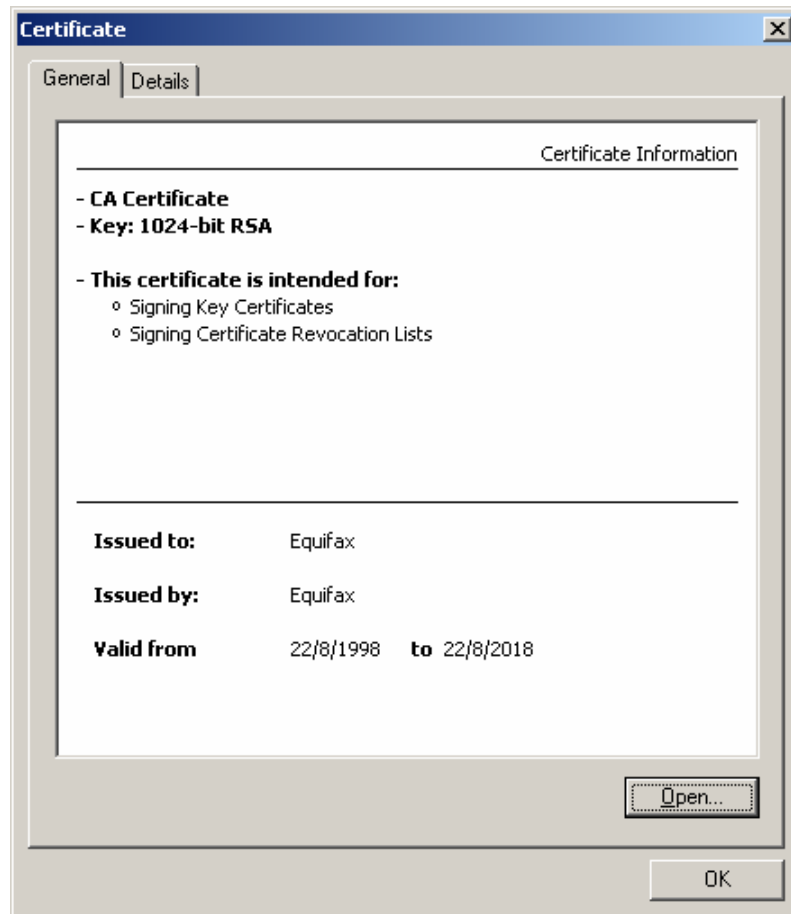
Under Win32 cryptlib provides user interface functionality via the cryptlib user interface library `cryptui.dll`, which contains functions to display certificate objects and to generate keys and obtain information needed to create or obtain a certificate. The certificate display function takes the contents of a certificate object and displays the various fields to the user in a standard resizable, tabbed dialog, adjusting the format and contents as required by the certificate object. For example a certificate chain would be displayed as a collection of certificates, where each certificate has its contents broken down and displayed as described above.

Displaying Certificates

To display a certificate object, you use `cryptUIDisplayCert`, passing in the handle of the certificate object to display and the handle of the owner window, or NULL if the window has no owner:

```
cryptUIDisplayCert( cryptCertificate, hWnd );
```

A certificate might look as follows when displayed by `cryptUIDisplayCert`:



If you set the certificate parameter for `cryptUIDisplayCert` to `CRYPT_UNUSED`, it will allow the user to choose a certificate file to load with a standard file open dialog:

```
cryptUIDisplayCert( CRYPT_UNUSED, hWnd );
```

Key/Certificate Generation

The key generation function is a powerful operation that encompasses much of the functionality covered in the chapters on key and certificate management, allowing the generation of keys for the full range of public-key algorithms supported by cryptlib, with support for the use of crypto devices such as smart cards and Fortezza cards. In

addition this function obtains from the user all the information needed to create a certificate or certification request ready for submission to a CA for signing.

The user interface is a standard wizard that takes the user through the steps of choosing an algorithm, key size, password, and various identification components needed for a certificate such as a name and email address. The general idea behind using the wizard is:

```
create a certificate object to contain the certificate information;
add any fixed certificate details if required;
call the key generation wizard;
make any required changes to the certificate contents;
use the returned key to sign the certificate object;
store the key and/or certificate in a keyset using the returned
password;
```

One stage in the `cryptUIGenerateKey` key generation process might look as follows:



In the simplest case, which involves generating a key with a certificate request ready for submission to a CA, you'd do the following:

```
CRYPT_CERTIFICATE cryptCertRequest;
CRYPT_CONTEXT cryptContext;
password[ CRYPT_MAX_TEXTSIZE + 1 ];

/* Generate the cert request */
cryptCreateCert( &cryptCertRequest, CRYPT_UNUSED,
    CRYPT_CERTTYPE_CERTREQUEST );

/* Generate the key and fill in the cert request via the key
generation wizard */
cryptUIGenerateKey( CRYPT_UNUSED, &cryptContext, cryptCertRequest,
    password, hWnd );

/* Sign the cert request */
cryptSignCert( cryptCertRequest, cryptContext );
```

Once the key has been generated by cryptlib it needs to be saved to a private key keyset as described in “Certificates and Certificate Management” on page 234. The key can also be generated using a smart card or other crypto device, in which case the first parameter is the handle to the device object:

```
cryptUIGenerateKey( cryptDevice, &cryptContext, cryptCertRequest,
    password, hWnd );
```

Since the key is in this case generated and securely stored in the crypto device, there's no need (or indeed possibility) to store it in a keyset.

The code presented so far has assumed that the user will be filling in all of the certificate request details such as the country, location, and organisation. If you want to use pre-set values for any of the certificate object components, you can fill these in before calling **cryptUIGenerateKey**. For example to default to using the company name Foo Corporation located in Canada with the certificate object you would use:

```
CRYPT_CERTIFICATE cryptCertRequest;
CRYPT_CONTEXT cryptContext;
password[ CRYPT_MAX_TEXTSIZE + 1 ];

/* Generate the cert request and fill in pre-set values */
cryptCreateCert( &cryptCertRequest, cryptUser,
    CRYPT_CERTTYPE_CERTREQUEST );
cryptSetAttributeString( cryptCertRequest,
    CRYPT_CERTINFO_ORGANISATIONNAME, "Foo Corporation", 15 );
cryptSetAttributeString( cryptCertRequest, CRYPT_CERTINFO_COUNTRYNAME,
    "CA", 2 );

/* Generate the key and fill in the cert request via the key
   generation wizard using the pre-set organisation and country name
   */
cryptUIGenerateKey( CRYPT_UNUSED, &cryptContext, cryptCertRequest,
    password, hWnd );

/* Sign the cert request */
cryptSignCert( cryptCertRequest, cryptContext );
```

In addition to a certification request it's possible to use other types of certificate objects like CMP or SCEP requests and standard certificates with **cryptUIGenerateKey**. For example if you wanted to create a self-signed CA certificate you would create a CRYPT_CERTTYPE_CERTIFICATE object instead of a CRYPT_CERTTYPE_CERTREQUEST one and set the CRYPT_CERTINFO_CA attribute to true to indicate that this is a CA certificate. Once the key has been generated and the other certificate details filled in, you can sign the certificate in the same manner as a cert request and save the result to a cryptlib private key keyset as described in "Certificates and Certificate Management" on page 234.

Encryption Devices and Modules

cryptlib's standard cryptographic functionality is provided through its built-in implementations of the required algorithms and mechanisms, however in some cases it may be desirable to use external implementations contained in cryptographic hardware or portable cryptographic devices like smart cards or PCMCIA cards. Examples of external implementations are:

- Cryptographic hardware accelerators
- PCMCIA crypto cards such as Fortezza cards
- Cryptographic smart cards
- Datakeys
- PKCS #11 crypto tokens
- Dallas iButtons
- Software encryption modules

The most common use for an external implementation is one where the hardware provides secure key storage and management functions, or where it provides specific algorithms or performance that may not be available in software.

Using an external implementation involves conceptually plugging in the external hardware or software alongside the built-in capabilities provided by cryptlib and then creating cryptlib objects (for example encryption contexts) via the device. The external cryptographic implementation is viewed as a logical device, although the "device" may be just another software implementation.

Note that the crypto device interface is intended for use with fairly complete crypto modules and devices capable of performing their own key and data storage, key management, and handling of crypto mechanisms. If all you want to do is replace one (or more) of cryptlib's built-in encryption, signing, or hash algorithms with crypto hardware, a native crypto core, or your own implementation, you're better off using the crypto plugin capability described in "The Crypto Plugin Interface" on page 385. At that level all you need to do is unplug the built-in algorithm implementation and plug in your own replacement, which is much simpler than working with the device-level interface.

Creating/Destroying Device Objects

Devices are accessed as device objects that work in the same general manner as other cryptlib objects. You open a connection to a device using **cryptDeviceOpen**, specifying the user who is to own the device object or CRYPT_UNUSED for the default, normal user, the type of device you want to use and the name of the particular device if required or null if there's only one device type possible. This opens a connection to the device. Once you've finished with the device, you use **cryptDeviceClose** to sever the connection and destroy the device object:

```
CRYPT_DEVICE cryptDevice;

cryptDeviceOpen( &cryptDevice, cryptUser, deviceType, deviceName );

/* Use the services provided by the device */

cryptDeviceClose( cryptDevice );
```

The available device types are:

Device	Description
CRYPT_DEVICE_FORTEZZA	Fortezza PCMCIA card.
CRYPT_DEVICE_PKCS11	PKCS #11 crypto token. These devices are accessed via their names, see the

Device	Description
	section on PKCS #11 devices for more details.

Most of the devices are identified implicitly so there's no need to specify a device name and you can pass null as the name parameter (the exception is PKCS #11 devices, which are covered in more detail further on). Once you've finished with the device, you use **cryptDeviceClose** to deactivate it and destroy the device object. For example to work with a Fortezza card you would use:

```
CRYPT_DEVICE cryptDevice;

cryptDeviceOpen( &cryptDevice, cryptUser, CRYPT_DEVICE_FORTEZZA,
    NULL );

/* Use the services provided by the device */

cryptDeviceClose( cryptDevice );
```

If the device can't be accessed, cryptlib will return `CRYPT_ERROR_OPEN` to indicate that it couldn't establish a connection and activate the device. Note that the `CRYPT_DEVICE` is passed to **cryptDeviceOpen** by reference, as it modifies it when it activates the device. In all other routines in cryptlib, `CRYPT_DEVICE` is passed by value.

Some devices have built-in real-time clocks, if cryptlib detects that the device has a built-in clock it'll use the device clock to obtain the time for operations such as creating signed timestamps. Since device clocks can drift over time, cryptlib will perform a consistency check of the device time against the system time and will fall back to using the system time if the device time is too far out of step. In addition the debug build will throw an exception if it detects a problem with the device time.

Activating and Controlling Cryptographic Devices

Once cryptlib has established a connection to the device, you may need to authenticate yourself to it or perform some other control function with it before it will allow itself to be used. You can do this by setting various device attributes, specifying the type of action you want to perform on the device and any additional information that may be required. In the case of user authentication, the additional information will consist of a PIN or password that enables access. Many devices recognise two types of access code, a user-level code that provides standard access (for example for encryption or signing) and a supervisor-level code that provides extended access to device control functions, for example key generation and loading. An example of someone who may require supervisor-level access is a site security officer (SSO) who can load new keys into a device or re-enable its use after a user has been locked out.

Device Initialisation

By setting the `CRYPT_DEVINFO_INITIALISE` attribute, you can initialise the device. This clears keys and other information in the device and prepares it for use. In devices that support supervisor access you need to supply the initialisation or initial supervisor PIN when you call this function:

```
cryptSetAttributeString( cryptDevice, CRYPT_DEVINFO_INITIALISE,
    initialPin, initialPinLength );
```

Once you've initialised the device, you may need to set the supervisor PIN if the device uses a distinct initialisation PIN:

```
cryptSetAttributeString( cryptDevice,
    CRYPT_DEVINFO_SET_AUTHENT_SUPERVISOR, supervisorPin,
    supervisorPinLength );
```

At this point you can carry out device-specific initialisation actions while the device is still in the supervisor state. For example if you're working with a Fortezza card, you would load the CA root (PAA) certificate at this point, since it can only be loaded when the card is first moved into the supervisor-initialised state. Since this is the

ultimately-trusted certificate in the card, it can only be loaded when the card is in this state.

Once you've finished performing any optional further initialisation, you need to set a user PIN, unless the device uses a combined user/supervisor role:

```
cryptSetAttributeString( cryptDevice, CRYPT_DEVINFO_SET_AUTHENT_USER,  
    userPin, userPinLength );
```

Finally, you'll need to log on as a user with the PIN you've just set if the device doesn't do this automatically when you initially set the PIN:

```
cryptSetAttributeString( cryptDevice, CRYPT_DEVINFO_AUTHENT_USER,  
    userPin, userPinLength );
```

The exact initialisation details vary from device to device and driver to driver. Some devices don't distinguish between supervisor and user roles and so only have a single role and PIN. Some devices require a PIN to initialise the device and then set the supervisor PIN using a separate call, others set the supervisor PIN as part of the initialisation call. Some devices will automatically switch over to user mode when you set the user PIN while others require you to explicitly log on in user mode after setting the user PIN. Finally, some devices can't be initialised through PKCS #11 but require proprietary vendor software to initialise them.

When the device is initialised, it usually moves through a number of states going from uninitialised to supervisor initialised to user initialised, with strict restrictions on what can be done in each state. For example once a supervisor has set the user PIN, they can usually no longer change it, since the supervisor isn't supposed to be able to take on the user role and manipulate the device. This is why some devices automatically log the supervisor out once the user PIN has been set. In addition some maintenance operations such as loading initial trusted certificates can only be performed after the device has been initialised and is still in the initial supervisor-initialised state. Again, this prevents modification of trusted keys after the user has been given access to the device.

A general rule of thumb is that when you go through an initialisation you have to perform all of the steps in sequence without logging out in between, and once you've initialised the device you usually can't change any settings without re-initialising it and starting from scratch. Individual devices may diverge from this in places, but in general you shouldn't assume that you can go back later and change things once you've set them.

User Authentication

Before you can use the device you generally need to authenticate yourself to it with a PIN or password. To authenticate yourself as supervisor, set the CRYPT_DEVINFO_AUTHENT_SUPERVISOR attribute; to authenticate yourself as user, set the CRYPT_DEVINFO_AUTHENT_USER attribute. For example to authenticate yourself to the device using a PIN as a normal user you would use:

```
cryptSetAttributeString( cryptDevice, CRYPT_DEVINFO_AUTHENT_USER, pin,  
    pinLength );
```

To authenticate yourself to the device using a PIN for supervisor-level access you would use:

```
cryptSetAttributeString( cryptDevice,  
    CRYPT_DEVINFO_AUTHENT_SUPERVISOR, pin, pinLength );
```

If the PIN or password that you've supplied is incorrect, cryptlib will return CRYPT_ERROR_WRONGKEY. If the device doesn't support this type of access, it will return CRYPT_ERROR_PARAM2. Note that, as is traditional for most PIN and password checking systems, some devices may only allow a limited number of access attempts before locking out the user, requiring CRYPT_DEVINFO_AUTHENT_SUPERVISOR access to re-enable user access.

Device Zeroisation

The `CRYPT_DEVINFO_ZEROISE` attribute works much like `CRYPT_DEVINFO_INITIALISE` except that its specific goal is to clear any sensitive information such as encryption keys from the device (it's often the same as device initialisation, but sometimes will only specifically erase the keys and in some cases may even disable the device). In some devices you may need to supply a zeroisation PIN or the initial supervisor PIN when you call this function, otherwise you should set the data value to an empty string:

```
cryptSetAttributeString( cryptDevice, CRYPT_DEVINFO_ZEROISE, "", 0 );
```

Working with Device Objects

With the device activated and the user authenticated, you can use its cryptographic capabilities in encryption contexts as if it were a standard part of cryptlib. In order to specify the use of the cryptographic device rather than cryptlib's built-in functionality, cryptlib provides the **cryptDeviceCreateContext** and **cryptDeviceQueryCapability** functions that are identical to **cryptCreateContext** and **cryptQueryCapability** but take as an additional argument the handle to the device. For example to create a standard RSA encryption context you would use:

```
cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_RSA );
```

To create an RSA encryption context using an external cryptographic device you would use:

```
cryptDeviceCreateContext( cryptDevice, &cryptContext,
    CRYPT_ALGO_RSA );
```

After this you can use the encryption context as usual, both will function in an identical manner with cryptlib keeping track of whether the implementation is via the built-in functionality or the external device. In this way the use of any form of external hardware for encryption is completely transparent after the initial step of activating and initialising the hardware.

Note that, unlike the other functions that create cryptlib objects, **cryptDeviceCreateContext** doesn't require you to specify the identity of the user who is to own the context which is being created. This is because the device is already associated with a user, so there's no need to specify this again when creating an object within it.

For an example of how you might utilise external hardware, let's use a generic DES/triple DES hardware accelerator (identified by the label "DES/3DES accelerator") accessed as a PKS #11 device. To use the triple DES hardware instead of cryptlib's built-in triple DES implementation you would use:

```
CRYPT_DEVICE cryptDevice;
CRYPT_CONTEXT cryptContext;

/* Activate the DES hardware and create a context in it */
cryptDeviceOpen( &cryptDevice, cryptUser, CRYPT_DEVICE_PKCS11,
    "DES/3DES accelerator" );
cryptDeviceCreateContext( cryptDevice, &cryptContext,
    CRYPT_ALGO_3DES );

/* Generate a key in the DES hardware */
cryptGenerateKey( cryptContext );

/* Encrypt data using the hardware */
cryptEncrypt( cryptContext, data, dataLength );

/* Destroy the context and shut down the DES hardware */
cryptDestroyContext( cryptContext );
cryptDeviceClose( cryptDevice );
```

After the context has been created with **cryptDeviceCreateContext**, the use of the context is identical to a standard encryption context. There is no other (perceptual) difference between the use of a built-in implementation and an external implementation.

Key Storage in Crypto Devices

When you create a normal public-key context and load or generate a key into it, the context goes away when you destroy it or shut down cryptlib. If the context is created in a crypto device, the public and private keys from the context don't go away when the context is destroyed but are stored inside the device for later use. You can later recreate the context using the key stored in the device by treating the device as a keyset containing a stored key. For example to create an RSA key in a device you would use:

```
CRYPT_CONTEXT privKeyContext;

/* Create the RSA context, set a label for the key, and generate a key
into it */
cryptCreateContext( &privKeyContext, cryptUser, CRYPT_ALGO_RSA );
cryptSetAttributeString( privKeyContext, CRYPT_CTXINFO_LABEL, label,
    labelLength );
cryptGenerateKey( privKeyContext );

/* Destroy the context */
cryptDestroyContext( privKeyContext );
```

Although the context has been destroyed, the key itself is still held inside the device. To recreate the context at a later date, you can treat the device as if it were a keyset, using the label as the key ID:

```
CRYPT_CONTEXT privKeyContext;

cryptGetPrivateKey( cryptDevice, &privKeyContext, CRYPT_KEYID_NAME,
    label, NULL );
```

Since you've already authenticated yourself to the device, you don't need to specify a password.

Key storage in crypto devices has additional special considerations that are covered in "Considerations when Working with Devices" on page 355. The most notable of these is that many devices don't allow direct key loads into devices, and virtually all don't allow them to be extracted, so that the key has to be generated inside the device (as the example code given earlier shows) and can't leave the device except (for conventional encryption keys) in encrypted form.

Querying Device Information

Crypto devices come in a wide range of configurations and with varying capabilities, which can include facilities that bypass the normal device-handling operations described here. For example a device may have a built-in keypad or other authentication mechanism that bypasses the need to provide a PIN or password from software. In this case it's not necessary to log in to the device because the login process is handled via an external mechanism. You can determine whether a device is already logged in, or doesn't require a login, by reading the CRYPT_DEVINFO_LOGGEDIN attribute. If this is set to true (any nonzero value) then the device is already logged in, otherwise you need to provide a PIN or password to log in to the device:

```
int deviceLoggedIn;

/* Check whether we're logged in to the device and if not, log in */
cryptGetAttribute( cryptDevice, CRYPT_DEVINFO_LOGGEDIN,
    &deviceLoggedIn );
if( !deviceLoggedIn )
    /* Get PIN from user and log in */;
```

Since some devices represent removable tokens such as smart cards, it's possible for the user to unplug one token and plug in a new one in its place. To help you determine which token was plugged in at the time it was accessed with **cryptDeviceOpen**, you can read the device's CRYPT_DEVINFO_LABEL attribute, which returns the label or name of the token which is accessible via the device:


```

char label[ CRYPT_MAX_TEXTSIZE ];
int labelLength;

cryptGetAttributeString( cryptDevice, CRYPT_DEVINFO_LABEL, label,
    &labelLength );
label[ labelLength ] = '\0';

```

Once you've read the label you can use it to determine whether the required crypto token is available via the device.

Some readers and device interfaces aren't very good at detecting the removal of a crypto token, or the removal of a token and insertion of a new one. For example, many smart card readers only have a simple sensor to detect whether there's something present in the reader, but can't tell whether what's present is the original smart card or a piece of cardboard. In addition some low-level reader drivers can't report the presence (or absence) of a card to the higher-level code. cryptlib will try to contact the crypto token to check whether it's still present and active, but can only go as far as the underlying hardware and software will let it.

Considerations when Working with Devices

There are several considerations to be taken into account when using crypto devices, the major one being that requiring that crypto hardware be present in a system automatically limits the flexibility of your application. There are some cases where the use of certain types of hardware (for example Fortezza cards) may be required, but in many instances the reliance on specialised hardware can be a drawback.

The use of crypto devices can also complicate key management, since keys generated or loaded into the device usually can't be extracted again afterwards. This is a security feature that makes external access to the key impossible, and works in the same way as cryptlib's own storing of keys inside its security perimeter. This means that if you have a crypto device that supports (say) DES and RSA encryption, then to export an encrypted DES key from a context stored in the device, you need to use an RSA context also stored inside the device, since a context located outside the device won't have access to the DES context's key.

Another consideration that needs to be taken into account is the data processing speed of the device. In most cases it's preferable to use cryptlib's built-in implementation of an algorithm rather than the one provided by the device because the built-in implementation will be much faster. For example when hashing data prior to signing it, cryptlib's built-in hashing capabilities should be used in preference to any provided by the device, since cryptlib can process data at the full memory bandwidth using a processor clocked at several gigahertz while a crypto device has to move data over a slow I/O bus to be processed by a processor typically clocked at tens of megahertz or even a few megahertz. In addition when encrypting or decrypting data it's generally preferable to use cryptlib's high-speed encryption capabilities, particularly with devices such as smart cards and to a lesser extent PCMCIA cards, which are severely limited by their slow I/O throughput. As a general rule of thumb, if your system processor is running at 500 MHz or higher then it's always faster to perform the crypto in software rather than using crypto hardware. Because of this it's usual to only perform private-key operations in the crypto device.

A final consideration concerns the limitations of the encryption engine in the device itself. Although cryptlib provides a great deal of flexibility in its software crypto implementations, most hardware devices have only a single encryption engine through which all data must pass (possibly augmented by the ability to store multiple encryption keys in the device). What this means is that each time a different key is used, it has to be loaded into the device's encryption engine before it can be used to encrypt or decrypt data, a potentially time-consuming process. For example if two encryption contexts are created via a device and both are used alternately to encrypt data, the key corresponding to each context has to be loaded by the device into its encryption engine before the encryption can begin (while most devices can store multiple keys, few can keep more than one at a time ready for use in their encryption engine).

As a result of this, although cryptlib will allow you to create as many contexts via a device as the hardware allows, it's generally not a good idea to have more than a single context of each type in use at any one time. For example you could have a single conventional encryption context (using the device's crypto engine), a single digital signature context (using the device's public-key engine), and a single hash context (using the device's CPU or hash engine, or preferably cryptlib itself) active, but not two conventional encryption contexts (which would have to share the encryption engine) or two digital signature contexts (which would have to share the public-key engine).

Fortezza Cards

cryptlib provides complete Fortezza card management capabilities, allowing you to initialise and program a card, generate or load keys into it, add certificates for the generated/loaded keys, update and change PINs, and perform other management functions. This provides full certificate authority workstation (CAW) capabilities.

The steps involved in programming a blank Fortezza card are given in "Activating and Controlling Cryptographic Devices" on page 351. Once the card is in the SSO initialised state (after you've set the SSO PIN), you should install the CA root (PAA) certificate in the card, since this operation is only permitted in the SSO initialised state. The use of PAA certificates is somewhat specific to the use of Fortezza's by the US Government, you may want to simply load a dummy certificate at this point and use standard CA certificates with any keys that you'll be storing on the card.

Note that the Fortezza control firmware requires that all of the steps in the initialisation/programming process be performed in a continuous sequence of operations, without removing the card or closing the device. If you interrupt the process halfway through, you'll need to start again.

After the above programming process has completed, you can generate further keys into the device, load certificates, and so on. This provides the same functionality as a Fortezza CAW.

PKCS #11 Devices

Although most of the devices that cryptlib interfaces with have specialised, single-purpose interfaces, PKCS #11 provides a general-purpose interface that can be used with a wide selection of parameters and in a variety of ways. The following section covers the installation of PKCS #11 modules and documents the way in which cryptlib interfaces to PKCS #11 modules.

Installing New PKCS #11 Modules

You can install new PKCS #11 modules by setting the names of the drivers in cryptlib's configuration database. The module names are specified using the configuration options CRYPT_OPTION_DEVICE_PKCS11_DVR01 ... CRYPT_OPTION_DEVICE_PKCS11_DVR05, cryptlib will step through the list and load each module in turn. Once you've specified the module name, you need to commit the changes in order for cryptlib to use them the next time it's loaded. For example to use the Gemplus GemSAFE driver, you would use:

```
cryptSetAttributeString( CRYPT_UNUSED,
    CRYPT_OPTION_DEVICE_PKCS11_DVR01, "w32pk2ig.dll", 12 );
cryptSetAttribute( CRYPT_UNUSED, CRYPT_OPTION_CONFIGCHANGED, FALSE );
```

The first line of code updates the configuration information to point to the PKCS #11 driver DLL, and the second line makes the changes permanent by flushing the configuration information to disk.

Since the drivers are dynamically loaded on start-up by cryptlib, specifying a driver as a configuration option won't immediately make it available for use. To make the driver available, you have to restart cryptlib or the application using it so that cryptlib can load the driver on start-up, whereupon cryptlib will load the specified modules and make them available as CRYPT_DEVICE_PKCS11 devices. When the module

is loaded, cryptlib will query each module for the device name, this is the name that you should use to access it using **cryptDeviceOpen**.

Some devices don't implement all of their crypto functionality in the device but instead emulate it in software on the host PC. If you have a PKCS #11 module that does then it's better to use cryptlib's native crypto capabilities because they'll be more efficient than those in the driver and possibly more secure as well, depending on how carefully the driver has been written. In order to use only the real device capabilities (rather than those emulated on the host PC), you can set the configuration option `CRYPT_OPTION_DEVICE_PKCS11_HARDWAREONLY` to true (any nonzero value) as explained in "Working with Configuration Options" on page 359. If this option is set, cryptlib will only use capabilities that are provided by the crypto token any not any that are emulated in software.

Accessing PKCS #11 Devices

PKCS #11 devices are identified by the device name, for example the Litronix PKCS #11 driver identifies itself as "Litronix CryptOki Interface" so you would create a device object of this type with:

```
CRYPT_DEVICE cryptDevice;

cryptDeviceOpen( &cryptDevice, cryptUser, CRYPT_DEVICE_PKCS11,
    "Litronix CryptOki Interface" );
```

If you don't know the device name or there's only one device present, you can use the special device name `[Autodetect]` to have cryptlib auto-detect the device for you. If there's more than one device present, cryptlib will use the first one it finds:

```
CRYPT_DEVICE cryptDevice;

cryptDeviceOpen( &cryptDevice, cryptUser, CRYPT_DEVICE_PKCS11,
    "[Autodetect]" );
```

Some PKCS #11 devices allow the use of multiple physical or logical crypto tokens as part of a single device, for example a smart card reader device might have two slots that can each contain a smart card, or the reader itself might function as a crypto token alongside the smart card which is inserted into it. To identify a particular token in a device, you can specify its name after the device name, separated with a double colon. For example if the Litronix reader given in the example above contained two smart cards, you would access the one called "Signing smart card" with:

```
CRYPT_DEVICE cryptDevice;

cryptDeviceOpen( &cryptDevice, cryptUser, CRYPT_DEVICE_PKCS11,
    "Litronix CryptOki Interface::Signing smart card" );
```

Some PKCS #11 devices and drivers have special-case requirements that need to be taken into account when you use them. For example some removable tokens may require special handling for token changes if the reader doesn't support automatic insertion detection, some drivers may have problems if the application forks (under Unix), and so on. You should consult the vendor documentation for the crypto device and drivers that you'll be using to check for any special requirements that you need to meet when you use the device.

CryptoAPI

The following section is intended for forwards-compatibility with future versions of cryptlib. Although some portions of this interface may be implemented, they should not be relied upon in applications.

The CryptoAPI interface provides access to the encryption, signature, and hashing capabilities of the underlying CryptoAPI implementation. All of these facilities are already provided by cryptlib, so it's primary purpose is to provide access to PKCS #12/PFX private keys and certificates held in Windows' internal (proprietary) key store, and by extension keys imported to it from other applications. Using the CryptoAPI interface provides full access to all keys generated by and stored inside

Windows, while still allowing the use of all standard cryptlib functionality and facilities.

Since CryptoAPI is a software implementation managed entirely by the host operating system, there is no need to perform any initialisation, user authentication, or other operations like zeroisation, when using a CryptoAPI device. Initialisation was performed when the operating system was installed, and authentication is performed when the user logs in or the daemon or service that uses the keys is activated. This means that using the CryptoAPI device consists of no more than creating the device object and then utilising it in subsequent crypto operations. All keys and certificates that are accessed through the device will be ones stored in CryptoAPI, giving cryptlib full access to the host operating system's keys and crypto capabilities.

Miscellaneous Topics

This chapter covers various miscellaneous topics not covered in other chapters such as how to obtain information about the encryption capabilities provided by cryptlib, how to obtain information about a particular encryption context, and how to ensure that your code takes advantage of new encryption capabilities provided with future versions of cryptlib.

Querying cryptlib's Capabilities

cryptlib provides two functions to query encryption capabilities, one of which returns information about a given algorithm and mode and the other which returns information on the algorithm and mode used in an encryption context. In both cases the information returned is in the form of a `CRYPT_QUERY_INFO` structure, which is described in “`CRYPT_QUERY_INFO` Structure” on page 403.

You can interrogate cryptlib about the details of a particular encryption algorithm and mode using `cryptQueryCapability`:

```
CRYPT_QUERY_INFO cryptQueryInfo;

cryptQueryCapability( algorithm, &cryptQueryInfo );
```

If you just want to check whether a particular algorithm is available (without obtaining further information on them), you can set the query information parameter to null:

```
cryptQueryCapability( algorithm, NULL );
```

This will simply return a status value without trying to return algorithm information.

Working with Configuration Options

In order to allow extensive control over its security and operational parameters, cryptlib provides a configuration database that can be used to tune its operation for different environments using portable configuration files that function similarly to Unix `.rc` files. This allows cryptlib to be customised on a per-user basis (for example it can remember which key the user usually uses to sign messages and offer to use this key by default), allows a system administrator or manager to set a consistent security policy (for example mandating the use of 1024-or 2048 bit public keys on a company-wide basis instead of unsafe 512-bit keys), and provides information on the use of optional features such as smart card readers, encryption hardware, and cryptographically strong random number generators. The configuration options that affect encryption parameter settings are automatically applied by cryptlib to operations such as key generation and data encryption and signing.

The configuration database can be used to tune the way cryptlib works, with options ranging from algorithms and key sizes through to preferred public/private keys to use for signing and encryption and what to do when certain unusual conditions are encountered. The available options are listed below, with the data type associated with each value being either a boolean (B), numeric (N), or string (S) value:

Value	Type	Description
<code>CRYPT_OPTION_CERT - SIGNUNRECOGNISED-ATTRIBUTES</code>	B	Whether to sign a certificate containing unrecognised attributes. If this option is set to false, the attributes will be omitted from the certificate when it is signed. Default = false.

Value	Type	Description
CRYPT_OPTION_CERT_- COMPLIANCELEVEL	N	The amount of checking for standards-compliance to apply to certificates, certificate requests, and other certificate objects. Default = CRYPT_-COMPLIANCELEVEL_-STANDARD,
CRYPT_OPTION_CERT_- REQUIREPOLICY	B	Whether to require matching certificate policies for certificates in a cert chain once a CA sets a policy. Default = true.
CRYPT_OPTION_CERT_- UPDATEINTERVAL	N	The update interval in days for CRLs. Default = 90.
CRYPT_OPTION_CERT_- VALIDITY	N	The validity period in days for certificates. Default = 365.
CRYPT_OPTION_CMS_- DEFAULTATTRIBUTES CRYPT_OPTION_SMIME_- DEFAULTATTRIBUTES	B	Whether to add the default CMS/S/MIME attributes to signatures (these are alternative names for the same option, since S/MIME uses CMS as the underlying format). Default = true.
CRYPT_OPTION_- CONFIGCHANGED	B	Whether any configuration options have been changed from their original settings (see note below).
CRYPT_OPTION_DEVICE_- PKCS11_DVR01 ... CRYPT_OPTION_DEVICE_- PKCS11_DVR05	S	The module names of any PKCS #11 drivers that cryptlib should load on start-up.
CRYPT_OPTION_DEVICE_- PKCS11_HARDWAREONLY	B	Whether cryptlib should use only the hardware capabilities of the device and not capabilities emulated in software on the host PC by the PKCS #11 driver. Default = false.
CRYPT_OPTION_ENCR_ALGO	N	Encryption algorithm given as a conventional-encryption CRYPT_ALGO_TYPE. Default = CRYPT_ALGO_3DES.
CRYPT_OPTION_ENCR_HASH	N	Hash algorithm given as a hash CRYPT_ALGO_TYPE. Default = CRYPT_ALGO_SHA.
CRYPT_OPTION_ENCR_HASH	N	MAC algorithm given as a MAC CRYPT_ALGO_TYPE. Default = CRYPT_ALGO_HMAC_SHA.
CRYPT_OPTION_INFO_- COPYRIGHT	S	cryptlib copyright notice.

Value	Type	Description
CRYPT_OPTION_INFO_- DESCRIPTION	S	cryptlib description.
CRYPT_OPTION_INFO_- MAJORVERSION	N	cryptlib major and minor version numbers and stepping number.
CRYPT_OPTION_INFO_- MINORVERSION		
CRYPT_OPTION_INFO_- STEPPING		
CRYPT_OPTION_KEYING_ALGO	N	Key processing algorithm given as a hash CRYPT_ALGO_TYPE. Default = CRYPT_ALGO_SHA.
CRYPT_OPTION_KEYING_- ITERATIONS	N	Number of times to iterate the key-processing algorithm. Note that key processing when used for private-key encryption uses a much higher value than this general-purpose value. Default = 500.
CRYPT_OPTION_KEYS_LDAP_- CACERTNAME	S	The names of various LDAP attributes and object classes used for certificate storage/retrieval.
CRYPT_OPTION_KEYS_LDAP_- CERTNAME		
CRYPT_OPTION_KEYS_LDAP_- CRLNAME		
CRYPT_OPTION_KEYS_LDAP_- EMAILNAME		
CRYPT_OPTION_KEYS_LDAP_- FILTER		
CRYPT_OPTION_KEYS_LDAP_- OBJECTCLASS		
CRYPT_OPTION_MISC_- ASYNCINIT	B	Whether to bind in various drivers asynchronously when cryptlib is initialised. This performs the initialisation in a background thread rather than blocking on start-up until the initialisation has completed. Default = true.
CRYPT_OPTION_MISC_- SIDECHANNELPROTECTION	B	Whether to perform additional operations that add protection against some obscure (and rather unlikely) side-channel attacks on private keys. Enabling this option will slow down all private-key operations by up to 10%. Default = false.
CRYPT_OPTION_NET_HTTP_- PROXY	S	HTTP proxy used for accessing web pages. Default = none.
CRYPT_OPTION_NET_SOCKS_- SERVER	S	Socks server and user name used for Internet access. Default = none.
CRYPT_OPTION_NET_SOCKS_- USERNAME		

Value	Type	Description
CRYPT_OPTION_NET_- CONNECTTIMEOUT	N	Timeout in seconds when connecting to a remote server and when transferring data after a connection has been established. Default = 30 seconds for the connect timeout, 0 seconds for the read timeout, 2 seconds for the write timeout.
CRYPT_OPTION_NET_- READTIMEOUT		
CRYPT_OPTION_NET_- WRITETIMEOUT		
CRYPT_OPTION_PKC_ALGO	N	Public-key encryption algorithm given as a public-key CRYPT_ALGO_TYPE. Default = CRYPT_ALGO_RSA.
CRYPT_OPTION_PKC_KEYSIZE	N	Public-key encryption key size in bytes. Default = 128 (1024 bits).
CRYPT_OPTION_SELFTESTOK	N	The current algorithm self-test status (see note below).
CRYPT_OPTION_SIG_ALGO	N	Signature algorithm given as a public-key encryption CRYPT_ALGO_TYPE. Default = CRYPT_ALGO_RSA.
CRYPT_OPTION_SIG_KEYSIZE	N	Signature key size in bytes. Default = 128 (1024 bits).

CRYPT_OPTION_CONFIGCHANGED has special significance in that it contains the current state of the configuration options. If this value is FALSE, the current in-memory configuration options are still set to the same value that they had when cryptlib was started. If set to TRUE, one or more options have been changed and they no longer match the values saved in permanent storage such as a hard disk or flash memory. Writing this value back to FALSE forces the current in-memory values to be committed to permanent storage so that the two match up again.

CRYPT_OPTION_SELFTEST also has special significance, controlling cryptlib's built-in self-test functionality. If you want to perform a self-test of any cryptlib algorithm, you can set this attribute to the algorithm that you want to test. If the self-test succeeds, cryptlib will return an OK status, otherwise it'll return a failure error code. For example to perform the internal self-test of the DSA implementation you'd use:

```
cryptSetAttribute( cryptEnvelope, CRYPT_OPTION_SELFTESTOK,
                  CRYPT_ALGO_DSA );
```

To test all of the implementations, you can set the attribute to CRYPT_USE_DEFAULT. If one (or more) of the algorithm self-tests fails, you can use the per-algorithm test to determine which algorithm(s) failed the self-test.

In addition to these manually-triggered self-tests, cryptlib automatically tests its built-in SHA-1 and DES/3DES implementation and random number generator every time it starts, and won't start if there's a problem with any of them.

Querying/Setting Configuration Options

You can manipulate the configuration options by getting or setting the appropriate attribute values. Since these apply to all of cryptlib rather than to any specific object, you should set the object handle to CRYPT_UNUSED. For example to query the current default encryption algorithm you would use:

```
CRYPT_ALGO_TYPE cryptAlgo;

cryptGetAttribute( CRYPT_UNUSED, CRYPT_OPTION_ENCR_ALGO, &cryptAlgo );
```


To set the default encryption algorithm to CAST-128, you would use:

```
cryptSetAttribute( CRYPT_UNUSED, CRYPT_OPTION_ENCR_ALGO,
                  CRYPT_ALGO_CAST );
```

Some configuration options which contain values that apply to individual objects can also be set for that one object type rather than as a global setting. These options include timeouts for session objects, key size and key setup parameters for encryption contexts, and encryption and hash algorithms for envelopes. For example to set the encryption algorithm to be used when enveloping data in one particular envelope to IDEA, you would use:

```
cryptSetAttribute( cryptEnvelope, CRYPT_OPTION_ENCR_ALGO,
                  CRYPT_ALGO_IDEA );
```

A few of the options are used internally by cryptlib and are read-only (this is indicated in the options' description). These will return CRYPT_ERROR_PERMISSION if you try to modify them to indicate that you don't have permission to change this option.

Saving Configuration Options

The changes you make to the configuration options only last while your program is running or while cryptlib is loaded. In order to make the changes permanent, you can save them to a permanent storage medium such as a hard disk by setting the CRYPT_OPTION_CONFIGCHANGED option to FALSE, indicating that the in-memory settings will be synced to disk so that the two match up. cryptlib will automatically reload the saved options when it starts.

The location of the saved configuration options depend on the system type on which cryptlib is running:

System	Location
BeOS	\$(HOME)/cryptlib/cryptlib.p15
Unix	
DOS	./cryptlib.p15
OS/2	
MVS	CRYPTLIB P15
VM/CMS	
Tandem	\$system.system.cryptlib
Windows 3.x	Windows/cryptlib/cryptlib.p15
Windows 95/-	\Documents and Settings\user_name\Application Data\cryptlib\cryptlib.p15 or \Windows\All Users\Application Data\cryptlib\cryptlib.p15 or \Windows\Profiles\user_name\Application Data\cryptlib.p15 or \Users\user_name\AppData\-
98/ME	
Windows NT/-	
2000/XP/-	
Vista	Roaming (this varies depending on the OS type and version, and is determined by the Windows application data CSIDL)
Windows CE	

Where the operating system supports it, cryptlib will set the security options on the configuration information so that only the person who created it (and, usually, the system administrator) can access it. For example under Unix the file access bits are set to allow only the file owner (and, by extension, the superuser) to access the file, and under Windows NT/2000/XP/Vista with NTFS the file ACLs are set so that only the user who owns it can access or change it.

Obtaining Information About Cryptlib

cryptlib provides a number of read-only configuration options that you can use to obtain information about the version of cryptlib that you're working with.

These options are:

Value	Type	Description
CRYPT_OPTION_INFO_- MAJORVERSION	N	The cryptlib major and minor version numbers and release stepping. For cryptlib 3.1 the major version number is 3 and the minor version number is 1. For beta release 2 the stepping is 2.
CRYPT_OPTION_INFO_- MINORVERSION		
CRYPT_OPTION_INFO_- STEPPING		
CRYPT_OPTION_INFO_- DESCRIPTION	S	A text string containing a description of cryptlib.
CRYPT_OPTION_INFO_- COPYRIGHT	S	The cryptlib copyright notice.

Random Numbers

Several cryptlib functions require access to a source of cryptographically strong random numbers. The random-data-gathering operation is controlled with the **cryptAddRandom** function, which can be used to either inject your own random information into the internal randomness pool or to tell cryptlib to poll the system for random information. To add your own random data (such as keystroke timings when the user enters a password) to the pool, use:

```
cryptAddRandom( buffer, bufferLength );
```

In addition to user-supplied and built-in randomness sources, cryptlib will check for a `/dev/random`, EGD, or PRNGD-style randomness driver (which continually accumulates random data from the system) and will use this as a source of randomness. If running on a system with a hardware random number source (provided by some CPUs and chipsets), cryptlib will also make use of the hardware random number source. cryptlib can also make use of additional entropy seeding information on embedded systems without inherent entropy sources, see “Porting to Devices without Randomness/Entropy Sources” on page 379 for more information.

cryptlib includes in its built-in generator an ANSI X9.17 / ANSI X9.31 generator for FIPS 140 certification purposes. Full technical details of the generator are given in the reference in “Recommended Reading” on page 15.

Gathering Random Information

cryptlib can also gather its own random data by polling the system for random information. There are two polling methods you can use, a fast poll that returns immediately and retrieves a moderate amount of random information, and a slow poll that may take some time but that retrieves much larger amounts of random information. A fast poll is performed with:

```
cryptAddRandom( NULL, CRYPT_RANDOM_FASTPOLL );
```

In general you should sprinkle these throughout your code to build up the amount of randomness in the pool.

A slow poll is performed with:

```
cryptAddRandom( NULL, CRYPT_RANDOM_SLOWPOLL );
```

The effect of this call varies depending on the operating system. Under DOS the call returns immediately (see below). Under Windows 3.x the call will get all the information it can in about a second, then return (there is usually more information present in the system than can be obtained in a second). Under BeOS, OS/2, and on the Macintosh, the call will get all the information it can and then return. Under Unix, Windows 95/98/ME, Windows NT/2000/XP/Vista, and Windows CE the call will spawn one or more separate processes or threads to perform the polling and will return immediately while the poll continues in the background.

Before the first use of a high-level function such as envelopes, secure sessions, or calling **cryptGenerateKey** or **cryptExportKey** you must perform at least one slow

poll (or, in some cases, several fast polls — see below) in order to accumulate enough random information for use by cryptlib. On most systems cryptlib will perform a non-blocking randomness poll, so you can usually do this by calling the slow poll routine when your program starts. This ensures that the random information will have accumulated by the time you need it:

```
/* Program start-up */

cryptAddRandom( NULL, CRYPT_RANDOM_SLOWPOLL );

/* Other code, slow poll runs in the background */

cryptGenerateKey( cryptContext );
```

If you forget to perform a slow poll beforehand, the high-level function will block until the slow poll completes. The fact that the call is blocking is usually fairly obvious, because your program will stop for the duration of the randomness poll. If no reliable random data is available then the high-level function that requires it will return the error `CRYPT_ERROR_RANDOM`.

Obtaining Random Numbers

You can obtain random data from cryptlib by using an encryption context with an algorithm that produces byte-oriented output (for example a block cipher employed in a stream mode like CFB or OFB). To obtain random data, create a context, generate a key into it, and use the context to generate the required quantity of output by encrypting the contents of a buffer. Since the encryption output is random, it doesn't matter what the buffer initially contains. For example you can use the AES algorithm in CFB mode to generate random data with:

```
CRYPT_CONTEXT cryptContext;

cryptCreateContext( &cryptContext, cryptUser, CRYPT_ALGO_AES );
cryptSetAttribute( cryptContext, CRYPT_CTXINFO_MODE, CRYPT_MODE_CFB );
cryptGenerateKey( cryptContext );
cryptEncrypt( cryptContext, randomDataBuffer, randomDataLength );
cryptDestroyContext( cryptContext );
```

This will fill the data buffer with the required number of random bytes.

Working with Newer Versions of cryptlib

Your software can automatically support new encryption algorithms as they are added to cryptlib if you check for the range of supported algorithms instead of hard-coding in the values that existed when you wrote the program. In order to support this, cryptlib predefines the values `CRYPT_ALGO_FIRST_CONVENTIONAL` and `CRYPT_ALGO_LAST_CONVENTIONAL` for the first and last possible conventional encryption algorithms, `CRYPT_ALGO_FIRST_PKC` and `CRYPT_ALGO_LAST_PKC` for the first and last possible public-key encryption algorithms, `CRYPT_ALGO_FIRST_HASH` and `CRYPT_ALGO_LAST_HASH` for the first and last possible hash algorithms, and `CRYPT_ALGO_FIRST_MAC` and `CRYPT_ALGO_LAST_MAC` for the first and last possible MAC algorithms. By checking each possible algorithm value within this range using **cryptQueryCapability**, your software can automatically incorporate any new algorithms as they are added. For example to scan for all available conventional encryption algorithms you would use:

```
CRYPT_ALGO_TYPE cryptAlgo;

for( cryptAlgo = CRYPT_ALGO_FIRST_CONVENTIONAL;
    cryptAlgo <= CRYPT_ALGO_LAST_CONVENTIONAL;
    cryptAlgo++ )
    if( cryptStatusOK( cryptQueryCapability( cryptAlgo, NULL ) ) )
        /* Perform action using algorithm */;
```

The action you would perform would typically be building a list of available algorithms and allowing the user to choose the one they preferred. The same can be done for the public-key, hash, and MAC algorithms.

If your code follows these guidelines, it will automatically handle any new encryption algorithms that are added in newer versions of cryptlib. If you are using the shared library or DLL form of cryptlib, your software's encryption capabilities will be automatically upgraded every time cryptlib is upgraded.

Error Handling

Each function in cryptlib performs extensive parameter and error checking (although monitoring of error codes has been omitted in the code samples for readability). In addition each of the built-in encryption algorithms can perform a self-test procedure that checks the implementation using standard test vectors and methods given with the algorithm specification (typically FIPS publications, ANSI or IETF standards, or standard reference implementations). This self-test is used to verify that each encryption algorithm is performing as required.

The macros `cryptStatusError()` and `cryptStatusOK()` can be used to determine whether a return value denotes an error condition, for example:

```
CRYPT_CONTEXT cryptContext;
int status;

status = cryptCreateContext( &cryptContext, cryptUser,
    CRYPT_ALGO_IDEA );
if( cryptStatusError( status ) )
    /* Perform error processing */;
```

The error codes that can be returned are grouped into a number of classes that cover areas such as function parameter errors, resource errors, and data access errors.

The first group contains a single member, the “no error” value:

Error code	Description
CRYPT_OK	No error.

The next group contains parameter error codes that identify erroneous parameters passed to cryptlib functions:

Error code	Description
CRYPT_ERROR_ - PARAM1...	There is a problem with a parameter passed to a cryptlib function. The exact code depends on the parameter in error.
CRYPT_ERROR_ - PARAM7	

The next group contains resource-related errors such as a certain resource not being available or initialised:

Error code	Description
CRYPT_ERROR_ - FAILED	The operation, for example a public-key encryption or decryption, failed.
CRYPT_ERROR_ - INITED	The object or attribute that you have tried to initialise has already been initialised previously.
CRYPT_ERROR_ - MEMORY	There is not enough memory available to perform this operation.
CRYPT_NOSECURE	cryptlib cannot perform an operation at the requested security level (for example allocated pages can't be locked into memory to prevent them from being swapped to disk, or an LDAP connection can't be established using SSL).
CRYPT_ERROR_ - NOTINITED	The object or attribute that you have tried to use hasn't been initialised yet, or a resource which is required isn't available.
CRYPT_ERROR_ - RANDOM	Not enough random data is available for cryptlib to perform the requested operation.

The next group contains cryptlib security violations such as an attempt to use the wrong object for an operation or to use an object for which you don't have access permission:

Error code	Description
CRYPT_ERROR_-COMPLETE	An operation that consists of multiple steps (such as a message hash) is complete and cannot be continued.
CRYPT_ERROR_-INCOMPLETE	An operation that consists of multiple steps (such as a message hash) is still in progress and requires further steps before it can be regarded as having completed.
CRYPT_ERROR_-INVALID	The public/private key context or certificate object or attribute is invalid for this type of operation.
CRYPT_ERROR_-NOTAVAIL	The requested operation is not available for this object (for example an attempt to load an encryption key into a hash context, or to decrypt a Diffie-Hellman shared integer with an RSA key).
CRYPT_ERROR_-PERMISSION	You don't have permission to perform this type of operation (for example an encrypt-only key being used for a decrypt operation, or an attempt to modify a read-only attribute).
CRYPT_ERROR_-SIGNALLED	<p>An external event such as a signal from a hardware device caused a change in the state of the object. For example if a smart card is removed from a card reader, all the objects that had been loaded or derived from the data on the smart card would return CRYPT_ERROR_-SIGNALLED if you tried to use them.</p> <p>Once an object has entered this state, the only available option is to destroy it, typically using cryptDestroyObject.</p>
CRYPT_ERROR_-TIMEOUT	The operation timed out, either because of a general timeout while accessing an object such as a network connection or data file, or because the object was in use for another operation such as asynchronous key generation or a key database lookup operation.
CRYPT_ERROR_-WRONGKEY	The key being used to decrypt or verify the signature on a piece of data is incorrect.

The next group contains errors related to the higher-level encryption functions such as enveloping, secure session, and key export/import and signature generation/checking functions:

Error code	Description
CRYPT_ERROR_-BADDATA	The data item (typically encrypted or signed data, or a key certificate) was corrupt, or not all of the data was present, and it can't be processed.
CRYPT_ERROR_-OVERFLOW	<p>There is too much data for this function to work with. For an enveloping function, you need to call cryptPopData before you can add any more data to the envelope.</p> <p>For a certificate function this means the amount of data you have supplied is more than what is allowed for the field you are trying to store it in.</p> <p>For a public-key encryption or signature function this means there is too much data for this public/private key to encrypt/sign. You should either use a larger</p>

Error code	Description
	public/private key (in general a 1024-bit or larger key should be sufficient for most purposes) or less data (for example by reducing the key size in the encryption context passed to cryptExportKey).
CRYPT_ERROR_-SIGNATURE	The signature or integrity check value didn't match the data.
CRYPT_ERROR_-UNDERFLOW	There is too little data in the envelope or session for cryptlib to process (for example only a portion of a data item may be present, which isn't enough for cryptlib to work with).

The next group contains data/information access errors, usually arising from keyset, certificate, or device container object accesses:

Error code	Description
CRYPT_ERROR_-DUPLICATE	The given item is already present in the container object.
CRYPT_ERROR_-NOTFOUND	The requested item (for example a key being read from a key database or a certificate component being extracted from a certificate) isn't present in the container object.
CRYPT_ERROR_-OPEN	The container object (for example a keyset or configuration database) couldn't be opened, either because it wasn't found or because the open operation failed.
CRYPT_ERROR_-READ	The requested item couldn't be read from the container object.
CRYPT_ERROR_-WRITE	The item couldn't be written to the container object or the data object couldn't be updated (for example a key couldn't be written to a keyset, or couldn't be deleted from a keyset).

The next group contains errors related to data enveloping:

Error code	Description
CRYPT_ENVELOPE_RESOURCE	A resource such as an encryption key or password needs to be added to the envelope before cryptlib can continue processing the data in it.

Extended Error Reporting

Sometimes the standard cryptlib error codes aren't capable of returning full details on the large variety of possible error conditions that can be encountered. This is particularly true for complex objects such as certificates or ones that are tied to other software or hardware which is outside cryptlib's control. These objects include database or directory keyset objects, crypto devices, and secure sessions. For example if there is a problem checking a certificate object, cryptlib will return a generic CRYPT_ERROR_INVALID status. If there is a missing object attribute that must be set before an object can be used, cryptlib will return a CRYPT_ERROR_NOTINITED status.

In order to obtain more information on the problem you can read the CRYPT_ATTRIBUTE_ERRORLOCUS attribute to obtain the locus of the error (the attribute that caused the problem) and the CRYPT_ATTRIBUTE_ERRORTYPE attribute to identify the type of problem that occurred. These error attributes are present in all objects and can often provide more extensive information on why an operation with the object failed, for example if a function returns CRYPT_ERROR_NOTINITED

then the `CRYPT_ATTRIBUTE_ERRORLOCUS` attribute will tell you which object attribute hasn't been initialised.

The error types are:

Error Type	Description
<code>CRYPT_ERRTYPE_-ATTR_ABSENT</code>	The attribute is required but not present in the object.
<code>CRYPT_ERRTYPE_-ATTR_PRESENT</code>	The attribute is already present in the object, or present but not permitted for this type of object.
<code>CRYPT_ERRTYPE_-ATTR_SIZE</code>	The attribute is smaller than the minimum allowable or larger than the maximum allowable size.
<code>CRYPT_ERRTYPE_-ATTR_VALUE</code>	The attribute is set to an invalid value.
<code>CRYPT_ERRTYPE_-CONSTRAINT</code>	The attribute violates some constraint for the object, or represents a constraint which is being violated, for example a validity period or key usage or certificate policy constraint.
<code>CRYPT_ERRTYPE_-ISSUERCONSTRAINT</code>	The attribute violates a constraint set by an issuer certificate, for example the issuer may set a name constraint which is violated by the certificate object's <code>subjectName</code> or <code>subject altName</code> .

For example to obtain more information on why an attempt to sign a certificate failed you would use:

```
CRYPT_ATTRIBUTE_TYPE errorLocus;
CRYPT_ERRTYPE_TYPE errorType;

status = cryptSignCert( cryptCertificate, cryptCAKey );
if( cryptStatusError( status ) )
{
    cryptGetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_ERRORLOCUS,
        &errorLocus );
    cryptGetAttribute( cryptCertificate, CRYPT_ATTRIBUTE_ERRORTYPE,
        &errorType );
}
```

The error type and locus information comes from `cryptlib` itself, and relates to errors with object usage identified by `cryptlib`. In addition to the `cryptlib` error information, keyset and session objects and objects tied to devices often provide internal error information which is passed to them from the underlying software, hardware, or a remote client or server application. The object-specific error code and message are accessible as the `CRYPT_ATTRIBUTE_INT_ERRORCODE` and `CRYPT_ATTRIBUTE_INT_ERRORMESSAGE` attributes. For example to obtain more information on why an attempt to read a key from an SQL Server database failed you would use:

```
CRYPT_KEYSET cryptKeyset;
CRYPT_HANDLE publicKey
int status;

status = cryptGetPublicKey( &cryptKeyset, &publicKey,
    CRYPT_KEYID_NAME, "John Doe" );
if( cryptStatusError( status ) )
{
    int errorCode, errorStringLength;
    char *errorString;

    errorString = malloc( ... );
    cryptGetAttribute( cryptKeyset, CRYPT_ATTRIBUTE_INT_ERRORCODE,
        &errorCode );
}
```



```
cryptGetAttributeString( cryptKeyset,  
    CRYPT_ATTRIBUTE_INT_ERRORMESSAGE, errorString,  
    &errorStringLength );  
}
```

Note that the error information being returned is passed through by cryptlib from the underlying software or hardware, and will be specific to the implementation. For example if the software that underlies a keyset database is SQL Server then the data returned will be the SQL Server error code and message. Since the returned data is low-level, internal error information coming from the underlying software and will often be information provided by a third-party or remote client or server application, the contents of the error code and message can vary somewhat but the error message will typically contain some indication of what the problem is.

In some cases the access attempt will be blocked by the cryptlib security kernel, and never gets to the object itself. This typically occurs when cryptlib returns a `CRYPT_ERROR_PERMISSION` error, in which the kernel has prevented a disallowed access type. In this case neither the extended error information nor the internal error code and string will be set, since the object never saw the access attempt.

Embedded Systems

cryptlib has been designed to be usable in embedded designs that lack many facilities that are normally found on standard systems, both in terms of resources (memory, network I/O) and in system functionality (a filesystem, dynamic memory allocation). If you're running in a resource-constrained environment such as an embedded system, you first need to decide what cryptlib services you require and disable any unnecessary options, as described in "Customised and Cut-down cryptlib Versions" on page 24. This will reduce the cryptlib code footprint to the minimum required for your particular situation.

As a general rule of thumb if you're on a resource-constrained system you should turn off anything that uses networking, which includes secure sessions (`USE_SESSIONS`), and HTTP and LDAP keyset access (`USE_HTTP`, `USE_LDAP`). You probably also want to turn off crypto devices, (`USE_PKCS11` and `USE_FORTEZZA`), since the embedded system is unlikely to have PKCS #11 crypto hardware attached to it. You probably won't be using database keysets (`USE_DBMS`), and unless you're using PGP keyrings you can turn that off as well (`USE_PGPKEYS`). PGP keyrings are particularly problematic because their structure requires that they be processed via a lookahead buffer because it's not possible to determine how much more data associated with the key is to follow. If you're running in a memory-constrained environment and are thinking of using PGP keys, you should consider using the PKCS #15 format (the cryptlib native keyset type) instead, since this doesn't have this problem.

For envelopes, you probably want to turn off compressed enveloping (`USE_COMPRESSION`) since zlib needs to allocate a series of fairly sizeable buffers in order to operate (256KB for compression, 32KB for decompression, compared to only 8KB used for the envelope buffer itself). If you're not using PGP, you can turn that off as well (`USE_PGP`). Finally, there are a considerable range of other options that you can turn off to save memory and space, see `misc/config.h` for more details.

In addition to the code size tuning and if you're targeting a new embedded system that isn't already supported by cryptlib, you need to make any necessary system-specific adaptations to cryptlib to match the characteristics of the embedded device that you're working with. These adaptations are described following the discussion of supported systems below.

Embedded OS Types

Many embedded OSes, and in particular real-time OSes (RTOSes) are highly modular, and can be heavily customised to suit particular applications. Because of this high degree of customisability, it's not possible for cryptlib to automatically assume that a given OS capability will be available. As a result, the default cryptlib build for a particular embedded OS/RTOS uses a fairly minimal set of OS capabilities that should work in most configurations. If you have extended OS facilities available, you can use the cryptlib configuration file `misc/config.h` to enable any additional capabilities that you may need. It's a good idea to contact the cryptlib developers before you build cryptlib on one of the more modular, configurable embedded OSes like AMX, eCOS, μ C/OS-II, μ ITRON, VxWorks, or XMK. Notes for individual OSes are given below.

AMX

AMX is a highly configurable kernel with most functionality set to the minimum level in the default build in order to conserve space. To run cryptlib you need to enable the time/date manager (so that cryptlib can check timestamps on the data that it's processing), and time-slicing if you're running multiple tasks within cryptlib. For task synchronisation cryptlib uses AMX semaphores, but doesn't require any further AMX facilities like mailboxes, event groups, or buffer pools.

ChorusOS

ChorusOS provides a standard Posix file API and BSD sockets networking API that matches the one used by cryptlib's generic Unix configuration. No special operational considerations are required for cryptlib in this environment.

DOS

DOS isn't strictly speaking an embedded OS but its facilities are limited enough that for cryptlib's purposes it functions as one. Since standard real-mode DOS has very little memory available, you should shrink the object table (via `CONFIG_CONSERVE_MEMORY` and/or `CONFIG_NUM_OBJECTS`) to the smallest size that you can work with. In addition you should disable all unused functionality to conserve as much code and data space as possible. Finally, since DOS has no reliable entropy source, you should use the `CONFIG_RANDSEED` mechanism to enable the use of an external random number seed file.

eCOS

Unlike most other embedded OSes, eCOS requires that all data structures used by the kernel be statically allocated by the user. This means that cryptlib has to allocate storage for all semaphores, mutexes, tasks, and other eCOS objects either at compile time or (at the latest) when it's loaded/initialised. This entails allocating the storage required by eCOS for each object when cryptlib allocates its kernel object table, rather than allocating the storage on-demand when an object is created. If memory is at a premium, you should shrink the object table (via `CONFIG_CONSERVE_MEMORY` and/or `CONFIG_NUM_OBJECTS`) to the smallest size that you can work with, since each object entry has to include space for eCOS kernel data.

Typical eCOS configurations include a full TCP/IP stack and file I/O services. cryptlib uses the Posix section 5/6 file I/O layer, the universal low-level I/O layer that's supported by all filesystem drivers. The TCP/IP stack is a standard BSD-derived stack and its use is enabled by default in the eCOS build.

μC/OS-II

To run cryptlib under μC/OS-II you need to enable mutexes (`OS_MUTEX_EN` and `OS_MUTEX_DEL_EN`) for task synchronisation and tasks (`OS_TASK_CREATE_EN` and `OS_TASK_DEL_EN`) if you're running multiple tasks within cryptlib. μC/OS-II makes a task's priority do double duty as the task ID, so there's no way to uniquely identify a task over the long term. If you change a task's priority using `OSTaskChangePrio()`, you'll also change its task ID. This means that if you've bound a cryptlib object to a task for access control purposes (see "Object Security" on page 42), it'll no longer be accessible once the task priority change changes its task ID. If your tasks change their IDs in this manner, you shouldn't bind objects to particular task IDs.

Embedded Linux

Embedded Linux is a standard Unix environment. No special operational considerations are required for cryptlib in this environment.

μITRON

μITRON has a file interface (ITRON/FILE) derived from the BTRON persistent object store interface, but the only documentation for this is for BTRON and it's only available in Japanese. Because of the inability to obtain either documentation or an implementation to code against, cryptlib only contains stubs for file I/O functionality. If your μITRON system provides this file interface, please contact the cryptlib developers.

μITRON also has a TCP/IP interface, but it doesn't seem to be widely used and the only documentation available is in Japanese. Because of this the use of TCP/IP under μITRON is disabled by default in `misc/config.h`, if you have a μITRON TCP/IP implementation you can use it to replace the existing TCP/IP interface in `io/tcp.c`.

PalmOS

When you install the PalmOS SDK, the include path for the PalmOS compiler may not cover the directory containing the standard ANSI/ISO C headers. These headers are found in the `posix` subdirectory of the standard PalmOS include directory, you can either configure the include path to include this directory or specify it in the makefile with the `-I` compiler option.

If you're building cryptlib using the PalmOS SDK compiler, all compiler warning messages are enabled by default and can't be reset to a more normal level. Because of this maximum warning level, you'll get a stream of compiler messages when you build cryptlib, in particular erroneous used-before-initialised messages. This is normal, and can be ignored.

If you're building cryptlib using the PRC toolchain, the PalmOS headers contain gcc-specific directives that try to pull in gcc headers that lie outside the PalmOS SDK path. If the path to these additional headers isn't configured, you can either configure the include path to include the directories needed by gcc or specify it in the makefile with the `-idirafter` compiler option.

QNX Neutrino

QNX Neutrino is a standard Unix environment, and in general no special operational considerations are required for cryptlib in this environment. The one exception is in the choice of networking environments. QNX Neutrino provides three network stack configurations, the standard TCP/IP stack, the enhanced TCP/IP stack, and a low-resource version of the standard stack. cryptlib works with all of these stacks, and will try and use the most sophisticated features provided by the system. If you're using one of the more restricted networking stacks (for example the tiny TCP/IP stack with no IPv6 support) you may need to change the settings in `io/tcp.h` to reflect this.

RTEMS

RTEMS provides a standard Posix file API and BSD sockets networking API that matches the one used by cryptlib's generic Unix configuration. No special operational considerations are required for cryptlib in this environment.

uClinux

uClinux is an embedded OS intended for use on hardware without memory protection, allowing it to be run on systems that couldn't otherwise run a standard Linux build. To conserve memory, you may want to configure uClinux to use the `page_alloc2/kmalloc2` allocator instead of the somewhat wasteful standard power-of-two Linux allocator, which is intended for use on systems with virtual memory support. cryptlib's memory allocation strategy fits neatly with the `page_alloc2` allocator to minimise memory usage.

By default the uClinux toolchains tend to allocate extremely small stacks of only 4KB, which is inadequate for all but the most trivial applications. To provide an adequate stack, you need to either set `FLTFLAGS=-s stacksize` and export `FLTFLAGS` to the makefile before building your application, or run `flthdrs -s stacksize` on your executable after building it.

Windows CE

Windows CE is a standard Windows environment. No special operational considerations are required for cryptlib in this environment.

VxWorks

VxWorks includes a TCP/IP stack and file I/O services. cryptlib uses the `ioLib` file I/O mechanisms, the universal low-level I/O layer that's supported by all filesystem drivers. The VxWorks TCP/IP stack has changed somewhat over time and is sometimes replaced by more functional third-party alternatives or may not be present at all if VxWorks has been configured without it. Because of this, the use of TCP/IP

services isn't enabled by default. If you need networking services, you can enable them in `misc/config.h`, and may need to perform VxWorks-specific network initialisation (for example calling `selectInit`) if your application doesn't already do so.

Xilinx XMK

XMK is highly configurable kernel with several functions disabled in the default build. To run cryptlib you need to enable mutexes (`config_pthread_mutex`) for thread synchronisation, the yield interface (`config_yield`) for thread management, and timers (`config_time`) for time handling. In addition if you're starting threads within cryptlib, you need to either increase the default thread stack size (`pthread_stack_size`) or set a larger stack size when you start the internal thread.

Xilinx XMK provides an emulated Posix filesystem API, however in order to reduce code size cryptlib uses the native XMK memory filesystem (MFS) interface to access stored data in RAM, ROM, or flash memory. If you need to store data such as configuration options or private keys, you need to enable MFS support in your XMK build.

XMK includes a minimal network stack (LibXilNet), however this only provides server functionality (so it's not possible to implement a network client) and doesn't support timers, so that each send or receive will block forever until data arrives or is sent. Because of these limitations, you need to use a third-party network stack in order to use cryptlib's networking capabilities under XMK.

Embedded cryptlib Configuration Options

You can use the standard cryptlib makefile to cross-compile the code for any of the embedded targets. If you're building for a new target type, you first need to add the new target type at the end of the makefile in the "Embedded Systems" section. The cryptlib naming convention for preprocessor symbols used to identify targets is to use `__target_name__`, which then enables system-specific behaviour in the code. For example if you were adding a new target type to build the code for an Atmel TDMI ARM core, you'd use `-D__ATMEL__` as the necessary compile option (some compilers will define the necessary symbols automatically).

The cryptlib makefile and source code auto-detect various system parameters at compile time, if you're cross-compiling for a new target type that you've defined yourself you'll need to override this so that you're building with the parameters for your target rather than for the host system. In addition you can enable various build options for systems with limited resources as described earlier. The values that you may need to define to handle these system-specific options are:

Option	Description
<code>__target_name__</code>	The target type that you're building for.
<code>CONFIG_LITTLE_ENDIAN</code> <code>CONFIG_BIG_ENDIAN</code>	The CPU endianness of the target system.
<code>CONFIG_CONSERVE_MEMORY</code>	Define if the target has limited memory available to reduce the default sizes of buffers and data structures. "Limited" means less than about 256KB of RAM.
<code>CONFIG_DEBUG_MALLOC</code>	Define to dump memory usage diagnostics to the console. You generally wouldn't use this option on the target system, but only on the host during development.
<code>CONFIG_NO_CERTIFICATES</code>	Define to disable the use of certificate objects. If you define this you also need

Option	Description
	to disable the use of secure sessions, which requires certificates. Some envelope types and keysets that work with certificates will also be affected.
CONFIG_NO_DEVICES	Define to disable the use of crypto device objects.
CONFIG_NO_DYNALLOC	Define to change cryptlib's handling of on-demand memory allocation as described in "Porting to Devices without Dynamic Memory Allocation" on page 377.
CONFIG_NO_ENVELOPES	Define to disable the use of envelope objects. Some secure session types that work with envelopes will also be affected.
CONFIG_NO_ERRORMSG	Don't include long descriptive error messages in the code, which reduces code size.
CONFIG_NO_KEYSETS	Define to disable the use of keyset objects. This also disables the ability to store configuration options to persistent storage, since these are stored in a file keyset. Some secure session and envelope types that work with keysets will also be affected.
CONFIG_NO_SESSIONS	Define to disable the use of secure session objects.
CONFIG_NO_STDIO	Define if the target has no filesystem/stdio support.
CONFIG_NUM_OBJECTS= <i>n</i>	The number of objects that cryptlib reserves room for, defaulting to 1024 without CONFIG_CONSERVE_MEMORY defined or 128 with.
CONFIG_RANDSEED CONFIG_RANDSEED_QUALITY	Define to use external random seed data. Define to set the value of the random seed data, as a percentage figure from 10-100 percent.
CONFIG_SLOW_CPU	Define to disables some of the more CPU-intensive self-tests that are performed on cryptlib startup. The exact definition of a "slow" CPU is somewhat variable, but as a rule of thumb if you're using a 16-bit CPU or one clocked at under 100MHz or so then you probably want to enable this define to speed up the startup process.

Finally, cryptlib includes a considerable amount of other configurability that you can take advantage of if you need to use it in an environment that imposes particular restrictions on resource usage. If you're working with an embedded system, you should contact the cryptlib developers with more details on any specific requirements that you may have.

Once you've got the necessary options set up, you can build the code. If you're building for a completely new target, cryptlib will detect this and print messages at

the various locations in the code where you need to add system-specific adaptations such as support for reading/writing to flash memory segments in `io/file.c`. Alternatively, you can edit `io/file.c` before you try to build the code, look for all the locations where `CONFIG_NO_STDIO` is referenced, and add the necessary support there rather than having cryptlib warn you about it during the build process.

Debugging with Embedded cryptlib

Since you'll be using the same code on your host system as you will in the target, by far the easiest way to develop and debug your application is to do it on the host using your preferred development tools. By enabling the same build options as you would on the target (except for the CPU endianness override) you can exactly duplicate the conditions on the target embedded system and perform all of your application development on the host rather than having to cross-compile, upload code, and work with the target's debugging facilities (if there are any).

Porting to Devices without a Filesystem

If the device you're working with lacks a filesystem, you'll need to work with `io/file.c` to add an adaptation layer to handle the underlying storage abstraction that you're using. In embedded devices this usually consists of blocks of flash memory or occasionally battery-backed RAM, identified either by name/label or an integer value or tag. cryptlib supports the use of named/tagged memory segments if you build it with the `CONFIG_NO_STDIO` option, and will assemble in-memory (RAM) pseudo-files on which it performs all I/O until the file is committed to backing store, whereupon it'll perform an atomic transfer of the pseudo-file to flash to minimise wear on the flash memory cells. It's thus possible to manipulate these (pseudo-)files arbitrarily without causing excessive wear on the underlying storage medium.

Porting to Devices without Dynamic Memory Allocation

If your system lacks dynamic memory allocation, or has so little memory that it's necessary to conserve it as much as possible, you first need to build cryptlib with the `CONFIG_CONSERVE_MEMORY` option. This reduces the default sizes of some buffers, and sets the initial size of cryptlib's internal object table to 128 objects instead of the usual 1024. You can further tune the amount of memory used by the system object table by setting the `CONFIG_NUM_OBJECTS` setting to the maximum number of objects that you'll need. This value must be a power of 2, and can't be less than 8. For single-purpose use in an embedded device (for example when used specifically for enveloping messages rather than as a general-purpose tool where anything is possible), you can usually get by with 32 or even 16 objects. Depending on other options such as whether you use certificate trust settings or not and whether your system has a 16- or 32-bit word size, the cryptlib kernel and built-in system objects consume between 6 and 12 KB of memory.

As a rough rule of thumb, each non-public-key encryption context consumes around 200 bytes (along with any extra memory needed by the algorithm's expanded encryption key), each public-key encryption context consumes around 1500 bytes (depending again on algorithm-specific parameters such as the algorithm type and key size), file keysets (which are buffered in memory as mentioned earlier) consume 600 bytes plus the size of the keyset file (usually around 1.3 KB for a standard 1024-bit RSA key and accompanying certificate and 3 KB for the key and a 3-certificate chain), envelopes consume 1.2KB plus 16 KB for enveloping and 8KB for de-enveloping (the extra size is due to the built-in envelope buffer), and certificates consume an amount of memory that isn't easily predictable in advance since they consist of an arbitrary number of arbitrarily-sized objects. This makes it very difficult to estimate their eventual memory usage, but a rule of thumb is about 2 KB used for a typical certificate. Note that the certificate object consumption has very little to do with the key size, but is mostly dependent on the number and size of all the other X.509 components that are present in the certificate.

Memory Allocation Strategy

cryptlib allocates memory in strict FIFO manner, so that creating an object and then destroying it again rolls back memory to the exact state it was in before the object was created. This ensures that it's possible to run cryptlib on a system without dynamic memory allocation by using a simple high-water-mark pointer that tracks the last memory position used, and falls back to its earlier position when the memory is "freed". Because of this memory usage strategy, cryptlib, although it does acquire memory as required, doesn't need real dynamic memory allocation and can function perfectly well if given a single fixed block of memory and told to use that.

cryptlib allocates either very little or no memory during its normal operation. That is, memory is allocated once at object creation or activation, after which cryptlib stays within the already-allocated bounds unless it encounters some object that it needs to store for later use. For example if it finds a certificate while processing S/MIME data it'll need to acquire a block of memory to store the certificate for later access by the caller.

cryptlib Memory Usage

Almost all of the information that cryptlib processes has the potential to include arbitrary-length data, and occasionally arbitrary amounts of arbitrary-length data. Certificates are a particular example of this, as mentioned earlier. cryptlib's strategy for handling these situations is to use stack memory to handle the data if possible, but if the item being processed exceeds a certain size, to temporarily grab a larger block of memory from the heap to allow the item to be processed, freeing it again immediately after use.

In normal use this overflow handling is never invoked, however since cryptlib can always run into data items of unusual size (constructed either accidentally or maliciously), you need to decide whether you want to allow this behaviour or not. Allowing it means that you can process unusual data items, but may make you vulnerable to deliberate resource-starvation attacks. Conversely, denying it makes you immune to excessive memory usage when trying to process data maliciously constructed to require extra memory to process, but will also make it impossible to process data that just happens to have unusual characteristics. In general, cryptlib will be able to process any normal data without requiring dynamically allocated memory, so if you know in advance which types of data you'll be processing and are concerned about possible resource-starvation attacks, you can disable the opportunistic allocation of larger working areas by using the `CONFIG_NO_DYNALLOC` build option.

cryptlib includes a number of internal lookup tables used for certificate decoding, algorithm information lookup, error parsing, and so on. These are all declared `static const` to tell the compiler to place them in the read-only code segment (held in ROM) rather than the initialised data segment (held in RAM). If your compiler doesn't automatically do this for you (almost all do), you'll need to play with compiler options to ensure that the tables are stored in ROM rather than RAM.

Many cryptlib functions store detailed error information as descriptive text strings that can be retrieved through the `CRYPT_ATTRIBUTE_INT_ERRORMESSAGE` attribute. Since storage for these detailed text messages consumes ROM space, you may want to disable them to save space, or only enable them in the debug build but not the release build. To disable descriptive error messages (only error codes will be returned), define `CONFIG_NO_ERRORMSG`.

Tracking Memory Usage

In order to track memory usage and determine what'll be required on your target system, you can use the `CONFIG_DEBUG_MALLOC` option to dump diagnostics about memory usage to the console. This will allow you to see approximately how much memory a certain operation will require, and let you play with rearranging operations to reduce memory consumption. For example having two objects active simultaneously rather than using them one after the other will result in a total memory

consumption equal to the sum of their sizes rather than only the size of the larger of the two objects.

The memory usage diagnostics will reveal the FIFO nature of the memory allocation that cryptlib uses to try to minimise its overall footprint. You can use the sequence numbers after each allocate and free to track the order in which things are used.

Porting to Devices without Randomness/Entropy Sources

cryptlib requires a source (or more generally multiple sources) of randomness/entropy for the generation of encryption keys and similar data, as described in “Random Numbers” on page 364. On some embedded systems there may not be enough entropy available to safely generate these keys. You can provide this additional entropy yourself through the use of the `CONFIG_RANDOMSEED` option, which enables the use of stored random data that contains additional random seed material. This is stored in the same location as the cryptlib configuration data (see “Working with Configuration Options” on page 359 for more details), and isn’t necessarily a file but can be a block of data in flash memory, data in battery-backed RAM, or whatever other mechanism your system uses for persistent storage. If you define `CONFIG_RANDOMSEED`, cryptlib will try and read the random seed data and use it as additional input to the internal randomness pool. This seed data should be at least 128 bits (16 bytes) long, something like 128 or 256 bytes is a better value. The source of the data is determined by your system configuration, if there’s a file system available it’ll be stored in a file called `randseed.dat`, if not it’ll be accessed via whatever persistent storage mechanism is configured for your system in `io/file.c`. When you build your embedded system, you should install the seed data from an external source, for example a hardware random number generator or a copy of cryptlib running on a secure system with a good source of randomness (the use of cryptlib to generate random data is covered in “Random Numbers” on page 364).

Since a significant portion of the input data for crypto key generation will be determined by the seed data if there are no other randomness sources available (cryptlib will always get at least *some* randomness from the environment, so the value will change each time it’s used), you should take as much care as possible to protect the seed data. Obviously you should use different seed data on each system, to prevent a compromise of one system from affecting any others. In addition if your system provides any protection mechanisms you should apply them to the seed data to try and safeguard it as much as possible. Finally, you should use the ability to add user-supplied randomness described in “Random Numbers” on page 364 to periodically add any situation-specific data that you may have available. For example if your embedded device is being used for voice or video transmission you can add segments of the compressed audio or video data, and if your device performs a sensor/monitoring function you can add the sensor data. Since most embedded devices have at least some interaction with the surrounding environment, there’s usually a source of additional randomness available.

Once you have your seed data set up, you need to decide how much overall randomness it contributes to the system. You can set this value as a percentage between 10 and 100 percent via the `CONFIG_RANDOMSEED_QUALITY` configuration option. If you don’t set a value, cryptlib will assume a figure of 80%, meaning that it needs to obtain an additional 20% of randomness from the environment before it’ll generate keys. Note that this setting is merely a safety level, it doesn’t mean that cryptlib will gather randomness until it reaches 100% and then stop (it never stops gathering randomness), merely that it won’t generate keys when the randomness value is below 100%.

Database and Networking Plugins

In order to communicate with databases that are used as certificate stores and with different network types, cryptlib uses a plugin interface that allows it to talk to any type of database back-end and network protocol. The database plugin provides five functions that are used to interface to the back-end, two functions to open and close the connection to the back-end, two to send data to and read data from it, and one to fetch extended error information if a problem occurs. The plugin typically runs as a Unix daemon which is accessed via an RPC mechanism, however for the ODBC and generic database interfaces the code is compiled directly into cryptlib. If you prefer to have your plugin as part of cryptlib you can compile it in as a generic database interface. The advantage of using an RPC mechanism instead of compiling the plugin code directly into cryptlib is that cryptlib itself (and the machine that cryptlib is running on) don't need to contain any database interface code, since everything can be done on the database server.

The network plugin interface also provides five functions, two to initialise and shut down the connection, two to read and write data, and one to check that the networking interface provided by the interface has been correctly initialised. The network plugin allows cryptlib to use any kind of network interface, either a customised form of the built-in BSD sockets interface or a completely different network mechanism such as SNA or X.25.

The crypto plugin interface is slightly different, and provides direct access to cryptlib's internal encryption capability interface. Replacing a built-in software encryption capability with (say) a hardware crypto core involves unplugging the built-in software implementation and replacing it with the corresponding hardware core interface.

The Database Plugin Interface

The database plugin interface is used when cryptlib receives a user request to access a database of type CRYPT_KEYSET_PLUGIN or CRYPT_KEYSET_PLUGIN_STORE (and by extension for the various CRYPT_KEYSET_ODBC and CRYPT_KEYSET_DATABASE types as well, although these are preconfigured and don't require any further setup). The first thing that cryptlib does is call the `initDbxSession()` function in `keyset/dbms.c`, which connects the generic database type to the actual database plugin (for example an Oracle, Sybase, or PostgreSQL interface). There are three standard plugin types defined, one for ODBC, one for generic built-in databases, and a skeleton generic database network plugin that can communicate with a stub server that talks to the actual database. If you need any other plugin type for a particular database, you can create it as required.

The structure of the plugin is as follows:

```
#include "keyset/keyset.h"

/* Plugin functions: openDatabase(), closeDatabase(), performUpdate(),
   performQuery(), performErrorQuery() */

int initDispatchDatabase( DBMS_INFO *dbmsInfo )
{
    dbmsInfo->openDatabaseBackend = openDatabase;
    dbmsInfo->closeDatabaseBackend = closeDatabase;
    dbmsInfo->performUpdateBackend = performUpdate;
    dbmsInfo->performQueryBackend = performQuery;
    dbmsInfo->performErrorQueryBackend = performErrorQuery;

    return( CRYPT_OK );
}
```

`keyset/keyset.h` contains the keyset-related defines that are used in the code, and the dispatcher initialisation function sets up function pointers to the database access routines, which are explained in more detail below. State information about a session with the database is contained in the `DBMS_STATE_INFO` structure which is

defined in `keyset/keyset.h`. This contains both shared information such as the last error code and the status of the session, and back-end -specific information such as connection handles and temporary data areas. When you create a plugin for a new database type, you should add any variables that you need to the database-specific section of the `DBMS_STATE_INFO` structure. When cryptlib calls your plugin functions, it will pass in the `DBMS_STATE_INFO` that you can use to store state information.

Database Plugin Functions

The database plugin functions that you need to provide are as follows:

```
static int openDatabase( DBMS_STATE_INFO *dbmsInfo, const char *name,
                        const int options, int *featureFlags )
```

This function is called to open a session with the database. The parameters are the name of the database to open the session to and a set of option flags that apply to the session. The name parameter is a composite value that depends on the underlying database being used, usually this is simply the database name, but it can also contain a complete user name and password in the format `user:pass@server`. Other combinations are `user:pass` (only a database user name and password) or `user@server` (only a user name and server).

The option flags will be set to either `CRYPT_KEYOPT_NONE` or `CRYPT_KEYOPT_READONLY`, many servers can optimise accesses if they know that no updates will be performed so your code should try and communicate this to the server if possible. The function should return a set of database feature flags indicating its capabilities in the `featureFlags` parameter. These will be either `DBMS_HAS_BINARYBLOBS` if the database can store binary data blobs rather than requiring that data be base64-encoded, and `DBMS_HAS_NONE` if it has no special capabilities. The plugin should provide binary blob access if the database supports this (almost all do) since this increases data handling efficiency and reduces storage requirements.

```
static void closeDatabase( DBMS_STATE_INFO *dbmsInfo )
```

This function is called to shut down the session with the database.

```
static int performUpdate( DBMS_STATE_INFO *dbmsInfo, const char
                          *command, const void *boundData, const int boundDataLength, const
                          time_t boundDate, const DBMS_UPDATE_TYPE updateType )
```

This function is called to send data to the database. The parameters are an SQL command, optional binary blob data and a date, and an update type indicator that indicates which type of update is being performed. If the `boundData` value is non-null then this parameter and the `boundDataLength` contain a binary blob which is to be added as part of the SQL command. If the `boundDate` value is nonzero then this parameter contains the date and time which is to be added as part of the SQL command as an SQL DATETIME value. For example the function can be called with:

```
performUpdate( ..., "INSERT INTO certificates VALUES ( '...', '...', ...
                  '...' )", NULL, 0, 0 );
performUpdate( ..., "INSERT INTO certificates VALUES ( '...', '...', ... ? )",
                  data, length, 0 );
performUpdate( ..., "INSERT INTO certificates VALUES ( ?, '...', ... ? )",
                  data, length, date );
```

In the first case all data is contained in the SQL command. In the second case there is a binary data blob associated with the SQL command whose position is indicated by the “?” placeholder. After sending the SQL command to the database, you also need to send the (`data`, `length`) value. In the third case there is a binary data blob and a date value associated with the SQL command, with the positions again indicated by the “?” placeholders. The date value is always first in the sequence of placeholders, and the data blob is always second (even if the data blob parameter appears before the date parameter in the list of function parameters). After sending the SQL command to the database, you also need to send the date and then the (`data`, `length`) values. The date value needs to be converted into whatever format the database expects for a

DATETIME value. The exact format depends on the database back-end, which is why it's not present in the SQL command.

The update types are as follows:

Update Type	Description
DBMS_UPDATE_- ABORT	Abort a transaction. This state is communicated to the database through an SQL statement such as ABORT TRANSACTION or ROLLBACK or ABORT, or via a function call that indicates that the transaction begun earlier should be aborted or rolled back.
DBMS_UPDATE_- BEGIN	Begin a transaction. This state is communicated to the database through an SQL statement such as BEGIN TRANSACTION or BEGIN WORK or BEGIN, or via a function call that indicates that transaction semantics are in effect for the following SQL statements.
DBMS_UPDATE_- COMMIT	Commit a transaction. This state is communicated to the database through an SQL statement such as END TRANSACTION or COMMIT WORK or COMMIT, or via a function call that indicates that the transaction should be committed and that transaction semantics are no longer in effect after the statement has been submitted.
DBMS_UPDATE_- CONTINUE	Continue an ongoing transaction.
DBMS_UPDATE_- NORMAL	Standard data update.

The DBMS_UPDATE_BEGIN/CONTINUE/COMMIT combination is used to perform an atomic update on the database. The sequence of calls is as follows:

```
performUpdate( ..., "INSERT INTO certificates VALUES ( ... )",  
              certificate, certLength, certDate, DBMS_UPDATE_BEGIN );  
performUpdate( ..., "INSERT INTO certLog VALUES ( ... )", certificate,  
              certLength, currentDate, DBMS_UPDATE_CONTINUE );  
performUpdate( ..., "DELETE FROM certRequests WHERE keyID = keyID",  
              NULL, 0, 0, DBMS_UPDATE_COMMIT );
```

The first call begins the transaction and submits the initial portion, the ongoing calls submit successive portions of the transaction, and the final call submits the last portion and commits the transaction. If there's a problem, the last call in the transaction will use an update type of DBMS_UPDATE_ABORT. Note that it's important to ensure that performUpdate itself is atomic, for example if there's an error inside the function then it needs to back out of the transaction (if one is in progress) rather than simply returning immediately to the caller. This requires careful tracking of the state of the transaction and handling of error conditions.

```
static int performQuery( DBMS_STATE_INFO *dbmsInfo, const char  
                        *command, char *data, int *dataLength, const char *boundData, const  
                        int boundDataLength, time_t boundDate, const DBMS_CACHEDQUERY_TYPE  
                        queryEntry, const DBMS_QUERY_TYPE queryType )
```

This function is called to fetch data from the database. The parameters are an SQL command, an optional buffer to store the result, optional bound query data and date parameters, a query caching indicator (explained further on) and a query type indicator that indicates which type of query is being performed. The query types are as follows:

Query Type	Description
DBMS_QUERY_- CANCEL	Cancel an ongoing query. This terminates an ongoing query begun by sending a DBMS_QUERY_START query.

Query Type	Description
DBMS_QUERY_- CANCEL	Cancel an ongoing query. This terminates an ongoing query begun by sending a DBMS_QUERY_START query.
DBMS_QUERY_- CHECK	Perform a presence check that simply returns a present/not present indication without returning any data. This allows the query to be optimised since there's no need to actually fetch any data from the back-end. All that's necessary is that a status indication be returned that indicates whether the requested data is available to be fetched or not.
DBMS_QUERY_- CONTINUE	Continue a previous ongoing query. This returns the next entry in the result set generated by sending a DBMS_QUERY_START query.
DBMS_QUERY_- NORMAL	Standard data fetch.
DBMS_QUERY_- START	Begin an ongoing query. This submits a query to the back-end without returning any data. The result set is read one entry at a time by sending DBMS_QUERY_CONTINUE messages.

The DBMS_QUERY_START/CONTINUE/CANCEL combination is used to fetch a collection of entries from the database. The sequence of calls is as follows:

```
performQuery( ..., "SELECT certData FROM certificates WHERE key = ?",
              NULL, NULL, boundData, boundDataLength, 0, DBMS_CACHEDQUERY_NONE,
              DBMS_QUERY_START );

do
    status = performQuery( ..., NULL, buffer, &length, NULL, 0, 0,
                          DBMS_CACHEDQUERY_NONE, DBMS_QUERY_CONTINUE );
while( cryptStatusOK( status ) );
```

The first call submits the query and the ongoing calls fetch successive entries in the result set until an error status is returned (usually this is CRYPT_ERROR_-COMPLETE to indicate that there are no more entries in the result set).

In order to allow for more efficient execution of common queries, cryptlib allows them to be cached by the database back-end for re-use in the future. This allows the back-end to perform the task of SQL parsing and validation against the system catalog, query optimisation, and access plan generation just once when the first query is executed rather than having to re-do it for each query. cryptlib provides hints about cached queries by specifying a cache entry number when it submits the query.

Uncached queries are given an entry number of DBMS_CACHEDQUERY_NONE (these will be little-used query types that it's not worth caching), queries where caching are worthwhile are given an entry number from 1 to 5. The submitted SQL for these queries will never change over subsequent calls, so it's only necessary to perform the parsing and processing once when the query is submitted for the first time. Any subsequent requests can be satisfied using the previously parsed query held at the back-end. In the above example, if the query were submitted with a caching indicator of DBMS_CACHEDQUERY_URI, you could prepare a query for "SELECT certData FROM certificates WHERE uri = ?" the first time that the query is submitted and then re-use the prepared query every time another query with the caching indicator DBMS_CACHEDQUERY_URI is used.

Note that some databases may return a (potentially large) result set in response to a query for a single result using DBMS_QUERY_NORMAL, for example by returning further results after the first one is read or by disallowing further queries until all results have been processed. In this case it will be necessary to limit the query response size either by setting a size limit before submitting the query or by explicitly cancelling a query if more than one result is returned. In addition since cryptlib expects all data to be SQL text strings (or binary data for certificates if the database

supports it) you may need to convert some data types such as integer values to text equivalents when returning them in response to a query.

```
static void performErrorQuery( DBMS_STATE_INFO *dbmsInfo, int
                             *errorCode, char *errorMessage )
```

This function is called to return extended error information when an error occurs. Whenever either `performQuery()` or `performUpdate()` return an error status, this function will be called to obtain further information. The information returned is specific to the database back-end and can include the back-end-specific error code and a text string describing the error. If this information isn't available, you should leave it empty.

An example of a plugin interface is `keyset/odbc.c`, which implements the full functionality required by cryptlib. In addition to the standard functions included below, you may also need to include an SQL rewrite function that changes the contents of SQL queries to match the SQL dialect used by your database. This is a simple function that just substitutes one text string in the query for another. The most common conversion changes the name of the binary blob type (if the database supports it) from the built-in "BLOB" to whatever value is required by the database. Again, see `keyset/odbc.c` for an example of the SQL rewrite process.

The Network Plugin Interface

The network plugin interface is used to provide a transport-layer service to the higher-level cryptlib protocols that require network access capabilities. Network management is handled by the cryptlib I/O streams module `io/stream.c`. The stream I/O system implements a multi-layer architecture with the transport-layer service in the lowest layer, optional I/O buffering layered above that, optional application-layer handling (for example HTTP) above that, and finally the cryptlib protocols such as CMP, RTCS, SCEP, OCSP and TSP above that. Other protocols such as SSH and SSL, which don't require any of the intermediate layers, talk directly to the transport layer.

By replacing the transport-layer interface, you can run cryptlib communications over any type of transport interface. Currently cryptlib provides two types of built-in transport provider, a generic BSD sockets provider and a provider that uses a cryptlib session as the transport layer, making it possible to run (for example) RTCS over SSL or CMP over SSH. You can also use the plugin functionality to provide custom I/O handling that goes beyond that provided by the standard sockets-based interface. For example if you need to use event-based I/O or OS-specific mechanisms such as I/O completion ports, you can provide this capability through the use of custom I/O handlers in the network plugin interface.

The network plugin interface is handled through function pointers to the various transport-layer functions. By setting these to point to functions in the appropriate plugins, it's possible to use any type of networking or communications interface for the transport layer. To set these pointers, the cryptlib I/O stream system calls `setAccessMethodXXX()`, which in the case of BSD sockets is `setAccessMethodTCP()`.

When calling a transport-layer interface function, cryptlib passes in a `STREAM` structure which is defined in `io/stream.h`. This contains information which is required by the transport layer such as socket handles and network communications timeout information.

Network Plugin Functions

The network plugin functions that you need to provide are as follows:

```
static BOOLEAN transportOKFunction( void )
```

This function is called before using any transport-layer functions to query whether the transport layer is OK. It should return `TRUE` if it's safe to call the other transport-layer functions or `FALSE` otherwise, for example because the requested network interface drivers aren't loaded.

```
static int transportConnectFunction( STREAM *stream, const char
    *server, const int port )
```

This function is called to establish a connection, either by connecting to a remote system or by waiting for a connection from a remote system (the exact type depends on whether the stream is acting as a client or server stream). The `server` parameter is the name of the local interface or remote server, and the `port` parameter is the port number to listen on or connect to.

```
static void transportDisconnectFunction( STREAM *stream )
```

This function is called to shut down a connection with a remote client or server.

```
static int transportReadFunction( STREAM *stream, BYTE *buffer, const
    int length, const int flags )
```

This function is called to read data from a remote client or server. The behaviour of this function differs slightly depending on read timeout handling. For blocking reads, it should read as many bytes as are indicated in the `length` parameter, returning an error if less bytes are read. For nonblocking reads it should read as many bytes as are available (which may be zero) and return. In either case if the read succeeds it returns a byte count.

Normally the read should wait for data to appear for the number of seconds indicated by the timeout value stored in the stream I/O structure. However, it's possible to override this with the `flags` parameter, which can contain the following flags:

Flag	Description
TRANSPORT_FLAG_-NONBLOCKING	Perform a nonblocking read, overriding the timeout value in the stream I/O structure if necessary.
TRANSPORT_FLAG_-BLOCKING	Perform a blocking read, overriding the timeout value in the stream I/O structure if necessary.

These flags are used in cases where it's known that a certain number of bytes must be read in order to continue, or when the higher-level stream buffering functions want to perform a speculative read-ahead.

```
static int transportWriteFunction( STREAM *stream, const BYTE *buffer,
    const int length, const int flags )
```

This function is used to write data to a remote client or server. The `flags` parameter is currently unused and should be set to `TRANSPORT_FLAG_NONE`.

The Crypto Plugin Interface

The crypto plugin interface is used to replace or supplement cryptlib's built-in encryption capabilities with external implementations such as crypto hardware or dedicated crypto cores. When cryptlib initialises itself, it calls a sequence of encryption capability initialisation functions declared in `device/system.c`, which return information on each encryption capability available to cryptlib. This capability information is returned in a `CAPABILITY_INFO` structure, defined in `device/capabil.h`, that contains details such as the algorithm name, block size, key size details, and a set of function pointers to the interface for the algorithm. For example one of these would point to the function to load a key (if the algorithm uses keys), one to the function to encrypt data (if it's an encryption algorithm), and so on. For a software implementation, the encryption function would simply constitute the encryption algorithm. For a hardware implementation, the encryption function would pass the data on to the encryption hardware for processing.

The get-capability function is the only externally visible interface to an encryption capability. The easiest way to understand the interface is by looking at an example. If you look in the `context` directory you'll find the implementations for all of cryptlib's built-in capabilities. Take as an example the DES capability implementation, implemented in `context/ctx_des.c`. The `getDESCapability()` function simply returns a pointer to the initialised `CAPABILITY_INFO` structure

containing algorithm information and function pointers for each of the DES capabilities. This is the simplest case, in more sophisticated implementations you could (for example) check for the presence of encryption hardware and return an appropriate `CAPABILITY_INFO` structure for the hardware instead of the software implementation, or vary the algorithm parameters based on what your implementation is capable of. You can even implement completely new (and/or proprietary) algorithms in this manner.

To add support for your own implementation, it's easiest to use one of the `context/ctx_XXX.c` modules as a template for your implementation, and replace the code in the module with your own code or the interface to the crypto hardware or crypto core. If you're replacing a built-in algorithm (rather than adding a new one), you can retain some of the existing functions such as the `selfTest()` function, since these function independently of the underlying implementation.

Algorithms and Standards Conformance

This chapter describes the characteristics of each algorithm used in cryptlib and any known restrictions on their use. Since cryptlib originates in a country that doesn't allow software patents, there are no patent restrictions on the code in its country of origin. Known restrictions in other countries are listed below and all possible care has been taken to ensure that no other infringing technology is incorporated into the code, however since the author is a cryptographer and not an IP lawyer users are urged to consult IP lawyers in the country of intended use if they have any concerns over potential restrictions.

All algorithms, security methods, and data encoding systems used in cryptlib either comply with one or more national and international banking and security standards, or are implemented and tested to conform to a reference implementation of a particular algorithm or security system. Compliance with national and international security standards is automatically provided when cryptlib is integrated into an application. The algorithm standards that cryptlib follows are listed below. A further list of non-algorithm-related standards that cryptlib complies with are given at the start of this document.

AES

AES is a 128-bit block cipher with a 128-bit key and has the cryptlib algorithm identifier `CRYPT_ALGO_AES`.

AES has been implemented as per:

FIPS PUB 197, "Advanced Encryption Standard", 2001.

The AES code has been validated against the test vectors given in:

FIPS PUB 197, "Advanced Encryption Standard", 2001.

Blowfish

Blowfish is a 64-bit block cipher with a 448-bit key and has the cryptlib algorithm identifier `CRYPT_ALGO_BLOWFISH`.

Blowfish has been implemented as per:

"Description of a New Variable-Length Key, 64-bit Block Cipher (Blowfish)", Bruce Schneier, "Fast Software Encryption", *Lecture Notes in Computer Science No. 809*, Springer-Verlag 1994.

The Blowfish modes of operation are given in:

ISO/IEC 8372:1987, "Information Technology — Modes of Operation for a 64-bit Block Cipher Algorithm".

ISO/IEC 10116:1997, "Information technology — Security techniques — Modes of operation for an n-bit block cipher algorithm".

The Blowfish code has been validated against the Blowfish reference implementation test vectors.

CAST-128

CAST-128 is a 64-bit block cipher with a 128-bit key and has the cryptlib algorithm identifier `CRYPT_ALGO_CAST`.

CAST-128 has been implemented as per:

RFC 2144, "The CAST-128 Encryption Algorithm", Carlisle Adams, May 1997.

The CAST-128 modes of operation are given in:

ISO/IEC 8372:1987, "Information Technology — Modes of Operation for a 64-bit Block Cipher Algorithm".

ISO/IEC 10116:1997, “Information technology — Security techniques — Modes of operation for an n-bit block cipher algorithm”.

The CAST-128 code has been validated against the RFC 2144 reference implementation test vectors.

DES

DES is a 64-bit block cipher with a 56-bit key and has the cryptlib algorithm identifier `CRYPT_ALGO_DES`. Note that this algorithm is no longer considered secure and should not be used. It is present in cryptlib only for compatibility with legacy applications.

Although cryptlib uses 64-bit DES keys, only 56 bits of the key are actually used. The least significant bit in each byte is used as a parity bit (cryptlib will set the correct parity values for you, so you don’t have to worry about this). You can treat the algorithm as having a 64-bit key, but bear in mind that only the high 7 bits of each byte are actually used as keying material.

Loading a key will return a `CRYPT_ERROR_PARAM3` error if the key is a weak key. **cryptExportKey** will export the correct parity-adjusted version of the key.

DES has been implemented as per:

ANSI X3.92, “American National Standard, Data Encryption Algorithm”, 1981.

FIPS PUB 46-2, “Data Encryption Standard”, 1994.

FIPS PUB 74, “Guidelines for Implementing and Using the NBS Data Encryption Standard”, 1981.

ISO/IEC 8731:1987, “Banking — Approved Algorithms for Message Authentication — Part 1: Data Encryption Algorithm (DEA)”.

The DES modes of operation are given in:

ANSI X3.106, “American National Standard, Information Systems — Data Encryption Algorithm — Modes of Operation”, 1983.

FIPS PUB 81, “DES Modes of Operation”, 1980.

ISO/IEC 8372:1987, “Information Technology — Modes of Operation for a 64-bit Block Cipher Algorithm”.

ISO/IEC 10116:1997, “Information technology — Security techniques — Modes of operation for an n-bit block cipher algorithm”.

The DES MAC mode is given in:

ANSI X9.9, “Financial Institution Message Authentication (Wholesale)”, 1986.

FIPS PUB 113, “Computer Data Authentication”, 1984.

ISO/IEC 9797:1994, “Information technology — Security techniques — Data integrity mechanism using a cryptographic check function employing a block cipher algorithm”.

The DES code has been validated against the test vectors given in:

NIST Special Publication 500-20, “Validating the Correctness of Hardware Implementations of the NBS Data Encryption Standard”.

Triple DES

Triple DES is a 64-bit block cipher with a 112/168-bit key and has the cryptlib algorithm identifier `CRYPT_ALGO_3DES`.

Although cryptlib uses 128, or 192-bit DES keys (depending on whether two- or three-key triple DES is being used), only 112 or 168 bits of the key are actually used. The least significant bit in each byte is used as a parity bit (cryptlib will set the correct parity values for you, so you don’t have to worry about this). You can treat

the algorithm as having a 128 or 192-bit key, but bear in mind that only the high 7 bits of each byte are actually used as keying material.

Loading a key will return a CRYPT_ERROR_PARAM3 error if the key is a weak key. **cryptExportKey** will export the correct parity-adjusted version of the key.

Triple DES has been implemented as per:

ANSI X9.17, “American National Standard, Financial Institution Key Management (Wholesale)”, 1985.

ANSI X9.52, “Triple Data Encryption Algorithm Modes of Operation”, 1999.

FIPS 46-3, “Data Encryption Standard (DES)”, 1999.

ISO/IEC 8732:1987, “Banking — Key Management (Wholesale)”.

The triple DES modes of operation are given in:

ISO/IEC 8372:1987, “Information Technology — Modes of Operation for a 64-bit Block Cipher Algorithm”.

ISO/IEC 10116:1997, “Information technology — Security techniques — Modes of operation for an n-bit block cipher algorithm”.

The DES code has been validated against the test vectors given in:

NIST Special Publication 800-20, “Modes of Operation Validation System for the Triple Data Encryption Algorithm”.

Diffie-Hellman

Diffie-Hellman is a key exchange algorithm with a key size of up to 4096 bits and has the cryptlib algorithm identifier CRYPT_ALGO_DH.

Diffie-Hellman was formerly covered by a patent in the US, this has now expired.

DH has been implemented as per:

PKCS #3, “Diffie-Hellman Key Agreement Standard”, 1991.

ANSI X9.42, “Public Key Cryptography for the Financial Services Industry — Agreement of Symmetric Keys Using Diffie-Hellman and MQV Algorithms”, 2000.

DSA

DSA is a digital signature algorithm with a key size of up to 1024 bits and has the cryptlib algorithm identifier CRYPT_ALGO_DSA.

DSA is covered by US patent 5,231,668, with the patent held by the US government. This patent has been made available royalty-free to all users world-wide. The US Department of Commerce is not aware of any other patents that would be infringed by the DSA. US patent 4,995,082, “Method for identifying subscribers and for generating and verifying electronic signatures in a data exchange system” (“the Schnorr patent”) relates to the DSA algorithm but only applies to a very restricted set of smart-card based applications and does not affect the DSA implementation in cryptlib.

DSA has been implemented as per:

ANSI X9.30-1, “American National Standard, Public-Key Cryptography Using Irreversible Algorithms for the Financial Services Industry”, 1993.

FIPS PUB 186, “Digital Signature Standard”, 1994.

Elgamal

Elgamal is a public-key encryption/digital signature algorithm with a key size of up to 4096 bits and has the cryptlib algorithm identifier CRYPT_ALGO_ELGAMAL.

Elgamal was formerly covered (indirectly) by a patent in the US, this has now expired.

Elgamal has been implemented as per

“A public-key cryptosystem based on discrete logarithms”, Taher Elgamal, *IEEE Transactions on Information Theory*, **Vol.31, No.4** (1985), p.469.

HMAC-MD5

HMAC-SHA1

HMAC-RIPEMD-160

HMAC-MD5, HMAC-SHA1, and HMAC-RIPEMD-160 are MAC algorithms with a key size of up to 1024 bits and have the cryptlib algorithm identifiers CRYPT_ALGO_HMAC_MD5, CRYPT_ALGO_HMAC_SHA, and CRYPT_ALGO_HMAC_RIPEMD160.

HMAC-MD5 has been implemented as per:

RFC 2104, “HMAC: Keyed-Hashing for Message Authentication”, Hugo Krawczyk, Mihir Bellare, and Ran Canetti, February 1997.

The HMAC-MD5 code has been validated against the test vectors given in:

“Test Cases for HMAC-MD5 and HMAC-SHA-1”, Pau-Chen Cheng and Robert Glenn, March 1997.

HMAC-SHA1 has been implemented as per:

FIPS PUB 198, “The Keyed-Hash Message Authentication Code (HMAC)”, 2002.

RFC 2104, “HMAC: Keyed-Hashing for Message Authentication”, Hugo Krawczyk, Mihir Bellare, and Ran Canetti, February 1997.

The HMAC-SHA1 code has been validated against the test vectors given in:

“Test Cases for HMAC-MD5 and HMAC-SHA-1”, Pau-Chen Cheng and Robert Glenn, March 1997.

IDEA

IDEA is a 64-bit block cipher with a 128-bit key and has the cryptlib algorithm identifier CRYPT_ALGO_IDEA.

IDEA is covered by patents in Austria, France, Germany, Italy, Japan, the Netherlands, Spain, Sweden, Switzerland, the UK, and the US. A statement from the patent owners is included below.

IDEA Patent Notice

IDEA is protected by International copyright law and in addition has been patented in the USA, several countries in Europe (Austria, France, Germany, Italy, Netherlands, Spain, Sweden, Switzerland, United Kingdom), and filed in Japan.

Ascom Systec Ltd., 5506 Mägenwil, Switzerland, holds the patent rights. MediaCrypt AG, 8005 Zurich, Switzerland holds all the relevant rights from Ascom related to the worldwide licensing of the IDEA algorithm.

Any use of the algorithm for Commercial Purposes is subject to a license from MediaCrypt AG and any misuse of the algorithm will be prosecuted.

Commercial Purposes shall mean any revenue generating purpose including but not limited to

- (i) using the algorithm for company internal purposes
- (ii) incorporating an application software containing the algorithm into any hardware and/ or software and distributing such hardware and/or software and/or providing services related thereto to others

- (iii) using a product containing an application software that uses the algorithm not covered by an IDEA license

Free use for private purposes:

The free use of software and/or hardware containing the algorithm is strictly limited to non revenue generating data transfer between private individuals, i.e., not serving commercial purposes. Requests by freeware developers to obtain a royalty-free license to spread an application program containing the algorithm not for commercial purposes must be directed to MediaCrypt.

Special offer for shareware developers:

Selling any software and/or hardware containing the algorithm is subject to a product license. However, there is a special waiver for shareware developers. Such waiver eliminates the up front fees as well as royalties for the first USD 10,000 gross sales of the product containing the algorithm, if and only if:

- 1) The product is being sold for a minimum of USD 10.00 and a maximum of USD 50.00.
- 2) The source code for the shareware product is available to the public. Beyond USD 10,000 gross sales from the shareware product the standard terms and conditions for product licenses shall apply.

IDEA has been implemented as per:

“Device for the Conversion of a Digital Block and the Use Thereof”, James Massey and Xuejia Lai, International Patent PCT/CH91/00117, 1991.

“Device for the Conversion of a Digital Block and Use of Same”, James Massey and Xuejia Lai, US Patent #5,214,703, 1993.

“On the Design and Security of Block Ciphers”, Xuejia Lai, ETH Series in Information Processing, Vol.1, Hartung-Gorre Verlag, 1992.

ISO/IEC 9979, “Data Cryptographic Techniques — Procedures for the Registration of Cryptographic Algorithms”.

The IDEA modes of operation are given in:

ISO/IEC 8372:1987, “Information Technology — Modes of Operation for a 64-bit Block Cipher Algorithm”.

ISO/IEC 10116:1997, “Information technology — Security techniques — Modes of operation for an n-bit block cipher algorithm”.

The IDEA code has been validated against the ETH reference implementation test vectors.

MD2

MD2 is a message digest/hash algorithm with a digest/hash size of 128 bits and has the cryptlib algorithm identifier CRYPT_ALGO_MD2. Although no weaknesses have been found in this algorithm, the algorithm is considered obsolete and should not be used any more except for legacy application support. It is disabled by default.

MD2 has been implemented as per:

RFC 1319, “The MD2 Message Digest Algorithm”, Burt Kaliski, 1992.

The MD2 code has been validated against the RFC 1319 reference implementation test vectors.

MD4

MD4 is a message digest/hash algorithm with a digest/hash size of 128 bits and has the cryptlib algorithm identifier CRYPT_ALGO_MD4. Note that this algorithm is no longer considered secure and should not be used. It is present in cryptlib only for compatibility with legacy applications, and is disabled by default.

MD4 has been implemented as per:

RFC 1320, “The MD4 Message Digest Algorithm”, Ronald Rivest, 1992.

The MD4 code has been validated against the RFC 1320 reference implementation test vectors.

MD5

MD5 is a message digest/hash algorithm with a digest/hash size of 128 bits and has the cryptlib algorithm identifier CRYPT_ALGO_MD5. Note that this algorithm is no longer considered secure and should not be used. It is present in cryptlib only for compatibility with legacy applications.

MD5 has been implemented as per:

RFC 1321, “The MD5 Message Digest Algorithm”, Ronald Rivest, 1992.

The MD5 code has been validated against the RFC 1321 reference implementation test vectors.

RC2

RC2 is a 64-bit block cipher with a 1024-bit key and has the cryptlib algorithm identifier CRYPT_ALGO_RC2. Although no weaknesses have been found in this algorithm, the algorithm is considered obsolete and should not be used any more except for legacy application support. It is disabled by default.

The term “RC2” is trademarked in the US. It may be necessary to refer to it as “an algorithm compatible with RC2” in products that use RC2 and are distributed in the US.

The RC2 code is implemented as per:

“The RC2 Encryption Algorithm”, Ronald Rivest, RSA Data Security Inc, 1992.

RFC 2268, “A Description of the RC2 Encryption Algorithm”, Ronald Rivest, 1998.

The RC2 modes of operation are given in:

ISO/IEC 8372:1987, “Information Technology — Modes of Operation for a 64-bit Block Cipher Algorithm”.

ISO/IEC 10116:1997, “Information technology — Security techniques — Modes of operation for an n-bit block cipher algorithm”.

The RC2 code has been validated against RSADSI BSAFE test vectors.

RC4

RC4 is an 8-bit stream cipher with a key of up to 1024 bits and has the cryptlib algorithm identifier CRYPT_ALGO_RC4. Some weaknesses have been found in this algorithm, and it’s proven to be extremely difficult to employ in a safe manner. For this reason it should not be used any more except for legacy application support, and is disabled by default.

The term “RC4” is trademarked in the US. It may be necessary to refer to it as “an algorithm compatible with RC4” in products that use RC4 and are distributed in the US. Common practice is to refer to it as ArcFour.

The RC4 code is implemented as per:

“The RC4 Encryption Algorithm”, Ronald Rivest, RSA Data Security Inc, 1992.

The RC4 code has been validated against RSADSI BSAFE and US Department of Commerce test vectors.

RC5

RC5 is a 64-bit block cipher with an 832-bit key and has the cryptlib algorithm identifier CRYPT_ALGO_RC5.

RC5 is covered by US patent 5,724,428, “Block Encryption Algorithm with Data-Dependent Rotation”, issued 3 March 1998. The patent is held by RSA Data Security Inc. 100 Marine Parkway, Redwood City, California 94065, ph.+1 415 595-8782, fax +1 415 595-1873, and the algorithm cannot be used commercially in the US without a license.

The RC5 code is implemented as per:

“The RC5 Encryption Algorithm”, Ronald Rivest, “Fast Software Encryption II”, Lecture Notes in Computer Science No.1008, Springer-Verlag 1995.

RFC 2040, “The RC5, RC5-CBC, RC5-CBC-Pad, and RC5-CTS Algorithms”, Robert Baldwin and Ronald Rivest, October 1996.

The RC5 modes of operation are given in:

ISO/IEC 8372:1987, “Information Technology — Modes of Operation for a 64-bit Block Cipher Algorithm”.

ISO/IEC 10116:1997, “Information technology — Security techniques — Modes of operation for an n-bit block cipher algorithm”.

The RC5 code has been validated against the RC5 reference implementation test vectors.

RIPEMD-160

RIPEMD-160 is a message digest/hash algorithm with a digest/hash size of 160 bits and has the cryptlib algorithm identifier CRYPT_ALGO_RIPEMD160.

The RIPEMD-160 code has been implemented as per:

“RIPEMD-160: A strengthened version of RIPEMD”, Hans Dobbertin, Antoon Bosselaers, and Bart Preneel, “Fast Software Encryption III”, *Lecture Notes in Computer Science No.1008*, Springer-Verlag 1995.

ISO/IEC 10118-3:1997, “Information Technology — Security Techniques — Hash functions — Part 3: Dedicated hash functions”.

The RIPEMD-160 code has been validated against the RIPEMD-160 reference implementation test vectors.

RSA

RSA is a public-key encryption/digital signature algorithm with a key size of up to 4096 bits and has the cryptlib algorithm identifier CRYPT_ALGO_RSA.

RSA was formerly covered by a patent in the US, this has now expired.

The RSA code is implemented as per:

ANSI X9.31-1, “American National Standard, Public-Key Cryptography Using Reversible Algorithms for the Financial Services Industry”, 1993.

ISO IEC 9594-8/ITU-T X.509, “Information Technology — Open Systems Interconnection — The Directory: Authentication Framework”.

PKCS #1, “RSA Encryption Standard”, 1991.

SHA

SHA is a message digest/hash algorithm with a digest/hash size of 160 bits and has the cryptlib algorithm identifier CRYPT_ALGO_SHA.

The SHA code has been implemented as per:

ANSI X9.30-2, “American National Standard, Public-Key Cryptography Using Irreversible Algorithms for the Financial Services Industry”, 1993.

FIPS PUB 180, “Secure Hash Standard”, 1993.

FIPS PUB 180-1, “Secure Hash Standard”, 1994.

ISO/IEC 10118-3:1997, “Information Technology — Security Techniques — Hash functions — Part 3: Dedicated hash functions”.

RFC 3174, “US Secure Hash Algorithm 1 (SHA1)”, 2001

The SHA code has been validated against the test vectors given in:

FIPS PUB 180, “Secure Hash Standard”, 1993.

The SHA1 code has been validated against the test vectors given in:

FIPS PUB 180-1, “Secure Hash Standard”, 1994.

SHA2

SHA2 is a message digest/hash algorithm with a digest/hash size of 256 bits and has the cryptlib algorithm identifier CRYPT_ALGO_SHA2.

The SHA2 code has been implemented as per:

FIPS PUB 180-2, “Secure Hash Standard”, 2002.

The SHA2 code has been validated against the test vectors given in:

FIPS PUB 180-2, “Secure Hash Standard”, 2002.

Skipjack

Skipjack is a 64-bit block cipher with an 80-bit key and has the cryptlib algorithm identifier CRYPT_ALGO_SKIPJACK. Although no weaknesses have been found in this algorithm, the algorithm is considered obsolete and should not be used any more except for legacy application support. It is disabled by default.

The Skipjack code has been implemented as per:

“Skipjack and KEA Algorithm Specifications, Version 2.0”, National Security Agency, 28 May 1998.

“Capstone (MYK-80) Specifications”, R21 Informal Technical Report, R21-TECH-30-95, National Security Agency, 14 August 1995.

Data Types and Constants

This section describes the data types and constants used by cryptlib.

CRYPT_ALGO_TYPE

The CRYPT_ALGO_TYPE is used to identify a particular encryption algorithm. More information on the individual algorithm types can be found in “Algorithms” on page 367.

Value	Description
CRYPT_ALGO_AES	AES
CRYPT_ALGO_BLOWFISH	Blowfish
CRYPT_ALGO_CAST	CAST-128
CRYPT_ALGO_DES	DES. This algorithm is no longer considered secure and should not be used except for legacy application support.
CRYPT_ALGO_3DES	Triple DES
CRYPT_ALGO_IDEA	IDEA
CRYPT_ALGO_RC2	RC2. Although no weaknesses have been found in this algorithm, it should not be used any more except for legacy application support.
CRYPT_ALGO_RC4	RC4
CRYPT_ALGO_RC5	RC5
CRYPT_ALGO_SKIPJACK	Skipjack. Although no weaknesses have been found in this algorithm, it should not be used any more except for legacy application support.
CRYPT_ALGO_DH	Diffie-Hellman
CRYPT_ALGO_DSA	DSA
CRYPT_ALGO_ELGAMAL	Elgamal
CRYPT_ALGO_RSA	RSA
CRYPT_ALGO_MD2	MD2. Although no weaknesses have been found in this algorithm, it should not be used any more except for legacy application support.
CRYPT_ALGO_MD4	MD4. This algorithm is no longer considered secure and should not be used except for legacy application support.
CRYPT_ALGO_MD5	MD5. This algorithm is no longer considered secure and should not be used except for legacy application support.

Value	Description
CRYPT_ALGO_RIPEMD160	RIPE-MD 160
CRYPT_ALGO_SHA	SHA/SHA-1
CRYPT_ALGO_SHA2	SHA2/SHA-256/SHA-384/SHA-512
CRYPT_ALGO_HMAC_MD5	HMAC-MD5
CRYPT_ALGO_HMAC_RIPEMD160	HMAC-RIPEMD-160
CRYPT_ALGO_HMAC_SHA	HMAC-SHA
CRYPT_ALGO_VENDOR1 CRYPT_ALGO_VENDOR2 CRYPT_ALGO_VENDOR3	Optional vendor-defined algorithms.
CRYPT_ALGO_FIRST_CONVENTIONAL CRYPT_ALGO_LAST_CONVENTIONAL	First and last possible conventional encryption algorithm type.
CRYPT_ALGO_FIRST_PKC CRYPT_ALGO_LAST_PKC	First and last possible public-key algorithm type.
CRYPT_ALGO_FIRST_HASH CRYPT_ALGO_LAST_HASH	First and last possible hash algorithm type.
CRYPT_ALGO_FIRST_MAC CRYPT_ALGO_LAST_MAC	First and last possible MAC algorithm type.

CRYPT_ATTRIBUTE_TYPE

The CRYPT_ATTRIBUTE_TYPE is used to identify the attribute associated with a cryptlib object. Object attributes are introduced in “Working with Object Attributes” on page 34 and are used extensively throughout this manual.

CRYPT_CERTFORMAT_TYPE

The CRYPT_CERTFORMAT_TYPE is used to specify the format for exported certificate objects. More information on exporting certificate objects is given in “Importing/Exporting Certificates” on page 306.

Value	Description
CRYPT_CERTFORMAT_CERTCHAIN	Certificate object encoded as a PKCS #7 certificate chain. This encoding is only possible for objects that are certificates or certificate chains.
CRYPT_CERTFORMAT_CERTIFICATE	Certificate object encoded according to the ASN.1 distinguished encoding rules (DER).
CRYPT_CERTFORMAT_TEXT_CERTCHAIN CRYPT_CERTFORMAT_TEXT_CERTIFICATE	Base64-encoded text format. The certificate object is encoded as for the basic CRYPT_CERTFORMAT_type format, and an extra layer of base64 encoding with BEGIN/END CERTIFICATE tags is added. This format is required by some web browsers and applications.

CRYPT_CERTTYPE_TYPE

The CRYPT_CERTTYPE_TYPE is used to specify the type of a certificate object when used with **cryptCreateCert**. More information on certificates and certificate objects is given in “Certificates and Certificate Management” on page 234.

Value	Description
CRYPT_CERTTYPE_- ATTRIBUTE_CERT	Attribute certificate.
CRYPT_CERTTYPE_CERTCHAIN	PKCS #7 certificate chain.
CRYPT_CERTTYPE_- CERTIFICATE	Certificate.
CRYPT_CERTTYPE_- CERTREQUEST	PKCS #10 certification request.
CRYPT_CERTTYPE_CMS_- ATTRIBUTES	PKCS #7/CMS attributes.
CRYPT_CERTTYPE_CRL	CRL
CRYPT_CERTTYPE_OCSP_- REQUEST	OCSP request and response.
CRYPT_CERTTYPE_OCSP_- RESPONSE	
CRYPT_CERTTYPE_RTCS_- REQUEST	RTCS request and response.
CRYPT_CERTTYPE_RTCS_- RESPONSE	
CRYPT_CERTTYPE_PKIUSER	PKI user information.
CRYPT_CERTTYPE_REQUEST_- CERT	CRMF certificate request/revocation request.
CRYPT_CERTTYPE_REQUEST_- REVOCATION	

CRYPT_DEVICE_TYPE

The CRYPT_DEVICE_TYPE is used to specify encryption hardware or an encryption device such as a PCMCIA or smart card. More information on encryption devices is given in “Encryption Devices and Modules” on page 350.

Value	Description
CRYPT_DEVICE_FORTEZZA	Fortezza card.
CRYPT_DEVICE_PKCS11	PKCS #11 crypto token.

CRYPT_FORMAT_TYPE

The CRYPT_FORMAT_TYPE is used to identify a data format type for exported keys, signatures, and encryption envelopes. Of the formats supported by cryptlib, the cryptlib native format is the most flexible and is the recommended format unless you require compatibility with a specific security standard. More information on the different formats is given in “Data Enveloping” on page 143, “Exchanging Keys” on page 278, and “Signing Data” on page 284.

Value	Description
CRYPT_FORMAT_CRYPTLIB	cryptlib native format.
CRYPT_FORMAT_PGP	PGP format.

Value	Description
CRYPT_FORMAT_CMS	PKCS #7/CMS format.
CRYPT_FORMAT_PKCS7	
CRYPT_FORMAT_SMIME	As CMS but with S/MIME-specific behaviour.

CRYPT_KEYID_TYPE

The `CRYPT_KEYID_TYPE` is used to identify the type of key identifier which is being passed to `cryptGetPublicKey` or `cryptGetPrivateKey`. More information on using these functions to read keys from keysets is given in “Reading a Key from a Keyset” on page 226

Value	Description
CRYPT_KEYID_NAME	The name of the key owner.
CRYPT_KEYID_EMAIL	The email address of the key owner.

CRYPT_KEYOPT_TYPE

The `CRYPT_KEYOPT_TYPE` is used to contain keyset option flags passed to `cryptKeysetOpen`. The keyset options may be used to optimise access to keysets by enabling cryptlib to perform enhanced transaction management in cases where, for example, read-only access to a database is desired. Because this can improve performance when accessing the keyset, you should always specify whether you will be using the keyset in a restricted access mode when you call `cryptKeysetOpen`. More information on using these options when opening a connection to a keyset is given in “Creating/Destroying Keyset Objects” on page 219

Value	Description
CRYPT_KEYOPT_CREATE	Create a new keyset. This option is only valid for writeable keyset types, which includes keysets implemented as databases and cryptlib key files.
CRYPT_KEYOPT_NONE	No special access options (this option implies read/write access).
CRYPT_KEYOPT_READONLY	Read-only keyset access. This option is automatically enabled by cryptlib for keyset types that have read-only restrictions enforced by the nature of the keyset, the operating system, or user access rights. Unless you specifically require write access to the keyset, you should use this option since it allows cryptlib to optimise its buffering and access strategies for the keyset.

CRYPT_KEYSET_TYPE

The `CRYPT_KEYSET_TYPE` is used to identify a keyset type (or, more specifically, the format and access method used to access a keyset) when used with `cryptKeysetOpen`. Some keyset types may be unavailable on some systems. More information on keyset types is given in “Keyset Types” on page 218.

Value	Description
CRYPT_KEYSET_FILE	A flat-file keyset, either a cryptlib key file or a PGP/OpenPGP key ring.
CRYPT_KEYSET_HTTP	URL specifying the location of a certificate or CRL.
CRYPT_KEYSET_LDAP	LDAP directory service.
CRYPT_KEYSET_PLUGIN	Generic database network plugin.
CRYPT_KEYSET_DATABASE	Generic RDBMS interface.
CRYPT_KEYSET_ODBC	Generic ODBC interface.
CRYPT_KEYSET_DATABASE_STORE	As for the basic keyset types, but representing a certificate store for use by a CA rather than a simple keyset. The user who creates and updates these keyset types must be a CA user.
CRYPT_KEYSET_ODBC_STORE	
CRYPT_KEYSET_PLUGIN_STORE	

CRYPT_MODE_TYPE

The CRYPT_MODE_TYPE is used to identify a particular conventional encryption mode. More information on the individual modes can be found in “Algorithms” on page 367.

Value	Description
CRYPT_MODE_ECB	ECB
CRYPT_MODE_CBC	CBC
CRYPT_MODE_CFB	CFB
CRYPT_MODE_OFB	OFB

CRYPT_OBJECT_TYPE

The CRYPT_OBJECT_TYPE is used to identify the type of an exported key or signature object that has been created with **cryptExportKey** or **cryptCreateSignature**. More information on working with these objects is given in “Querying an Exported Key Object” on page 281, and “Querying a Signature Object” on page 285.

Value	Description
CRYPT_OBJECT_ENCRYPTED_KEY	Conventionally exported key object.
CRYPT_OBJECT_KEYAGREEMENT	Key agreement object.
CRYPT_OBJECT_PKCENCRYPTED_KEY	Public-key exported key object.
CRYPT_OBJECT_SIGNATURE	Signature object.

CRYPT_SESSION_TYPE

The CRYPT_SESSION_TYPE is used to identify a secure session type when used with **cryptCreateSession**. More information on sessions is given in “Secure Sessions” on page 190.

Value	Description
CRYPT_SESSION_CMP	CMP client/server session.
CRYPT_SESSION_CMP_SERVER	

Value	Description
CRYPT_SESSION_CMP CRYPT_SESSION_CMP_SERVER	CMP client/server session.
CRYPT_SESSION_OCSP CRYPT_SESSION_OCSP_SERVER	OCSP client/server session.
CRYPT_SESSION_RTCS CRYPT_SESSION_RTCS_SERVER	RTCS client/server session.
CRYPT_SESSION_SCEP CRYPT_SESSION_SCEP_SERVER	SCEP client/server session.
CRYPT_SESSION_SSH CRYPT_SESSION_SSH_SERVER	SSH client/server session.
CRYPT_SESSION_SSL CRYPT_SESSION_SSL_SERVER	SSL client/server session.
CRYPT_SESSION_TSP CRYPT_SESSION_TSP_SERVER	TSP client/server session.

Data Size Constants

The following values define various maximum lengths for data objects that are used in cryptlib. These can be used for allocating memory to contain the objects, or as a check to ensure that an object isn't larger than the maximum size allowed by cryptlib.

Constant	Description
CRYPT_MAX_HASHSIZE	Maximum hash size in bytes.
CRYPT_MAX_IVSIZE	Maximum initialisation vector size in bytes.
CRYPT_MAX_KEYSIZE	Maximum conventional-encryption key size in bytes.
CRYPT_MAX_PKCSIZE	Maximum public-key component size in bytes. This value specifies the maximum size of individual components, since public/private keys are usually composed of a number of components the overall size is larger than this.
CRYPT_MAX_TEXTSIZE	Maximum size of a text string (e.g. a public or private key owner name) in characters. This defines the string size in characters rather than bytes, so a Unicode string of size CRYPT_MAX_TEXTSIZE could be twice as long as an ASCII string of size CRYPT_MAX_TEXTSIZE. This value does not include the terminating null character in C strings.

Miscellaneous Constants

The following values are used for various purposes by cryptlib, for example to specify that default parameter values are to be used, that the given parameter is unused and can be ignored, or that a special action should be taken in response to seeing this parameter.

Constant	Description
CRYPT_KEYTYPE_PRIVATE CRYPT_KEYTYPE_PUBLIC	Whether the key being passed to <code>cryptInitComponents()</code> / <code>cryptSetComponent()</code> is a

Constant	Description
	public or private key.
CRYPT_RANDOM_FASTPOLL	The type of polling to perform to update the internal random data pool.
CRYPT_RANDOM_SLOWPOLL	
CRYPT_UNUSED	A value indicating that this parameter is unused and can be ignored.
CRYPT_USE_DEFAULT	A value indicating that the default setting for this parameter should be used.

Data Structures

This section describes the data structures used by cryptlib.

CRYPT_OBJECT_INFO Structure

The CRYPT_OBJECT_INFO structure is used with **cryptQueryObject** to return information about a data object created with **cryptExportKey** or **cryptCreateSignature**. Some of the fields are only valid for certain algorithm and mode combinations, or for some types of data objects. If they don't apply to the given algorithm and mode or context, they will be set to CRYPT_ERROR, null, or filled with zeroes as appropriate.

Field	Description
CRYPT_OBJECT_TYPE objectType	Data object type.
CRYPT_ALGO_TYPE cryptAlgo	Encryption/signature algorithm.
CRYPT_MODE_TYPE cryptMode	Encryption/signature mode.
CRYPT_ALGO_TYPE hashAlgo	The hash algorithm used to hash the data if the data object is a signature object.
unsigned char salt[CRYPT_MAX_HASHSIZE] int saltLength	The salt used to derive the export/import key if the object is a conventionally encrypted key object

CRYPT_PKCINFO_xxx Structures

The CRYPT_PKCINFO_xxx structures are used to load public and private keys (which contain multiple key components) into encryption contexts by setting them as the CRYPT_CTXINFO_KEY_COMPONENTS attribute. All fields are multi-precision integer values that are set using the **cryptSetComponent()** macro.

The CRYPT_PKCINFO_DLP structure is used to load keys for algorithms based on the discrete logarithm problem, which includes keys for Diffie-Hellman, DSA, and Elgamal. The structure contains the following fields:

Field	Description
p	Prime modulus.
q	Prime divisor. Some DH and Elgamal keys don't use this parameter, in which case you should set it to an all-zero value of the appropriate size. Note that omitting the q parameter means that cryptlib can't perform certain key validity checks that it otherwise performs when q is present.
g	Element of order q mod p.
x	Private random integer.
y	Public random integer, $g^x \bmod p$.

The CRYPT_PKCINFO_RSA structure is used to load RSA public-key encryption keys and contains the following fields:

Field	Description
n	Modulus.
e	Public exponent.
d	Private exponent. Some keys don't include this parameter, in which case you should set it to an all-zero value of the appropriate size. Note that if the d

Field	Description
	parameter is absent then the e1 and e2 values must be present.
p	Prime factor 1.
q	Prime factor 2.
u	CRT coefficient $q^{-1} \bmod p$.
e1	Private exponent 1 (PKCS #1), $d \bmod (p-1)$.
e2	Private exponent 2 (PKCS #1), $d \bmod (q-1)$.

The e1 and e2 components of CRYPT_PKCINFO_RSA may not be present in some keys. cryptlib will make use of them if they are present, but can also work without them. The loading of private keys is slightly slower if these values aren't present since cryptlib needs to generate them itself.

CRYPT_QUERY_INFO Structure

The CRYPT_QUERY_INFO structure is used with **cryptQueryCapability** to return information about an encryption algorithm or an encryption context or key-related certificate object (for example a public-key certificate or certification request). Some of the fields are only valid for certain algorithm types, or for some types of encryption contexts. If they don't apply to the given algorithm or context, they will be set to CRYPT_ERROR, null, or filled with zeroes as appropriate.

Field	Description
char algoName[CRYPT_MAX_TEXTSIZE]	Algorithm name.
int blockSize	Algorithm block size in bytes.
int minKeySize	The minimum, recommended, and maximum key size in bytes (if the algorithm uses a key).
int keySize	
int maxKeySize	

Function Reference

cryptAddCertExtension

The **cryptAddCertExtension** function is used to add a generic blob-type certificate extension to a certificate object.

```
int cryptAddCertExtension( const CRYPT_CERTIFICATE certificate, const char *oid, const int criticalFlag, const void *extension, const int extensionLength );
```

Parameters	<i>certificate</i>
	The certificate object to which to add the extension.
	<i>oid</i>
	The object identifier value for the extension being added, specified as a sequence of integers.
	<i>criticalFlag</i>
	The critical flag for the extension being added.
	<i>extension</i>
	The address of the extension data.
	<i>extensionLength</i>
	The length in bytes of the extension data.
Remarks	cryptlib directly supports extensions from X.509, PKIX, SET, SigG, and various vendors itself, so you shouldn't use this function for anything other than unknown, proprietary extensions.
See also	cryptGetCertExtension , cryptDeleteCertExtension .

cryptAddPrivateKey

The **cryptAddPrivateKey** function is used to add a user's private key to a keyset.

```
int cryptAddPrivateKey( const CRYPT_KEYSET keyset, const CRYPT_HANDLE cryptKey, const char *password );
```

Parameters	<i>keyset</i>
	The keyset object to which to write the key.
	<i>cryptKey</i>
	The private key to write to the keyset.
	<i>password</i>
	The password used to encrypt the private key.
Remarks	The use of a password to encrypt the private key is required when storing a private key to a keyset, but not to a crypto device such as a smart card or Fortezza card, since these provide their own protection for the key data.
See also	cryptAddPublicKey , cryptDeleteKey , cryptGetPrivateKey , cryptGetPublicKey .

cryptAddPublicKey

The **cryptAddPublicKey** function is used to add a user's public key or certificate to a keyset.

```
int cryptAddPublicKey( const CRYPT_KEYSET keyset, const CRYPT_CERTIFICATE certificate );
```

Parameters	<i>keyset</i>
	The keyset object to which to write the key.

certificate

The certificate to add to the keyset.

Remarks This function requires a key certificate object rather than an encryption context, since the certificate contains additional identification information which is used when the certificate is written to the keyset.

See also `cryptAddPrivateKey`, `cryptDeleteKey`, `cryptGetPrivateKey`, `cryptGetPublicKey`.

cryptAddRandom

The **cryptAddRandom** function is used to add random data to the internal random data pool maintained by cryptlib, or to tell cryptlib to poll the system for random information. The random data pool is used to generate session keys and public/private keys, and by several of the high-level cryptlib functions.

int cryptAddRandom(**const void** *randomData, **const int** randomDataLength);

Parameters *randomData*

The address of the random data to be added, or null if cryptlib should poll the system for random information.

randomDataLength

The length of the random data being added, or CRYPT_RANDOM_SLOWPOLL to perform an in-depth, slow poll or CRYPT_RANDOM_FASTPOLL to perform a less thorough but faster poll for random information.

cryptAsyncCancel

The **cryptAsyncCancel** function is used to cancel an asynchronous operation on an object.

int cryptAsyncCancel(**const CRYPT_HANDLE** cryptObject);

Parameters *cryptObject*

The object on which an asynchronous operation is to be cancelled.

Remarks Because of the asynchronous nature of the operation being performed the cancel may not take effect immediately. In the worst case it may take a second or two for the cancel command to be processed by the object.

See also `cryptAsyncQuery`, `cryptGenerateKeyAsync`.

cryptAsyncQuery

The **cryptAsyncQuery** function is used to obtain the status of an asynchronous operation on an object.

int cryptAsyncQuery(**const CRYPT_HANDLE** cryptObject);

Parameters *cryptObject*

The object to be queried.

Remarks **cryptAsyncQuery** will return CRYPT_ERROR_TIMEOUT if an asynchronous operation is in progress and the object is unavailable for use until the operation completes.

See also `cryptAsyncCancel`, `cryptGenerateKeyAsync`.

cryptCAAddItem

The **cryptCAAddItem** function is used to add a certificate object to a certificate store. Usually this would be a standard certificate, however this function can be used by CAs to add special items such as certificate requests and PKI user information.

```
int cryptCAAddItem( const CRYPT_KEYSET keyset, const CRYPT_CERTIFICATE certificate
);
```

Parameters *keyset*
The certificate store to which the item will be added.

certificate
The item to add to the certificate store.

See also **cryptCACertManagement, cryptCAGetItem.**

cryptCACertManagement

The **cryptCACertManagement** function is used to perform a CA certificate management operation such as a certificate issue, revocation, CRL issue, certificate expiry, or other operation with a certificate store.

```
int cryptCACertManagement( CRYPT_CERTIFICATE *cryptCert, const
CRYPT_CERTACTION_TYPE action, const CRYPT_KEYSET keyset, const
CRYPT_CONTEXT caKey, const CRYPT_CERTIFICATE certRequest );
```

Parameters *cryptCert*
The address of the certificate object to be created.

action
The certificate management operation to perform.

keyset
The certificate store to use to perform the action.

caKey
The CA key to use when performing the action, or CRYPT_UNUSED if no key is necessary for this action.

certRequest
The certificate request to use when performing the action, or CRYPT_UNUSED if no request is necessary for this action.

See also **cryptCAAddItem, cryptCAGetItem.**

cryptCAGetItem

The **cryptCAGetItem** function is used to read a certificate object from a certificate store. Usually this would be a standard certificate, however this function can be used by CAs to obtain special items such as certificate requests and PKI user information. The item to be fetched is identified either through the key owner's name or their email address.

```
int cryptCAGetItem( const CRYPT_KEYSET keyset, CRYPT_CERTIFICATE *certificate, const
CRYPT_CERTTYPE_TYPE certType, const CRYPT_KEYID_TYPE
keyIDtype, const void *keyID );
```

Parameters *keyset*
The certificate store from which to obtain the item.

certificate
The address of the certificate object to be fetched.

certType
The item type.

keyIDtype
The type of the key ID, either CRYPT_KEYID_NAME for the name or key label, or CRYPT_KEYID_EMAIL for the email address.

keyID
The key ID of the item to read.

See also cryptCACertManagement, cryptCAAddItem.

cryptCheckCert

The **cryptCheckCert** function is used to check the signature on a certificate object, or to verify a certificate object against a CRL or a keyset containing a CRL.

```
int cryptCheckCert( const CRYPT_CERTIFICATE certificate, const CRYPT_HANDLE
sigCheckKey );
```

Parameters *certificate*

The certificate container object that contains the certificate item to check.

sigCheckKey

A public-key context or key certificate object containing the public key used to verify the signature, or alternatively CRYPT_UNUSED if the certificate item is self-signed. If the certificate is to be verified against a CRL, this should be a certificate object or keyset containing the CRL. If the certificate is to be verified online, this should be a session object for the server used to verify the certificate.

See also cryptSignCert.

cryptCheckSignature

The **cryptCheckSignature** function is used to check the digital signature on a piece of data.

```
int cryptCheckSignature( const void *signature, const int signatureLength, const
CRYPT_HANDLE sigCheckKey, const CRYPT_CONTEXT hashContext );
```

Parameters *signature*

The address of a buffer that contains the signature.

signatureLength

The length in bytes of the signature data.

sigCheckKey

A public-key context or key certificate object containing the public key used to verify the signature.

hashContext

A hash context containing the hash of the data.

See also cryptCheckSignatureEx, cryptCreateSignature, cryptCreateSignatureEx, cryptQueryObject.

cryptCheckSignatureEx

The **cryptCheckSignatureEx** function is used to check the digital signature on a piece of data with extended control over the signature information.

```
int cryptCheckSignatureEx( const void *signature, const int signatureLength, const
CRYPT_HANDLE sigCheckKey, const CRYPT_CONTEXT hashContext,
CRYPT_HANDLE *extraData );
```

Parameters *signature*

The address of a buffer that contains the signature.

signatureLength

The length in bytes of the signature data.

sigCheckKey

A public-key context or key certificate object containing the public key used to verify the signature.

hashContext

A hash context containing the hash of the data.

extraData

The address of a certificate object containing extra information which is included with the signature, or null if you don't require this information.

See also `cryptCheckSignature`, `cryptCreateSignature`, `cryptCreateSignatureEx`, `cryptQueryObject`.

cryptCreateCert

The **cryptCreateCert** function is used to create a certificate object that contains a certificate, certification request, certificate chain, CRL, or other certificate-like object.

```
int cryptCreateCert( CRYPT_CERTIFICATE *cryptCert, const CRYPT_USER cryptUser, const CRYPT_CERTTYPE_TYPE certType );
```

Parameters *cryptCert*

The address of the certificate object to be created.

cryptUser

The user who is to own the certificate object or CRYPT_UNUSED for the default, normal user.

certType

The type of certificate item that will be created in the certificate object.

See also `cryptDestroyCert`.

cryptCreateContext

The **cryptCreateContext** function is used to create an encryption context for a given encryption algorithm.

```
int cryptCreateContext( CRYPT_CONTEXT *cryptContext, const CRYPT_USER cryptUser, const CRYPT_ALGO_TYPE cryptAlgo );
```

Parameters *cryptContext*

The address of the encryption context to be created.

cryptUser

The user who is to own the encryption context or CRYPT_UNUSED for the default, normal user.

cryptAlgo

The encryption algorithm to be used in the context.

See also `cryptDestroyContext`, `cryptDeviceCreateContext`.

cryptCreateEnvelope

The **cryptCreateEnvelope** function is used to create an envelope object for encrypting or decrypting, signing or signature checking, compressing or decompressing, or otherwise processing data.

```
int cryptCreateEnvelope( CRYPT_ENVELOPE *cryptEnvelope, const CRYPT_USER cryptUser, const CRYPT_FORMAT_TYPE formatType );
```

Parameters *cryptEnvelope*

The address of the envelope to be created.

cryptUser

The user who is to own the envelope object or CRYPT_UNUSED for the default, normal user.

formatType

The data format for the enveloped data.

See also `cryptDestroyEnvelope`.

cryptCreateSession

The **cryptCreateSession** function is used to create a secure session object for use in securing a communications link or otherwise communicating with a remote server or client.

```
int cryptCreateSession( CRYPT_SESSION *cryptSession, const CRYPT_USER cryptUser, const
    CRYPT_SESSION_TYPE sessionType );
```

Parameters *cryptSession*
The address of the session to be created.

cryptUser
The user who is to own the session object or CRYPT_UNUSED for the default, normal user.

sessionType
The type of the secure session.

See also `cryptDestroySession`.

cryptCreateSignature

The **cryptCreateSignature** function digitally signs a piece of data. The signature is placed in a buffer in a portable format that allows it to be checked using **cryptCheckSignature**.

```
int cryptCreateSignature( void *signature, const int signatureMaxLength, int *signatureLength,
    const CRYPT_CONTEXT signContext, const CRYPT_CONTEXT hashContext
    );
```

Parameters *signature*
The address of a buffer to contain the signature. If you set this parameter to null, **cryptCreateSignature** will return the length of the signature in *signatureLength* without actually generating the signature.

signatureMaxLength
The maximum size in bytes of the buffer to contain the signature data.

signatureLength
The address of the signature length.

signContext
A public-key encryption or signature context containing the private key used to sign the data.

hashContext
A hash context containing the hash of the data to sign.

See also `cryptCheckSignature`, `cryptCheckSignatureEx`, `cryptCreateSignatureEx`, `cryptQueryObject`.

cryptCreateSignatureEx

The **cryptCreateSignatureEx** function digitally signs a piece of data with extended control over the signature format. The signature is placed in a buffer in a portable format that allows it to be checked using **cryptCheckSignatureEx**.

```
int cryptCreateSignatureEx( void *signature, const int signatureMaxLength, int *signatureLength,
    const CRYPT_FORMAT_TYPE formatType, const CRYPT_CONTEXT
```

```
signContext, const CRYPT_CONTEXT hashContext, const
CRYPT_CERTIFICATE extraData );
```

- Parameters**
- signature*
The address of a buffer to contain the signature. If you set this parameter to null, **cryptCreateSignature** will return the length of the signature in *signatureLength* without actually generating the signature.
 - signatureMaxLength*
The maximum size in bytes of the buffer to contain the signature data.
 - signatureLength*
The address of the signature length.
 - formatType*
The format of the signature to create.
 - signContext*
A public-key encryption or signature context containing the private key used to sign the data.
 - hashContext*
A hash context containing the hash of the data to sign.
 - extraData*
Extra information to include with the signature or CRYPT_UNUSED if the format is the default signature format (which doesn't use the extra data) or CRYPT_USE_DEFAULT if the signature isn't the default format and you want to use the default extra information.
- See also** **cryptCheckSignature**, **cryptCheckSignatureEx**, **cryptCreateSignature**, **cryptQueryObject**.
-

cryptDecrypt

The **cryptDecrypt** function is used to decrypt or hash data.

```
int cryptDecrypt( const CRYPT_CONTEXT cryptContext, void *buffer, const int length );
```

- Parameters**
- cryptContext*
The encryption context to use to decrypt or hash the data.
 - buffer*
The address of the data to be decrypted or hashed.
 - length*
The length in bytes of the data to be decrypted or hashed.
- Remarks** Public-key encryption and signature algorithms have special data formatting requirements that need to be taken into account when this function is called. You shouldn't use this function with these algorithm types, but instead should use the higher-level functions **cryptCreateSignature**, **cryptCheckSignature**, **cryptExportKey**, and **cryptImportKey**.
- See also** **cryptEncrypt**.
-

cryptDeleteAttribute

The **cryptDeleteAttribute** function is used to delete an attribute from an object.

```
int cryptDeleteAttribute( const CRYPT_HANDLE cryptObject, const
CRYPT_ATTRIBUTE_TYPE attributeType );
```

- Parameters**
- certificate*
The object from which to delete the attribute.

attributeType

The attribute to delete.

Remarks Most attributes are always present and can't be deleted, in general only certificate attributes are deletable.

See also `cryptGetAttribute`, `cryptGetAttributeString`, `cryptSetAttribute`, `cryptSetAttributeString`.

cryptDeleteCertExtension

The **cryptDeleteCertExtension** function is used to delete a generic blob-type certificate extension from a certificate object.

```
int cryptDeleteCertExtension( const CRYPT_CERTIFICATE certificate, const char *oid );
```

Parameters *certificate*

The certificate object from which to delete the extension.

oid

The object identifier value for the extension being deleted, specified as a sequence of integers.

Remarks cryptlib directly supports extensions from X.509, PKIX, SET, SigG, and various vendors itself, so you shouldn't use this function for anything other than unknown, proprietary extensions.

See also `cryptAddCertExtension`, `cryptGetCertExtension`.

cryptDeleteKey

The **cryptDeleteKey** function is used to delete a key or certificate from a keyset or device. The key to delete is identified either through the key owner's name or their email address.

```
int cryptDeleteKey( const CRYPT_HANDLE cryptObject, const CRYPT_KEYID_TYPE
keyIDtype, const void *keyID );
```

Parameters *cryptObject*

The keyset or device object from which to delete the key.

keyIDtype

The type of the key ID, either `CRYPT_KEYID_NAME` for the name or key label, or `CRYPT_KEYID_EMAIL` for the email address.

keyID

The key ID of the key to delete.

See also `cryptAddPrivateKey`, `cryptAddPublicKey`, `cryptGetPrivateKey`, `cryptGetPublicKey`.

cryptDestroyCert

The **cryptDestroyCert** function is used to destroy a certificate object after use. This erases all keying and security information used by the object and frees up any memory it uses.

```
int cryptDestroyCert( const CRYPT_CERTIFICATE cryptCert );
```

Parameters *cryptCert*

The certificate object to be destroyed.

See also `cryptCreateCert`.

cryptDestroyContext

The **cryptDestroyContext** function is used to destroy an encryption context after use. This erases all keying and security information used by the context and frees up any memory it uses.

```
int cryptDestroyContext( const CRYPT_CONTEXT cryptContext );
```

Parameters *cryptContext*
The encryption context to be destroyed.

See also **cryptCreateContext**, **cryptDeviceCreateContext**.

cryptDestroyEnvelope

The **cryptDestroyEnvelope** function is used to destroy an envelope after use. This erases all keying and security information used by the envelope and frees up any memory it uses.

```
int cryptDestroyEnvelope( const CRYPT_ENVELOPE cryptEnvelope );
```

Parameters *cryptEnvelope*
The envelope to be destroyed.

See also **cryptCreateEnvelope**.

cryptDestroyObject

The **cryptDestroyObject** function is used to destroy a cryptlib object after use. This erases all security information used by the object, closes any open data sources, and frees up any memory it uses.

```
int cryptDestroyObject( const CRYPT_HANDLE cryptObject );
```

Parameters *cryptObject*
The object to be destroyed.

Remarks This function is a generic form of the specialised functions that destroy/close specific cryptlib object types such as encryption contexts and certificate and keyset objects. In some cases it may not be possible to determine the exact type of an object (for example the keyset access functions may return a key certificate object or only an encryption context depending on the keyset type), **cryptDestroyObject** can be used to destroy an object of an unknown type.

See also **cryptDestroyContext**, **cryptDestroyCert**, **cryptDestroyEnvelope**, **cryptDestroySession**, **cryptKeysetClose**.

cryptDestroySession

The **cryptDestroySession** function is used to destroy a session object after use. This close the link to the client or server, erases all keying and security information used by the session, and frees up any memory it uses.

```
int cryptDestroySession( const CRYPT_SESSION cryptSession );
```

Parameters *cryptSession*
The session to be destroyed.

See also **cryptCreateSession**.

cryptDeviceClose

The **cryptDeviceClose** function is used to destroy a device object after use. This closes the connection to the device and frees up any memory it uses.

```
int cryptDeviceClose( const CRYPT_DEVICE device );
```

Parameters *device*
The device object to be destroyed.

See also `cryptDeviceOpen`.

cryptDeviceCreateContext

The `cryptDeviceCreateContext` function is used to create an encryption context for a given encryption algorithm via an encryption device.

```
int cryptDeviceCreateContext( const CRYPT_DEVICE cryptDevice, CRYPT_CONTEXT
                             *cryptContext, const CRYPT_ALGO_TYPE cryptAlgo );
```

Parameters *cryptDevice*
The device object used to create the encryption context.

cryptContext
The address of the encryption context to be created.

cryptAlgo
The encryption algorithm to be used in the context.

See also `cryptCreateContext`, `cryptDestroyContext`.

cryptDeviceOpen

The `cryptDeviceOpen` function is used to establish a connection to a crypto device such as a crypto hardware accelerator or a PCMCIA card or smart card.

```
int cryptDeviceOpen( CRYPT_DEVICE *device, const CRYPT_USER cryptUser, const
                    CRYPT_DEVICE_TYPE deviceType, const char *name );
```

Parameters *device*
The address of the device object to be created.

cryptUser
The user who is to own the device object or `CRYPT_UNUSED` for the default, normal user.

deviceType
The device type to be used.

name
The name of the device, or null if a name isn't required.

See also `cryptDeviceClose`.

cryptDeviceQueryCapability

The `cryptDeviceQueryCapability` function is used to obtain information about the characteristics of a particular encryption algorithm provided by an encryption device. The information returned covers the algorithm's key size, data block size, and other algorithm-specific information.

```
int cryptDeviceQueryCapability( const CRYPT_DEVICE cryptDevice, const
                               CRYPT_ALGO_TYPE cryptAlgo, CRYPT_QUERY_INFO *cryptQueryInfo );
```

Parameters *cryptDevice*
The encryption device to be queried.

cryptAlgo
The encryption algorithm to be queried.

cryptQueryInfo
The address of a `CRYPT_QUERY_INFO` structure which is filled with the

information on the requested algorithm and mode, or null if this information isn't required.

Remarks Any fields in the `CRYPT_QUERY_INFO` structure that don't apply to the algorithm being queried are set to `CRYPT_ERROR`, null or zero as appropriate. To determine whether an algorithm is available (without returning information on them), set the query information pointer to null.

See also `cryptQueryCapability`.

cryptEncrypt

The **cryptEncrypt** function is used to encrypt or hash data.

```
int cryptEncrypt( const CRYPT_CONTEXT cryptContext, void *buffer, const int length );
```

Parameters

- cryptContext*
The encryption context to use to encrypt or hash the data.
- buffer*
The address of the data to be encrypted or hashed.
- length*
The length in bytes of the data to be encrypted or hashed.

Remarks Public-key encryption and signature algorithms have special data formatting requirements that need to be taken into account when this function is called. You shouldn't use this function with these algorithm types, but instead should use the higher-level functions **cryptCreateSignature**, **cryptCheckSignature**, **cryptExportKey**, and **cryptImportKey**.

See also `cryptDecrypt`.

cryptEnd

The **cryptEnd** function is used to shut down cryptlib after use. This function should be called after you have finished using cryptlib.

```
int cryptEnd( void );
```

Parameters None

See also `cryptInit`.

cryptExportCert

The **cryptExportCert** function is used to export an encoded signed public key certificate, certification request, CRL, or other certificate-related item from a certificate container object.

```
int cryptExportCert( void *certObject, const int certObjectMaxLength, int *certObjectLength, const CRYPT_CERTFORMAT_TYPE certFormatType, const CRYPT_CERTIFICATE certificate );
```

Parameters

- certObject*
The address of a buffer to contain the encoded certificate.
- certObjectMaxLength*
The maximum size in bytes of the buffer to contain the exported certificate.
- certObjectLength*
The address of the exported certificate length.
- certFormatType*
The encoding format for the exported certificate object.

certificate

The address of the certificate object to be exported.

Remarks The certificate object needs to have all the required fields filled in and must then be signed using **cryptSignCert** before it can be exported.

See also **cryptImportCert**.

cryptExportKey

The **cryptExportKey** function is used to share a session key between two parties by either exporting a session key from a context in a secure manner or by establishing a new shared key. The exported/shared key is placed in a buffer in a portable format that allows it to be imported back into a context using **cryptImportKey**.

If an existing session key is to be shared, it can be exported using either a public key or key certificate or a conventional encryption key. If a new session key is to be established, it can be done using a Diffie-Hellman encryption context.

```
int cryptExportKey( void *encryptedKey, const int encryptedKeyMaxLength, int
                  *encryptedKeyLength, const CRYPT_HANDLE exportKey, const
                  CRYPT_CONTEXT sessionKeyContext );
```

Parameters *encryptedKey*

The address of a buffer to contain the exported key. If you set this parameter to null, **cryptExportKey** will return the length of the exported key in *encryptedKeyLength* without actually exporting the key.

encryptedKeyMaxLength

The maximum size in bytes of the buffer to contain the exported key.

encryptedKeyLength

The address of the exported key length.

exportKey

A public-key or conventional encryption context or key certificate object containing the public or conventional key used to export the session key.

sessionKeyContext

An encryption context containing the session key to export (if the key is to be shared) or an empty context with no key loaded (if the key is to be established).

Remarks A session key can be shared in one of two ways, either by one party exporting an existing key and the other party importing it, or by both parties agreeing on a key to use. The export/import process requires an existing session key and a public/private or conventional encryption context or key certificate object to export/import it with. The key agreement process requires a Diffie-Hellman context and an empty session key context (with no key loaded) that the new shared session key is generated into.

See also **cryptExportKeyEx**, **cryptImportKey**, **cryptQueryObject**.

cryptExportKeyEx

The **cryptExportKeyEx** function is used to share a session key between two parties by either exporting a session key from a context in a secure manner or by establishing a new shared key, with extended control over the exported key format. The exported/shared key is placed in a buffer in a portable format that allows it to be imported back into a context using **cryptImportKey**.

If an existing session key is to be shared, it can be exported using either a public key or key certificate or a conventional encryption key. If a new session key is to be established, it can be done using a Diffie-Hellman encryption context.

```
int cryptExportKeyEx( void *encryptedKey, const int encryptedKeyMaxLength, int
                     *encryptedKeyLength, const CRYPT_FORMAT_TYPE formatType, const
                     CRYPT_HANDLE exportKey, const CRYPT_CONTEXT sessionKeyContext );
```

- Parameters**
- encryptedKey*
The address of a buffer to contain the exported key. If you set this parameter to null, **cryptExportKeyEx** will return the length of the exported key in *encryptedKeyLength* without actually exporting the key.
 - encryptedKeyMaxLength*
The maximum size in bytes of the buffer to contain the exported key.
 - encryptedKeyLength*
The address of the exported key length.
 - formatType*
The format for the exported key.
 - exportKey*
A public-key or conventional encryption context or key certificate object containing the public or conventional key used to export the session key.
 - sessionKeyContext*
An encryption context containing the session key to export (if the key is to be shared) or an empty context with no key loaded (if the key is to be established).
- Remarks** A session key can be shared in one of two ways, either by one party exporting an existing key and the other party importing it, or by both parties agreeing on a key to use. The export/import process requires an existing session key and a public/private or conventional encryption context or key certificate object to export/import it with. The key agreement process requires a Diffie-Hellman context and an empty session key context (with no key loaded) that the new shared session key is generated into.
- See also** **cryptExportKey**, **cryptImportKey**, **cryptQueryObject**.
-

cryptFlushData

The **cryptFlushData** function is used to flush data through an envelope or session object, completing processing and (for session objects) sending the data to the remote client or server.

```
int cryptFlushData( const CRYPT_HANDLE cryptHandle );
```

- Parameters**
- cryptHandle*
The envelope or session object to flush the data through.

See also **cryptPopData**, **cryptPushData**.

cryptGenerateKey

The **cryptGenerateKey** function is used to generate a new key into an encryption context.

```
int cryptGenerateKey( const CRYPT_CONTEXT cryptContext );
```

- Parameters**
- cryptContext*
The encryption context into which the key is to be generated.
- Remarks** Hash contexts don't require keys, so an attempt to generate a key into a hash context will return CRYPT_ERROR_NOTAVAIL.
- cryptGenerateKey** will generate a key of a length appropriate for the algorithm being used into an encryption context. If you want to specify the generation of a key of a particular length, you should set the CRYPT_CTXINFO_KEYSIZE attribute before calling this function.

The generation of large public-key encryption or digital signature keys can take quite some time. If the environment you are working in supports background processing, you should use **cryptGenerateKeyAsync** to generate the key instead.

See also [cryptGenerateKeyAsync](#).

cryptGenerateKeyAsync

The **cryptGenerateKeyAsync** function is used to asynchronously generate a new key into an encryption context.

```
int cryptGenerateKeyAsync( const CRYPT_CONTEXT cryptContext );
```

Parameters *cryptContext*

The encryption context into which the key is to be generated.

Remarks Hash contexts don't require keys, so an attempt to generate a key into a hash context will return CRYPT_ERROR_NOTAVAIL.

cryptGenerateKeyAsync will generate a key of a length appropriate for the algorithm being used into an encryption context. If you want to specify the generation of a key of a particular length, you should set the CRYPT_CTXINFO_ - KEYSIZE attribute before calling this function.

See also [cryptAsyncCancel](#), [cryptAsyncQuery](#).

cryptGetAttribute

The **cryptGetAttribute** function is used to obtain a boolean or numeric value, status information, or object from a cryptlib object.

```
int cryptGetAttribute( const CRYPT_HANDLE cryptObject, const CRYPT_ATTRIBUTE_TYPE attributeType, int *value );
```

Parameters *cryptObject*

The object from which to read the boolean or numeric value, status information, or object.

attributeType

The attribute which is being read.

value

The boolean or numeric value, status information, or object.

See also [cryptDeleteAttribute](#), [cryptGetAttributeString](#), [cryptSetAttribute](#), [cryptSetAttributeString](#).

cryptGetAttributeString

The **cryptGetAttributeString** function is used to obtain text or binary strings or time values from a cryptlib object.

```
int cryptGetAttributeString( const CRYPT_HANDLE cryptObject, const CRYPT_ATTRIBUTE_TYPE attributeType, void *value, int *valueLength );
```

Parameters *cryptObject*

The object from which to read the text or binary string or time value.

attributeType

The attribute which is being read.

value

The address of a buffer to contain the data. If you set this parameter to null, **cryptGetAttributeString** will return the length of the data in *attributeLength* without returning the data itself.

valueLength

The length of the data in bytes.

See also `cryptDeleteAttribute`, `cryptGetAttribute`, `cryptSetAttribute`,
`cryptSetAttributeString`.

cryptGetCertExtension

The `cryptGetCertExtension` function is used to obtain a generic blob-type certificate extension from a certificate object or public or private key with an attached certificate.

```
int cryptGetCertExtension( const CRYPT_CERTIFICATE certificate, const char *oid, int
                          *criticalFlag, void *extension, const int extensionMaxLength, int
                          *extensionLength );
```

Parameters *cryptObject*

The certificate or public/private key object from which to read the extension.

oid

The object identifier value for the extension being queried, specified as a sequence of integers.

criticalFlag

The critical flag for the extension being read.

extension

The address of a buffer to contain the data. If you set this parameter to null, `cryptGetCertExtension` will return the length of the data in *extensionLength* without returning the data itself.

extensionMaxLength

The maximum size in bytes of the buffer to contain the extension data.

extensionLength

The length in bytes of the extension data.

Remarks cryptlib directly supports extensions from X.509, PKIX, SET, SigG, and various vendors itself, so you shouldn't use this function for anything other than unknown, proprietary extensions.

See also `cryptAddCertExtension`, `cryptDeleteCertExtension`.

cryptGetPrivateKey

The `cryptGetPrivateKey` function is used to create an encryption context from a private key in a keyset or crypto device. The private key is identified either through the key owner's name or their email address.

```
int cryptGetPrivateKey( const CRYPT_HANDLE cryptHandle, CRYPT_CONTEXT
                       *cryptContext, const CRYPT_KEYID_TYPE keyIDtype, const void *keyID,
                       const char *password );
```

Parameters *cryptHandle*

The keyset or device from which to obtain the key.

cryptContext

The address of the context to be fetched.

keyIDtype

The type of the key ID, either `CRYPT_KEYID_NAME` for the name or key label, or `CRYPT_KEYID_EMAIL` for the email address.

keyID

The key ID of the key to read.

password

The password required to decrypt the private key, or null if no password is required.

Remarks **cryptGetPrivateKey** will return `CRYPT_ERROR_WRONGKEY` if an incorrect password is supplied. This can be used to determine whether a password is necessary by first calling the function with a null password and then retrying the read with a user-supplied password if the first call returns with `CRYPT_ERROR_WRONGKEY`.

See also **cryptAddPrivateKey**, **cryptAddPublicKey**, **cryptDeleteKey**, **cryptGetPublicKey**.

cryptGetPublicKey

The **cryptGetPublicKey** function is used to create an encryption context from a public key in a keyset or crypto device. The public key is identified either through the key owner's name or their email address.

```
int cryptGetPublicKey( const CRYPT_HANDLE cryptObject, CRYPT_HANDLE *publicKey,
                      const CRYPT_KEYID_TYPE keyIDtype, const void *keyID );
```

Parameters *cryptObject*

The keyset or device from which to obtain the key.

publicKey

The address of the context or certificate to be fetched.

keyIDtype

The type of the key ID, either `CRYPT_KEYID_NAME` for the name or key label, or `CRYPT_KEYID_EMAIL` for the email address.

keyID

The key ID of the key to read.

Remarks The type of object in which the key is returned depends on the keyset or device from which it is being read. Most sources will provide a key certificate object, but some will return only an encryption context containing the key. Both types of object can be used with cryptlib functions.

See also **cryptAddPrivateKey**, **cryptAddPublicKey**, **cryptDeleteKey**, **cryptGetPrivateKey**.

cryptImportCert

The **cryptImportCert** function is used to import an encoded certificate, certification request, CRL, or other certificate-related item into a certificate container object.

```
int cryptImportCert( const void *certObject, const int certObjectLength, const CRYPT_USER
                    cryptUser, CRYPT_CERTIFICATE *certificate );
```

Parameters *certObject*

The address of a buffer that contains the encoded certificate.

certObjectLength

The encoded certificate length.

cryptUser

The user who is to own the imported object or `CRYPT_UNUSED` for the default, normal user.

certificate

The certificate object to be created using the imported certificate data.

See also **cryptExportCert**.

cryptImportKey

The **cryptImportKey** function is used to share a session key between two parties by importing an encrypted session key that was previously exported with **cryptExportKey** into an encryption context.

If an existing session key being shared, it can be imported using either a private key or a conventional encryption key. If a new session key is being established, it can be done using a Diffie-Hellman encryption context.

```
int cryptImportKey( const void *encryptedKey, const int encryptedKeyLength, const
                   CRYPT_CONTEXT importContext, const CRYPT_CONTEXT
                   sessionKeyContext );
```

Parameters	<i>encryptedKey</i>
	The address of a buffer that contains the exported key created by cryptExportKey .
	<i>encryptedKeyLength</i>
	The length in bytes of the encrypted key data.
	<i>importContext</i>
	A public-key or conventional encryption context containing the private or conventional key required to import the session key.
	<i>sessionKeyContext</i>
	The context used to contain the imported session key.
Remarks	A session key can be shared in one of two ways, either by one party exporting an existing key and the other party importing it, or by both parties agreeing on a key to use. The export/import process requires an existing session key and a public/private or conventional encryption context or key certificate object to export/import it with. The key agreement process requires a Diffie-Hellman context and an empty session key context (with no key loaded) that the new shared session key is generated into.
See also	cryptExportKey , cryptExportKeyEx , cryptImportKey , cryptQueryObject .

cryptInit

The **cryptInit** function is used to initialise cryptlib before use. This function should be called before any other cryptlib function is called.

```
int cryptInit( void );
```

Parameters None

See also **cryptEnd**.

cryptKeysetClose

The **cryptKeysetClose** function is used to destroy a keyset object after use. This closes the connection to the key collection or keyset and frees up any memory it uses.

```
int cryptKeysetClose( const CRYPT_KEYSET keyset );
```

Parameters *keyset*
The keyset object to be destroyed.

See also **cryptKeysetOpen**.

cryptKeysetOpen

The **cryptKeysetOpen** function is used to establish a connection to a key collection or keyset.

```
int cryptKeysetOpen( CRYPT_KEYSET *keyset, const CRYPT_USER cryptUser, const
                    CRYPT_KEYSET_TYPE keysetType, const char *name, const
                    CRYPT_KEYOPT_TYPE options );
```

Parameters

- keyset*
The address of the keyset object to be created.
- cryptUser*
The user who is to own the keyset object or CRYPT_UNUSED for the default, normal user.
- keysetType*
The keyset type to be used.
- name*
The name of the keyset.
- options*
Option flags to apply when opening or accessing the keyset.

See also [cryptKeysetClose](#).

cryptPopData

The **cryptPopData** function is used to remove data from an envelope or session object.

```
int cryptPopData( const CRYPT_HANDLE envelope, void *buffer, const int length, int
                  *bytesCopied );
```

Parameters

- envelope*
The envelope or session object from which to remove the data.
- buffer*
The address of the data to remove.
- length*
The length of the data to remove.
- bytesCopied*
The address of the number of bytes copied from the envelope.

See also [cryptPushData](#).

cryptPushData

The **cryptPushData** function is used to add data to an envelope or session object.

```
int cryptPushData( const CRYPT_HANDLE envelope, const void *buffer, const int length, int
                  *bytesCopied );
```

Parameters

- envelope*
The envelope or session object to which to add the data.
- buffer*
The address of the data to add.
- length*
The length of the data to add.
- bytesCopied*
The address of the number of bytes copied into the envelope.

See also [cryptPopData](#).

cryptQueryCapability

The **cryptQueryCapability** function is used to obtain information about the characteristics of a particular encryption algorithm. The information returned covers the algorithm's key size, data block size, and other algorithm-specific information.

```
int cryptQueryCapability( const CRYPT_ALGO_TYPE cryptAlgo, CRYPT_QUERY_INFO
                        *cryptQueryInfo );
```

- Parameters**
- cryptAlgo*
The encryption algorithm to be queried.
 - cryptQueryInfo*
The address of a **CRYPT_QUERY_INFO** structure which is filled with the information on the requested algorithm and mode, or null if this information isn't required.
- Remarks**
- Any fields in the **CRYPT_QUERY_INFO** structure that don't apply to the algorithm being queried are set to **CRYPT_ERROR**, null or zero as appropriate. To determine whether an algorithm is available (without returning information on it), set the query information pointer to null.
- See also**
- cryptDeviceQueryCapability.**
-

cryptQueryObject

The **cryptQueryObject** function is used to obtain information about an exported key object created with **cryptExportKey** or a signature object created with **cryptCreateSignature**. It returns information such as the type and algorithms used by the object.

```
int cryptQueryObject( const void *objectData, const int objectDataLength,
                    CRYPT_OBJECT_INFO *cryptObjectInfo );
```

- Parameters**
- objectData*
The address of a buffer that contains the object created by **cryptExportKey** or **cryptCreateSignature**.
 - objectDataLength*
The length in bytes of the object data.
 - cryptObjectInfo*
The address of a **CRYPT_OBJECT_INFO** structure that contains information on the exported key or signature.
- Remarks**
- Any fields in the **CRYPT_OBJECT_INFO** structure that don't apply to the object being queried are set to **CRYPT_ERROR**, null or zero as appropriate.
- See also**
- cryptCheckSignature, cryptCreateSignature, cryptExportKey, cryptImportKey.**
-

cryptSetAttribute

The **cryptSetAttribute** function is used to add boolean or numeric information, command codes, and objects to a cryptlib object.

```
int cryptSetAttribute( const CRYPT_HANDLE cryptObject, const CRYPT_ATTRIBUTE_TYPE
                    attributeType, const int value );
```

- Parameters**
- cryptObject*
The object to which to add the value.
 - attributeType*
The attribute which is being added.
 - value*
The boolean or numeric value, command code, or object which is being added.

See also cryptDeleteAttribute, cryptGetAttribute, cryptGetAttributeString, cryptSetAttributeString.

cryptSetAttributeString

The **cryptSetAttributeString** function is used to add text or binary strings or time values to an object.

```
int cryptSetAttributeString( const CRYPT_HANDLE cryptObject, const
    CRYPT_ATTRIBUTE_TYPE attributeType, const void *value, const int
    valueLength );
```

Parameters

cryptObject
The object to which to add the text or binary string or time value.

attributeType
The attribute which is being added.

value
The address of the data being added.

valueLength
The length in bytes of the data being added.

See also cryptDeleteAttribute, cryptGetAttribute, cryptGetAttributeString, cryptSetAttribute.

cryptSignCert

The **cryptSignCert** function is used to digitally sign a public key certificate, CA certificate, certification request, CRL, or other certificate-related item held in a certificate container object.

```
int cryptSignCert( const CRYPT_CERTIFICATE certificate, const CRYPT_CONTEXT
    signContext );
```

Parameters

certificate
The certificate container object that contains the certificate item to sign.

signContext
A public-key encryption or signature context containing the private key used to sign the certificate.

Remarks Once a certificate item has been signed, it can no longer be modified or updated using the usual certificate manipulation functions. If you want to add further data to the certificate item, you have to start again with a new certificate object.

See also cryptCheckCert.

cryptUIDisplayCert

The **cryptUIDisplayCert** function displays a certificate object such as a certificate or certificate chain to the user.

```
int cryptUIDisplayCert( const CRYPT_CERTIFICATE certificate, const HWND hWnd );
```

Parameters

certificate
The certificate object to display.

hWnd
The handle of the owner window, or NULL if the certificate viewer dialog has no owner.

See also cryptUIGenerateKey.

cryptUIGenerateKey

The **cryptUIGenerateKey** function is used to generate a new key into an encryption context and obtain from the user the information required to create or obtain a certificate from a CA. This function presents the user with a key generation wizard that takes them through the key generation process and obtains the information needed for certificate creation.

```
int cryptUIGenerateKey( const CRYPT_DEVICE device, CRYPT_CONTEXT *cryptContext,  
                        const CRYPT_CERTIFICATE certificate, char *password, const HWND hWnd  
                        );
```

Parameters

device

The crypto device in which the key is to be generated, or CRYPT_UNUSED if no crypto device is being used.

cryptContext

The address of the encryption context into which the key is to be generated.

certificate

The certificate object that will be filled in with the user's details.

password

The password selected by the user.

hWnd

The handle of the owner window, or NULL if the certificate viewer dialog has no owner.

See also

cryptUIDisplayCert.

Acknowledgements

Alexey Kirichenko provided information on NtQuerySystemInfo for randomness-gathering under WinNT/Win2K to avoid the need to access the buggy Windows registry performance counters.

Brian Gladman wrote the AES code.

Chris Wedgwood and Paul Kendall helped write the Unix random data gathering routines.

endergone Zwiebeltüte helped debug the SSL/TLS implementation.

Eric Young and the OpenSSL team wrote the conventional encryption and hashing code and bignum library.

Jean-Loup Gailly and Mark Adler wrote the zlib compression code.

Joerg Plate did the Amiga port.

Markus F.X.J. Oberhumer did the 32-bit DOS port.

Matt Thomlinson and Blake Coverett helped fix up and debug the Win32 random data gathering routines.

Matthijs van Duin, Sascha Kratky, and Jeff Lamarche did the Macintosh port.

Nathan Hammond did the MVS port.

Osma Ahvenlampi did the PPC BeOS port.

Sami Tolvanen implemented the cryptlib GUI interface.

Sriram Ramachandran did the Cygwin port.

Steve Landers provided the Tcl bindings, with financial support from Eolas Technologies.

Stuart Woolford and Mario Korva did the OS/2 port.

Trevor Perrin did the C#, Java, and Python bindings.

Wolfgang Gothier did the Delphi and Visual Basic bindings and tracked down a number of *really* obscure probl[^]H[^]H[^]H[^]Hundocumented features.