



Applied Cryptography

CO 487



Alfred Menezes

Preface

Disclaimer Much of the information on this set of notes is transcribed directly/indirectly from the lectures of CO 487 during Winter 2021 as well as other related resources. I do not make any warranties about the completeness, reliability and accuracy of this set of notes. Use at your own risk.

Notations:

- $k \in_R K$ means that k is chosen uniformly and independently at random from K .
- $a \parallel b$ and a, b denote the concatenation between strings a and b .

For any questions, send me an email via <https://notes.sibeliusp.com/contact>.

You can find my notes for other courses on <https://notes.sibeliusp.com/>.

Sibeliusp Peng

Contents

Preface	1
List of Algorithms	7
1 Introduction	8
1.1 Secure Web Transactions	9
1.2 The TLS Protocol	9
1.3 Cryptography in Context	10
2 Symmetric-Key Cryptography	11
2.1 Basic concepts	11
2.1.1 The Simple Substitution Cipher	12
2.1.2 Security of SKES	12
2.1.3 Polyalphabetic Ciphers	15
2.2 The One-Time Pad	15
2.3 Stream Ciphers	16
2.4 The RC ₄ Stream Cipher	17
2.4.1 Case Study: Wired Equivalent Privacy	18
2.4.2 Fluhrer-Mantin-Shamir Attack	21
2.5 ChaCha ₂₀ Stream Cipher	22
2.5.1 ChaCha ₂₀ Quarter Round Function	23
2.6 Block Ciphers	24
2.6.1 Brief History of Block Ciphers	25
2.6.2 Some Desirable Properties of Block Ciphers	26
2.6.3 The Data Encryption Standard (DES)	27
2.6.4 Double-DES	28
2.6.5 Triple-DES	30
2.6.6 The Advanced Encryption Standard (AES)	31
2.6.7 Block Cipher Modes of Operation	38
3 Hash Functions	40
3.1 Introduction	40
3.1.1 Hash Functions from Block Ciphers	42
3.1.2 Desirable security properties for hash functions	42
3.1.3 Relationships between PR, $\mathcal{2}PR$, CR	44
3.2 Generic Attacks	46
3.2.1 Find Preimages	46
3.2.2 Find Collisions	46
3.2.3 VW Parallel Collision Search	47
3.3 Iterated Hash Functions (Merkle Meta-Method)	49

3.3.1	Collision Resistance of Iterated Hash Functions	50
3.3.2	Provable Security	50
3.4	MDx-Family of Hash Functions	51
3.5	SHA	51
3.5.1	SHA-1	51
3.5.2	SHA-2 Family	52
3.5.3	Description of SHA-256	52
3.5.4	SHA-3	54
3.5.5	NIST's Policy on Hash Functions	54
4	Message authentication code schemes	55
4.1	Introduction	55
4.1.1	Security Definition	56
4.2	Generic Attacks on MAC schemes	56
4.3	MACs Based on Block Ciphers	57
4.3.1	Security of CBC-MAC	57
4.3.2	Encrypted CBC-MAC (EMAC)	58
4.4	MACs Based on Hash Functions	58
4.4.1	Secret Prefix Method	58
4.4.2	Secret Suffix Method	59
4.4.3	Envelope Method	59
4.4.4	HMAC	59
4.5	Case study: GSM	60
4.5.1	GSM Security	61
5	Authentic Encryption	62
5.1	AES-GCM	63
5.1.1	CTR: CounTeR Mode of Encryption	63
5.1.2	AES-GCM Encryption, Decryption/ Authentication Procedure	64
5.1.3	Insights into Authentication Mechanism	66
5.1.4	Some Features of AES-GCM	67
5.2	Google Encryption	68
5.2.1	Google Data	68
5.2.2	Key Management Service (KMS)	69
5.2.3	Google's Key Hierarchy	69
6	Introduction to Public-Key Cryptography	72
6.1	Drawbacks with Symmetric-Key Cryptography	72
6.1.1	Key Establishment Problem	72
6.1.2	Key Management Problem	73
6.1.3	Non-Repudiation is Difficult to Achieve	73
6.2	Public-Key Cryptography	73
6.2.1	Merkle Puzzles	74
6.2.2	Public-Key Encryption	75
6.2.3	Digital Signatures	75
6.2.4	Hybrid Schemes	76
6.3	Algorithmic Number Theory	76
6.3.1	Complexity Theory Review	76
6.3.2	Basic Integer Operations	77
6.3.3	Basic Modular Operations	78
7	RSA	79

7.1	Basic RSA	79
7.1.1	RSA Encryption	79
7.1.2	Basic RSA Signature Scheme	81
7.2	Case Study: QQ Browser	81
7.2.1	Version 1	81
7.2.2	Version 2	82
7.3	Security of RSA Encryption	83
7.4	Integer Factorization	85
7.4.1	Review from complexity theory	85
7.4.2	Factoring Algorithms	85
7.4.3	History of Factoring	86
7.5	RSA Signature Scheme	87
7.5.1	Attack Model	87
7.5.2	Security of a Signature Scheme	88
7.6	RSA PKCS #1 v1.5 Signatures (1993)	88
7.6.1	Bleichenbacher's Attack (2006)	89
8	Elliptic Curve Cryptography	90
8.1	Elliptic Curves	90
8.1.1	Point Addition	92
8.2	Elliptic Curve Discrete Logarithm Problem (ECDLP)	94
8.3	Elliptic Curve Cryptography	96
8.3.1	Elliptic Curves in Practice	97
8.3.2	Modular Reduction	97
8.4	Elliptic Curve Diffie-Hellman (ECDH)	99
8.4.1	Unauthenticated ECDH	99
8.4.2	Authenticated ECDH	100
8.5	Case Study: ECDH in Google	100
8.5.1	TLS (as was commonly implemented)	100
8.5.2	TLS as implemented by Google	100
8.6	The Elliptic Curve Digital Signature Algorithm (ECDSA)	101
9	Bluetooth Security	104
9.1	Introduction	104
9.2	The Bluetooth security protocol	105
9.3	KNOB Attack	108
10	Key Management	109
10.1	Public Key Management	109
10.2	Certification Authorities (CAs)	110
10.3	Public-Key Infrastructures (PKI)	110
10.4	Case Study: TLS	111
10.4.1	TLS Handshake Protocol	111
10.4.2	TLS 1.2 Record Protocol	112
10.4.3	TLS 1.3	112
10.5	Public Key Management in TLS	112
10.5.1	Example of an X.509 Certificate	113
11	Random Bit Generation	116
11.1	Introduction	116
11.2	Cloudflare Random Bit Generation	117
11.3	Weak Random Bit Generation	117

11.4	Pseudorandom Bit Generation	118
12	FIDO U2F	119
12.1	U2F Protocol	119
12.2	Google's Titan Security Key	120
12.3	FIDO2	120
13	The Signal Protocol	121
13.1	Introduction	121
13.2	Signal Objectives	121
13.3	Signal Protocol	122
13.3.1	Registration	122
13.3.2	Root Key Establishment	122
13.3.3	Message Transmission	124
13.4	References	126
14	Post-Quantum Cryptography	127
14.1	Quantum Computers	127
14.2	The Threat of Quantum Computers	127
14.2.1	Shor	127
14.2.2	Grover	127
14.3	Quantum-Safe Candidates	130
15	Bitcoin	131
15.1	Paper Cash	131
15.1.1	Features of Paper Cash	131
15.2	Bitcoin	131
15.2.1	Elements of Bitcoin	133
15.2.2	Transactions	134
15.3	The Blockchain	135
15.4	Ethereum	137
16	Wrap-up	139
16.1	Boring Crypto	139
16.2	Cool Crypto	139

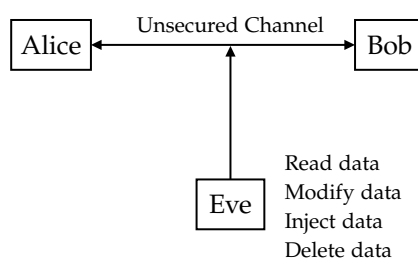
List of Algorithms

1	RC4 Key Scheduling Algorithm	17
2	RC4 Keystream Generator	18
3	WEP protocol: sending packet	19
4	WEP protocol: receiving packet	20
5	ChaCha20 Keystream Generator	24
6	Meet-In-The-Middle Attack on Double-DES	28
7	Encryption of Substitution-Permutation Networks	32
8	AES Encryption	37
9	AES Decryption	37
10	Davies-Meyer hash function	42
11	VW collision finding	48
12	Merkle's hash function	49
13	SHA-256 Preprocessing	53
14	CBC-MAC	57
15	Basic GSM security protocol	61
16	CTR: Encryption	63
17	CTR: Decryption	63
18	AES-GCM encryption/authentication	64
19	AES-GCM decryption/authentication	65
20	Hybrid scheme: encryption of signed message	76
21	Hybrid scheme: authenticity verification	76
22	Modular exponentiation naive algorithm 1	78
23	Modular exponentiation naive algorithm 2	78
24	Modular exponentiation: repeated square-and-multiply	78
25	RSA Key Generation	79
26	RSA Encryption	79
27	RSA Decryption	79
28	RSA Signature Generation	81
29	RSA Signature Verification	81
30	RSA PKCS #1 v1.5 signature generations	88
31	RSA PKCS #1 v1.5 signature verification	88
32	Bleichenbacher's Attack	89
33	Shank's algorithm	96
34	$c = a \cdot b \mod p$	98

35	ECDSA key generation	101
36	ECDSA signature generation	102
37	ECDSA signature verification	102
38	Bluetooth security protocol phase 1: public key exchange	106
39	Bluetooth security protocol phase 2: authentication stage 1	106
40	Bluetooth security protocol phase 3: authentication stage 2	107
41	Bluetooth security protocol phase 4: link key calculation	107
42	Bluetooth security protocol: authentication and encryption	107

Introduction

Cryptography is about securing communications in the presence of *malicious* adversaries.



Note that even if we call him “Eve”, he can do more than eavesdropping... The adversary is malicious, powerful and unpredictable.

Fundamental Goals of Cryptography

1. **Confidentiality:** Keeping data secret from all but those authorized to see it.
2. **Data integrity:** Ensuring data has not been altered by unauthorized means.
3. **Data origin authentication:** Corroborating the source of data.
4. **Non-repudiation:** Preventing an entity from denying previous commitments or actions.

Some examples of unsecured channel:

- Secure Browsing: The Internet
- Online Shopping: The Internet
- Automatic Software Upgrades: The Internet
- Cell Phone Service: Wireless
- Wi-Fi: Wireless
- Bluetooth: Wireless
- Messaging: Wired/Wireless

Communicating Parties

Alice and Bob are two communicating devices.

Alice	Bob	Communication channel
person	person	telephone cable
person	person	cellular network
person	web site	internet
iPhone	wireless	router wireless
iPhone	headphones	wireless
iPhone	service provider	cellular network
your car's brakes	another car	wireless
smart card	bank machine	financial network
smart meter	energy provider	wireless
military commander	satellite	space

1.1 Secure Web Transactions

Transport Layer Security (TLS): The cryptographic protocol used by web browsers for secure web transactions for secure access to amazon, gmail, hotmail, facebook etc.

TLS is used to assure an individual user (called a client) of the authenticity of the web site (called the server) he or she is visiting, and to establish a secure communications channel for the remainder of the session.

Symmetric-key cryptography: The client and server a priori share some secret information k , called a key.

They can subsequently engage in secure communications by encrypting their messages with AES and authenticating the resulting ciphertexts with HMAC.

How do they establish the shared secret key k ?

Public-key cryptography: Communicating parties a priori share some authenticated (but non-secret) information.

To establish a secret key, the client selects the secret session key k , and encrypts it with the server's RSA public key. Then only the server can decrypt the resulting ciphertext with its RSA private key to recover k .

How does the client obtain an authentic copy of the server's RSA public key?

Signature scheme: The server's RSA public key is signed by a Certifying Authority using the RSA signature scheme.

The client can verify the signature using the Certifying Authority's RSA public verification key which is embedded in its browser. Then, the client obtains an authentic copy of the server's RSA public key.

1.2 The TLS Protocol

1. When a client first visits a secured web page, the server transmits its certificate to the client.
 - The certificate contains the server's identifying information (e.g., web site name and URL) and RSA public key, and the RSA signature of a certifying authority.
 - The certifying authority (e.g., Verisign) is trusted to carefully verify the server's identity

before issuing the certificate.

2. Upon receipt of the certificate, the client verifies the signature using the certifying authority's public key, which is embedded in the browser. A successful verification confirms the authenticity of the server and of its RSA public key.
3. The client selects a random session key k , encrypts it with the server's RSA public key, and transmits the resulting ciphertext to the server.
4. The server decrypts the ciphertext to obtain the session key, which is then used with symmetric-key schemes to encrypt (e.g. with AES) and authenticate (e.g. with HMAC) all sensitive data exchanged for the remainder of the session.
5. The establishment of a secure link is indicated by a closed padlock in the browser. Clicking on this icon reveals the server's certificate and information about the certifying authority.

TLS is one of the most successful security technologies ever deployed. But is TLS really secure?

There are many potential security vulnerabilities:

1. The crypto is weak (e.g., AES, HMAC, RSA).
2. Quantum attacks on the underlying cryptography.
3. Weak random number generation.
4. Issuance of fraudulent certificates
 - In 2001, Verisign erroneously issued two Class 3 code-signing certificates to a person masquerading as a Microsoft representative.
 - Mistake due to human error.
5. Software bugs (both inadvertent and malicious).
6. Phishing attacks.
7. TLS only protects data during transit. It does not protect your data when collected at the server.

Many servers store large amounts of credit card data and other personal information.
8. The National Security Agency (NSA)

1.3 Cryptography in Context

Cybersecurity is comprised of the concepts, technical measures, and administrative measures used to protect networks, computers, programs and data from deliberate or inadvertent unauthorized access, disclosure, manipulation, loss or use. Also known as information security. Cybersecurity includes the study of computer security, network security and software security.

Note that Cryptography \neq Cybersecurity.

- Cryptography provides some mathematical tools that can assist with the provision of cybersecurity services. It is a small, albeit an indispensable, part of a complete security solution.
- Security is a chain
 - Weak links become targets; one flaw is all it takes.
 - Cryptography is usually not the weakest link. However, when the crypto fails the damage can be catastrophic.

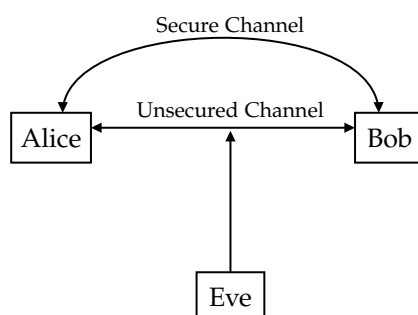
Symmetric-Key Cryptography

2.1 Basic concepts

symmetric-key encryption scheme

A **symmetric-key encryption scheme (SKES)** consists of:

- M - the plaintext space,
- C - the ciphertext space,
- K - the key space,
- a family of encryption functions: $E_k : M \rightarrow C, \forall k \in K$,
- a family of decryption functions: $D_k : C \rightarrow M, \forall k \in K$, such that $D_k(E_k(m)) = m$ for all $m \in M, k \in K$.



1. Alice and Bob agree on a secret key $k \in K$ by communicating over the secure channel.
2. Alice computes $c = E_k(m)$ and sends the ciphertext c to Bob over the unsecured channel.
3. Bob retrieves the plaintext by computing $m = D_k(c)$.

Some examples:

- WWII: Enigma Machine, Alan Turing,
https://en.wikipedia.org/wiki/Cryptanalysis_of_the_Enigma
- WWII: Lorenz Machine, Bill Tutte, Colossus

<http://tinyurl.com/COBillTutte>, <http://billtuttememorial.org.uk/>

2.1.1 The Simple Substitution Cipher

- M = all English msgs.
- C = all encrypted msgs.
- K = all permutations of the English alphabet.
- $E_k(m)$: Apply permutation k to m , one letter at a time.
- $D_k(c)$: Apply inverse permutations k^{-1} to c , one letter at a time.

Example:

$k =$	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
	D	N	X	E	S	K	O	J	T	A	F	P	Y	I	Q	U	B	R	Z	G	V	C	H	M	W	L

Encryption: $m = \text{the big dog}$, $c = E_k(\text{the big dog}) = GJS\ NTO\ EQO$.

Decryption: $c = GJS\ NTO\ EQO$, $m = E_k^{-1}(GJS\ NTO\ EQO) = \text{the big dog}$.

2.1.2 Security of SKES

Is the simple substitution cipher a secure SKES? but wait, what does it mean for a SKES to be secure? We need a **security definition**.

1. What are the computational powers of the adversary?
2. How does the adversary interact with the two communicating parties?
3. What is the adversary's goal?

Security model: Defines the computational abilities of the adversary, and how she interacts with the communicating parties.

Basic assumption: The adversary knows everything about the SKES, except the particular key k chosen by Alice and Bob. (Avoid security by obscurity!!) Security should only depend on the secrecy of the secret keying material, not on the secrecy of the algorithms that describe the cryptographic scheme.

Let's now consider the *computational power* of the adversary.

- **Information-theoretic security:** Eve has infinite computational resources.
- **Complexity-theoretic security:** Eve is a 'polynomial-time Turing machine'.
- **Computational security:** Eve has 36,768 Intel E5-2683 V4 cores running at 2.1 GHz at her disposal. See: [Graham](#) in the basement of MC

We say: Eve is "*computationally bounded*".

In this course, we will be concerned with computational security.

Next consider how the adversary can *interact* with Alice and Bob.

- Passive attacks:
 - **Ciphertext-only attack:** The adversary knows some ciphertext (that was generated by Alice or Bob).
 - **Known-plaintext attack:** The adversary also knows some plaintext and the corresponding

ciphertext. (stronger because more info than before)¹

- Active attacks: (even stronger)
 - **Chosen-plaintext attack:** The adversary can also *choose* some plaintext and obtains the corresponding ciphertext. This is at least as strong as known-plaintext attack because the attacker gets to choose the plain text for which it is given the corresponding ciphertext.
- Other attacks (not considered in this course):
 - Clandestine attacks: bribery, blackmail, etc.
 - Side-channel attacks: monitor the encryption and decryption equipment (timing attacks, power analysis attacks, electromagnetic-radiation analysis, etc.)

Finally, we will consider the *adversary's goal* (from strongest to weakest).

1. Recover the secret key.
2. Systematically recover plaintext from ciphertext (without necessarily learning the secret key).
3. Learn some partial information about the plaintext from the ciphertext (other than its length).

If the adversary can achieve 1 or 2, the SKES is said to be **totally insecure** (or **totally broken**).

If the adversary cannot learn any partial information about the plaintext from the ciphertext (except possibly its length), the SKES is said to be **semantically secure**. So the ciphertext hides all the meaning about the corresponding plaintext.

security of SKES

A symmetric-key encryption scheme is said to be **secure** if it is semantically secure against chosen-plaintext attack by a computationally bounded adversary.

Note that the definitions captures the computational abilities of the adversary, namely some concrete upper bound on its computational resources. It captures how the adversary interacts with Alice and Bob, namely by mounting a chosen-plaintext attack. And finally, it captures the goal of the adversary, namely breaking semantic security. Note also that in the definition the attacker's capabilities are *as strong as possible*, namely a chosen-plaintext attack and its goal is *as weak as possible* namely, breaking semantic security

From the definition, we can also deduce what it means to break an encryption scheme.

1. The adversary is given a challenge ciphertext c (generated by Alice or Bob using their secret key k).
2. During its computation, the adversary can select plaintexts and obtains (from Alice or Bob) the corresponding ciphertexts.
3. After a feasible amount of computation, the adversary obtains some information about the plaintext m corresponding to the challenge ciphertext c (other than the length of m).

If the attacker can win this game specified in these three steps, then we'll say that the encryption scheme is insecure.

¹In practice, Eve might get such plaintext because in many applications data is first formatted by prepending to it some header information whose contents might be predictable, for example email headers.

Desirable Properties of a SKES

1. Efficient algorithms should be known for computing E_k and D_k (i.e., for encryption and decryption).
2. The secret key should be small (but large enough to render exhaustive key search infeasible).
3. The scheme should be secure.
4. The scheme should be secure even against the designer of the system.

The last point is to ensure that the designer hasn't somehow inserted a backdoor into the encryption scheme whereby the adversary can learn a secret keying material by looking at ciphertext, while this method would be hard to find by other people. This fourth property can sometimes be very difficult to obtain in practice.

Now we can see whether the simple substitution cipher is secure: *Totally insecure against a chosen-plaintext attack*. This is because the adversary can simply give the plaintext message that's comprised of the English alphabet from a to z to Alice for encryption. Alice returns to the adversary the ciphertext, which it turns out is the secret key itself. And so by simply giving Alice a short plaintext and getting back the corresponding ciphertext the adversary has learnt Alice's secret key.

What about security under a ciphertext-only attack? Is exhaustive key search possible?

In this attack, we are given sufficient amounts of ciphertext c , decrypt c using each possible key until c decrypts to a plaintext message which "makes sense" (normal English). In principle, 30 characters of ciphertext are sufficient on average to yield a unique plaintext that is a sensible English message. In practice, a few hundred characters are needed.

If we tried exhaustive search, the number of keys to try is $26! \approx 2^{88}$, which is very large in the following sense: If the adversary uses 10^6 computers, each capable of trying 10^9 keys per second, then exhaustive key search takes about 10^4 years. So, exhaustive key search is infeasible.

In this course, a cryptographic task that requires

- 2^{40} operations is considered very easy.
- 2^{56} operations is considered easy.
- 2^{64} operations is considered feasible.
- 2^{80} operations is considered barely feasible.
- 2^{128} operations is considered infeasible.

The Bitcoin network is presently performing hash operations at the rate of $2^{66.4}$ per second (or $2^{91.3}$ per year).

The Landauer limit from thermodynamics suggests that exhaustively trying 2^{128} symmetric keys would require $\gg 3000$ gigawatts of power for one year (which is $\gg 100\%$ of the world's energy production). http://en.wikipedia.org/wiki/Brute-force_attack

security level of a cryptographic scheme

A cryptographic scheme is said to have a **security level** of ℓ bits if the fastest known attack on the scheme takes approximately 2^ℓ operations.

As of the year 2021, a security level of 128 bits is desirable in practice.

Let's return to our question of whether the simple substitution cipher is secure against a ciphertext-only attack. In fact, simple frequency analysis of ciphertext letters can be used to recover the key. One would simply find the most frequently occurring ciphertext letter; this would most likely correspond to the English letter *e* which is the most frequently occurring letter in the English alphabet. Hence, the simple substitution cipher is *totally insecure* even against a ciphertext-only attack.

2.1.3 Polyalphabetic Ciphers

Basic idea: Use several permutations, so a plaintext letter is encrypted to one of several possible ciphertext letters.

Example: Vigenère cipher

Secret key is an English word having no repeated letters. E.g., $k = \text{CRYPTO}$.

Example of encryption:

$$\begin{array}{rcccccccccccccccc}
 m & = & t & h & i & s & & i & s & & a & & m & e & s & s & a & g & e \\
 + & k & = & C & R & Y & P & & T & O & & C & & R & Y & P & T & O & C & R \\
 \hline
 c & = & V & Y & G & H & & B & G & & C & & D & C & H & L & O & I & V
 \end{array}$$

Here, $A = 0, B = 1, \dots, Z = 25$; addition of letters is mod 26.

Decryption is subtraction modulo 26: $m = c - k$.

Frequency distribution of ciphertext letters is flatter (than for a simple substitution cipher).

The Vigenère cipher is totally insecure against a chosen-plaintext attack. Not unexpectedly, the Vigenère cipher is also totally insecure against a ciphertext-only attack.

2.2 The One-Time Pad

The one-time pad is a symmetric-key encryption scheme invented by Vernam in 1917 for the telegraph system. The key is a random² string of letters. Example of encryption:

$$\begin{array}{rcccccccccccccccc}
 m & = & t & h & i & s & & i & s & & a & & m & e & s & s & a & g & e \\
 + & k & = & Z & F & K & W & & O & G & & P & & S & M & F & J & D & L & G \\
 \hline
 c & = & S & M & S & P & & W & Y & & P & & F & Q & X & C & D & R & K
 \end{array}$$

Note that the key is as long as the plaintext. One major disadvantage of the one-time pad is that the secret key is of the same length as the plaintext. This is especially inconvenient if Alice and Bob are exchanging long messages over a period of time, and so it might be tempting in practice to reuse the secret key.

The key should *not* be re-used: If $c_1 = m_1 + k$ and $c_2 = m_2 + k$, then $c_1 - c_2 = m_1 - m_2$. Thus $c_1 - c_2$ depends only on the plaintext (and not on the key) and hence can leak information about the plaintext. In particular, if m_1 is known, then m_2 can be easily computed.

In the one-time pad, we took the plaintext letters to be the english alphabet. From now on in the course, unless otherwise stated, messages and keys will be assumed to be binary/*bit strings*. This is a reasonable convention since computers process binary data.

²independently and uniformly at random

⊕

⊕ is bitwise exclusive-or (XOR) i.e., bitwise addition modulo 2. For example:

$$1011001010 \oplus 1001001001 = 0010000011.$$

Note that $x \oplus x = 0$ and $x \oplus y = y \oplus x$. Hence if $x = y \oplus z$, then $x \oplus y = z$.

So, for the one-time pad:

Encryption: $c = m \oplus k$.

Decryption: $m = c \oplus k$.

<https://cryptosmith.com/2008/05/31/stream-reuse/> is an example which reuses the same key...

Security of One-Time Pad

Perfect secrecy: The one-time pad is semantically secure against ciphertext-only attack by an adversary with infinite computational resources. This can be proven formally using concepts from information theory [Shannon 1949].

The bad news with perfect secrecy is that Shannon (1949) proved that if plaintexts are m -bit strings, then any symmetric-key encryption scheme with perfect secrecy must have $|K| \geq 2^m$. So, perfect secrecy (and the one-time pad) is fairly useless in practice. However, all is not lost. One can use one-time pads to inspire the construction of so-called *stream ciphers*.

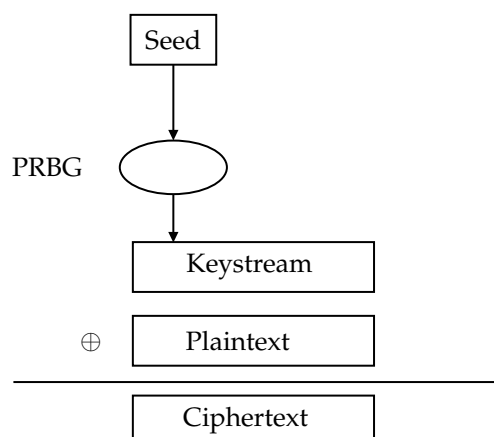
2.3 Stream Ciphers

Basic idea: Instead of using a random key in the one-time pad, use a “pseudorandom” key.

A **pseudorandom bit generator** (PRBG) is a deterministic algorithm that takes as input a (relatively small random) seed, and outputs a longer “pseudorandom” sequence called the keystream.

Using a PRBG for encryption (a stream cipher):

The seed is the secret key shared by Alice and Bob.



The XOR portion of the stream cipher emulates a one-time pad. Note though that we no longer have perfect secrecy because the keystream is no longer purely random but pseudo-random. Thus security depends on the quality of the PRBG.

There are two *security requirements* for the pseudo-random bit generator.

First, The keystream should be “indistinguishable” from a random sequence (the **indistinguishability requirement**).

If an adversary knows a portion c_1 of ciphertext and the corresponding plaintext m_1 , then she can easily find the corresponding portion $k_1 = c_1 \oplus m_1$ of the keystream. Thus, given portions of the keystream, it should be infeasible to learn any information about the rest of the keystream (the **unpredictability requirement**).

Aside: Don’t use UNIX random number generators (rand and srand) for cryptography!

- $X_0 = \text{seed}, X_{i+1} = aX_i + b \pmod n, i \geq 0.$
- a, b and n are fixed and public constants.

And so the UNIX routines rand and srand do not meet the unpredictability requirement.

One difficulty with using stream ciphers in practice is that we still have the requirement as we did with the one-time pad that keystream should never be reused.

Then we will see two examples of stream ciphers: Rc4 and ChaCha20.

2.4 The RC4 Stream Cipher

It was designed by Ron Rivest in 1987. He is the “R” in RSA. Until recently, was widely used in commercial products including Adobe Acrobat, Windows password encryption, Oracle secure SQL, TLS, etc.

- *Pros:* Extremely simple; extremely fast; variable key length. No catastrophic weakness has been found since 1987.
- *Cons:* Design criteria are proprietary; not much public scrutiny until the year 2001.

In the past 10 years, many weaknesses have been found in RC4 and so its use has rapidly declined in practice.

RC4 has two components: (i) a *key scheduling algorithm*, and (ii) a *keystream generator*.

In the following, $K[i], \bar{K}[i]$ and $S[i]$ are 8-bit integers (bytes).

Algorithm 1: RC4 Key Scheduling Algorithm

Input: Secret key $K[0], K[1], \dots, K[d-1]$. (Keylength is $8d$ bits.)

Output: 256-long array: $S[0], S[1], \dots, S[255]$.

```

1 for  $i = 0, \dots, 255$  do
2    $S[i] \leftarrow i$ 
3    $\bar{K}[i] \leftarrow K[i \bmod d]$ 
4  $j \leftarrow 0$ 
5 for  $i = 0, \dots, 255$  do
6    $j \leftarrow (\bar{K}[i] + S[i] + j) \bmod 256$ 
7   Swap( $S[i], S[j]$ )
```

The idea of key scheduling is to begin with the identity permutation capital S of the integers 0 through 255, and then apply 256 random swaps to this permutation with the hope that the resulting permutation is a random looking permutation. S is a “random-looking” permutation of $\{0, 1, 2, \dots, 255\}$ that is generated from the secret key.

In keystream generation, the array S from key scheduling is used to generate keystream of the desired

length. Alice would generate keystream of a certain length, add it to plaintext to get ciphertext. And Bob would generate the same keystream and add it to ciphertext to get plaintext.

Algorithm 2: RC4 Keystream Generator

Input: 256-long byte array: $S[0], S[1], \dots, S[255]$ produced by the RC4 Key Scheduling

Algorithm

Output: Keystream.

```

1  $i \leftarrow 0; j \leftarrow 0$ 
2 while keystream bytes are required do
3    $i \leftarrow (i + 1) \bmod 256$ 
4    $j \leftarrow (S[i] + j) \bmod 256$ 
5    $\text{Swap}(S[i], S[j])$ 
6    $t \leftarrow (S[i] + S[j]) \bmod 256$ 
7    $\text{Output}(S[t])$ 
  
```

ENCRYPTION: The keystream bytes are stored with the plaintext bytes to produce ciphertext bytes.

2.4.1 Case Study: Wired Equivalent Privacy

Wireless Security

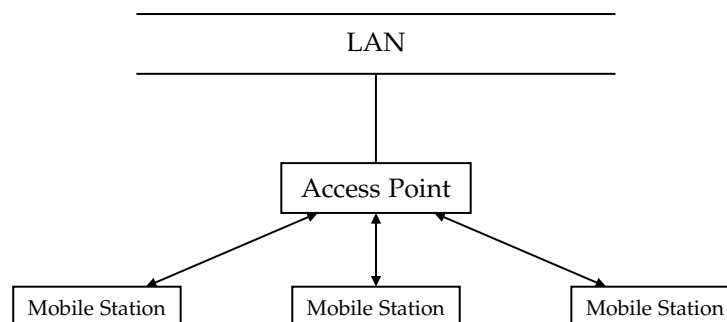
Wireless networks have become prevalent. Popular standards for wireless networks: IEEE 802.11 (longer range, higher speeds, commonly used for wireless LANs) and Bluetooth (short range, low speed).

New security concerns:

- More attack opportunities (no need for physical access).
- Attack from a distance (> 1 km with good antennae).
- No physical evidence of attack.

IEEE 802.11 Security

IEEE 802.11 standard for wireless LAN communications includes a protocol called **Wired Equivalent Privacy** (WEP). This standard was ratified in September 1999. This was a very exciting time because this was the first time that wi-fi was available to the consumer market. Multiple amendments: 802.11a (1999), 802.11b (1999), 802.11g (2003), 802.11i (2004), 802.11n (2009)... WEP's goal is (only) to protect link-level data during wireless transmission between mobile stations and access points. A mobile station might be your cell phone or your laptop, and an access point is a wi-fi router somewhere in the building.



WEP had three main *security goals*.

1. *Confidentiality*: Prevent casual eavesdropping. To achieve this, RC4 is used for encryption.
2. *Data Integrity*: Prevent tampering with transmitted messages. To achieve this, an ‘integrity checksum’ is used for WEP.
3. *Access Control*: Protect access to a wireless network infrastructure. This was achieved by discarding all packets that are not properly encrypted using WEP.

Description of WEP Protocol

Mobile stations share a secret key k with access point. Here k is either 40 bits or 104 bits in length. The IEEE standard does not specify how the key is to be distributed. But in practice, one shared key per LAN is common; this key is manually injected into each access point and mobile station; the key is not changed very frequently.

Plaintext messages are divided into packets of some fixed length (e.g., 1500 bytes). These packets were encrypted using RC4. Since a mobile station or access point might encrypt many packets in a given period of time, WEP had to be careful not to reuse RC4 keystreams.

Thus WEP uses a per-packet 24-bit *initialization vector* (IV) v to process each packet. WEP does not specify how the IVs are managed. In practice: either a random IV is generated for each packet; or the IV is set to 0 and incremented by 1 for each use.

To send a packet m , an entity (e.g., Alice) does the following:

Algorithm 3: WEP protocol: sending packet

- 1 Select a 24-bit IV v .
 - 2 Compute a 32-bit checksum: $S = \text{CRC}(m)$.
 - 3 Compute $c = (m \parallel S) \oplus \text{RC4}(v \parallel k)$.
 - 4 Send (v, c) over the wireless channel.
-

Note:

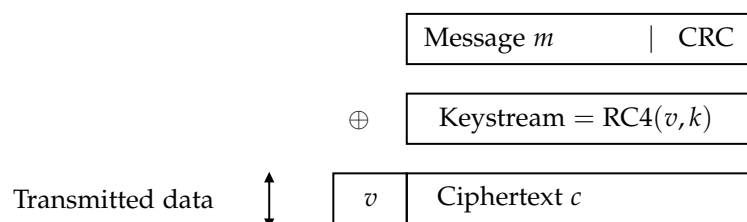
- 802.11 specifies that a CRC-32 checksum be used. CRC-32 is linear. That is, for any two messages m_1 and m_2 of the same bitlength,

$$\text{CRC}(m_1 \oplus m_2) = \text{CRC}(m_1) + \text{CRC}(m_2).$$

(The details of CRC-32 are not important to us)

- \parallel (and also a comma) denotes concatenation
- $(v \parallel k)$ is the key used the RC4 stream cipher.

Here’s a diagram of the encryption process.



The receiver of (v, c) does the following:

Algorithm 4: WEP protocol: receiving packet

- 1 Compute $(m \parallel S) = c \oplus \text{RC4}(v \parallel k)$.
 - 2 Compute $S' = \text{CRC}(m)$; reject the packet if $S' \neq S$.
-

Are confidentiality, data integrity, and access control achieved? Intuitively the answer is yes. No (Borisov, Goldberg & Wagner; 2001),

Problem 1: IV Collision

Suppose that two packets (v, c) and (v, c') use the same IV v . Let m, m' be the corresponding plaintexts. Then $c \oplus c' = (m \parallel S) \oplus (m' \parallel S')$. Thus, the eavesdropper can compute $m \oplus m'$. If m is known, the m' is immediately available. If m is not known, then one may be able to use the expected distribution of m and m' to discover information about them. (Some contents of network traffic is predictable.)

Since there are only 2^{24} choices for the IV, collisions are guaranteed after enough time - a few days on a busy network (5 Mbps). If IVs are randomly selected, then one can expect a collision after about 2^{12} ($= \sqrt{2^{24}}$) packets. This is due to the birthday paradox.

Birthday paradox

Suppose that an urn contains n numbered balls. Suppose that balls are drawn from the urn, one at a time, with replacement. The expected number of draws before a ball is selected for a second time (called a *collision*) is approximately $\sqrt{\pi n/2} \approx \sqrt{n}$.

Collisions are more likely if keys k are long-lived and the same key is used for multiple mobile stations in a network.

Conclusion: WEP does not provide a high degree of confidentiality.

Problem 2: Checksum is Linear

CRC-32 is used to check integrity. This is fine for random errors, but not for deliberate ones.

It is easy to make controlled changes to (encrypted) packets: Suppose (v, c) is an encrypted packet. Let $c = \text{RC4}(c \parallel k) \oplus (m \parallel S)$, where k, m, S are unknown to the adversary.

The adversary chooses a bit string Δ of the same length as m , where the ones in Δ correspond to the bits of m that she wishes to change. Let $m' = m \oplus \Delta$, where Δ is a bit string. And so in this way the attacker can make controlled changes to the unknown plaintext m .

Let $c' = c \oplus (\Delta \parallel \text{CRC}(\Delta))$. Then (v, c') is a valid encrypted packet for m' .

$$\begin{aligned}
 c' &= c \oplus (\Delta \parallel \text{CRC}(\Delta)) \\
 &= \text{RC4}(v \parallel k) \oplus (m \parallel S) \oplus (\Delta \parallel \text{CRC}(\Delta)) \\
 &= \text{RC4}(v \parallel k) \oplus (m \oplus \Delta \parallel \text{CRC}(m) \oplus \text{CRC}(\Delta)) \\
 &= \text{RC4}(c \parallel k) \oplus (m' \parallel \text{CRC}(m \oplus \Delta)) \\
 &= \text{RC4}(v \parallel k) \oplus (m' \parallel \text{CRC}(m'))
 \end{aligned}$$

Hence the receiver will accept c' as a validly encrypted WEP packet, and will accept m' as the plaintext. In this way, the adversary has successfully made a controlled change to the plaintext m without actually knowing the plaintext.

Conclusion: WEP does not provide data integrity.

Problem 3: Integrity Function is Unkeyed

Suppose that an attacker learns the plaintext m corresponding to a single encrypted packet (v, c) . Then, the attacker can compute the RC4 keystream $\text{RC4}(v \parallel k) = c \oplus (m \parallel \text{CRC}(m))$. Henceforth, the attacker can compute a valid encrypted packet for *any* plaintext m' of her choice: (v, c') , where $c' = \text{RC4}(v \parallel k) = c \oplus (m' \parallel \text{CRC}(m'))$. Note that this is a validly encrypted ciphertext for the plaintext m' .

Conclusion: WEP does not provide access control.

Optional Reading: Sections 1, 2, 3, 4.1, 4.2, 6 of “[Intercepting mobile communications: The insecurity of 802.11](#)”, by N. Borisov, I. Goldberg and D. Wagner.

2.4.2 Fluhrer-Mantin-Shamir Attack

Shortly after these attacks on WEP were disclosed, a *more devastating attack* was discovered. This was due to Fluhrer, Mantin & Shamir, 2001. Shamir is the “S” in RSA.

The attack makes the following three assumptions:

1. The same 104-bit key k is used for a long period of time. [Most products do this.]
2. The IV is incremented for each packet, or a random IV is selected for each packet. [Most products do this.]
3. The first plaintext byte of each packet (i.e. the first byte of each m) is known to the attacker. [Most wireless protocols prepend the plaintext with some header bytes which are non-secret.]

In the attack, a passive adversary who can collect about 5,000,000 encrypted packets can very easily recover k (and thus totally break the system). [Details not covered in this course.]

The attack can be easily mounted in practice: Can buy a \$100 wireless card and hack drivers to capture (encrypted) packets. On a busy wireless network (5Mbps), 5 million packets can be captured in a few hours, and then k can be immediately computed.

Implementation details: A. Stubblefield, J. Ionnidis, A. Rubin, “Using the Fluhrer, Mantin and Shamir attack to break WEP”, AT&T Technical Report, August 2001.

If you want to mount the WEP attack today, all you have to do is find a wireless network that still uses WEP and then go to one of these two websites: [Aircrack-ng](#), [WEPCrack](#). The latest enhancement of the Fluhrer-Mantin-Shamir attack is aircrack ptw. This can break WEP in under 60 seconds (only about 40,000 packets are needed).

WEP was blamed for the 2007 theft of 45 million credit-card numbers from T.J. Maxx. (T.J. Maxx is an American department store chain). A subsequent class action lawsuit settled for \$40,900,000.

See: <http://tinyurl.com/WEP-TJMaxx>

IEEE 802.11 Update

http://en.wikipedia.org/wiki/Wi-Fi_Protected_Access

WPA2 (Wi-Fi Protected Access) Implements the new IEEE 802.11i standard (2004). Uses the AES block cipher instead of RC4. Deployed ubiquitously in Wi-fi networks. However, deployments of WEP are still out there, check <https://wgle.net/stats>. KRACK attack disclosed in October 2017.

WPA3 Adopted in January 2018. This contains further improvements to WPA2, including counter-measures to the CRACK attack. Dragonblood attack disclosed in April 2019.

This cycle of finding attacks and then fixing them with counter-measures continues today.

The Fluhrer-Mantin-Shamir attack exploits known biases in the first few bytes of the keystream. The attack can be defeated by discarding the first few bytes (e.g., 768 bytes) of the keystream. [Details omitted]

As of 2013, approximately 50% of all TLS traffic was secured using RC4. 2013-2021: Because of several new weaknesses discovered, RC4 has been deprecated in applications such as TLS. (As of October 2019, $\approx 11.6\%$ of websites have RC4 enabled, and only $\approx 1.1\%$ actually use it.)

Conclusions: Don't use RC4. Instead use ChaCha20 or AES-CTR (more on these later).

Lessons Learned

1. Details matter: RC4 was considered to be a secure stream cipher. However, it was improperly used in WEP and the result was a highly insecure communications protocol. Do not assume that "obvious" ways of using cryptographic functions are secure.
2. Attacks only get better; they never get worse.
 - (a) Moore's law (1965): computers get twice as fast every two years.
 - (b) Known attacks are constantly being tweaked and improved.
 - (c) New attacks are constantly being invented.
3. Designing security is hard:
 - Designing cryptographic protocols is complicated and difficult.
 - Clearly state your security objectives.
 - Hire cryptography experts to design your security. Programming or engineering experts are not good enough.
 - Make your protocols available for public scrutiny.
4. There is a big demand in industry for "security engineers": People who have a deep understanding of applied cryptography and have excellent programming skills.

2.5 ChaCha20 Stream Cipher

It was designed by Dan Bernstein in 2008.

The ChaCha20 stream cipher is conceptually simple, word-oriented, and uses only simple arithmetic operations (integer addition modulo 2^{32} , xor, and left rotations³). As a result, it is extremely fast in software, and does not require any special hardware. To date, no security weaknesses have been found. ChaCha20 is widely deployed in practice, including in TLS.

Let's first describe ChaCha20's *initial state*. We will use the following notation:

- 256-bit key: $k = (k_1, k_2, \dots, k_8)$
- 96-bit nonce: $n = (n_1, n_2, n_3)$
- 128-bit constant: $f = (f_1, f_2, f_3, f_4)$
- 32-bit counter: c

A hexadecimal digit is a 4-bit number.

³These three arithmetic operations are included in the instruction sets of most modern processors.

The words are f_1, f_2, f_3, f_4 , written as hexadecimal strings of length 8. For example, $f_1 = 0x61707865$. Each of the eight hexadecimal digits is between 0 and 9 or a to f. These represent the decimal numbers between 0 and 15, or a binary string of length 4. And so this hexadecimal string of length 8 can be viewed as a binary string of length 32.

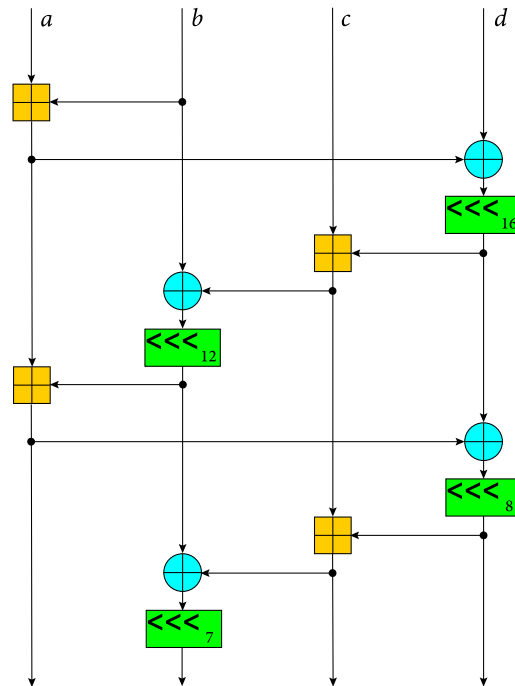
A nonce (or IV) is a non-repeating quantity (either a counter, or a randomly-generated string).

The initial state of ChaCha20 can be represented by this 4×4 array of 32-bit words:

f_1	f_2	f_3	f_4	=	S_1	S_2	S_3	S_4
k_1	k_2	k_3	k_4		S_5	S_6	S_7	S_8
k_5	k_6	k_7	k_8		S_9	S_{10}	S_{11}	S_{12}
c	n_1	n_2	n_3		S_{13}	S_{14}	S_{15}	S_{16}

2.5.1 ChaCha20 Quarter Round Function

It's called a quarter round function since it operates on a quarter of the state, which is a binary string of length 128 bits.



picture from wiki: https://commons.wikimedia.org/wiki/File:ChaCha_Cipher_Quarter_Round_Function.svg.

The purpose of the quarter round function is to mix up the bits of a, b, c , and d in a complicated nonlinear fashion. This is done with three arithmetic operations defined on 32-bit words, \oplus : XOR; and \boxplus : integer addition modulo 2^{32} ; and $\lll t$: left-rotate by t bit positions.

QR: Quarter Round Function**Input** Four 32-bit words a, b, c, d

$$\begin{aligned}
a &\leftarrow a \boxplus b, & d &\leftarrow d \oplus a, & d &\leftarrow d \lll 16 \\
c &\leftarrow c \boxplus d, & b &\leftarrow b \oplus c, & b &\leftarrow b \lll 12 \\
a &\leftarrow a \boxplus b, & d &\leftarrow d \oplus a, & d &\leftarrow d \lll 8 \\
c &\leftarrow c \boxplus d, & b &\leftarrow b \oplus c, & b &\leftarrow b \lll 7
\end{aligned}$$

Output a, b, c, d

Finally, here is a description of ChaCha20's *keystream generator*.

Algorithm 5: ChaCha20 Keystream Generator

-
- 1 Select a nonce n and initialize the counter c .
 - 2 **while** *keystream bytes are required* **do**
 - 3 Create the initial state S .
 - 4 Make a copy S' of S .
 - 5 Update S by repeating the following 10 times:

$QR(S_1, S_5, S_9, S_{13}) \quad QR(S_2, S_6, S_{10}, S_{14})$
 $QR(S_3, S_7, S_{11}, S_{15}) \quad QR(S_4, S_8, S_{12}, S_{16})$
 $QR(S_1, S_6, S_{11}, S_{16}) \quad QR(S_2, S_7, S_{12}, S_{13})$
 $QR(S_3, S_8, S_9, S_{14}) \quad QR(S_4, S_5, S_{10}, S_{15})$
 - 6 Output $S \oplus S'$ (64 keystream bytes).
 - 7 Increment the counter.
-

ENCRYPTION: The keystream bytes are xored with the plaintext bytes to produce ciphertext bytes. The nonce is included in the ciphertext.

When Bob receives the ciphertext and the nonce, he regenerates a keystream using the keystream generation algorithm just as Alice did, and then he xors the keystream bytes with the ciphertext bytes to produce the plaintext bytes.

We can now see why ChaCha20 is very efficient. 64 key stream bytes can be generated using only 960 computer instructions.

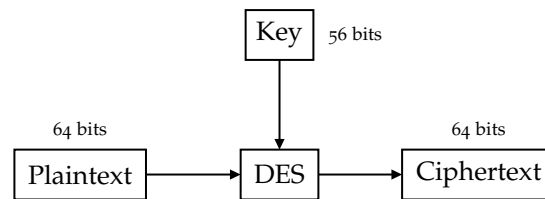
Unlike RC4, ChaCha20 incorporates a nonce and a counter into the encryption algorithm and thus the problems we saw with using RC4 in WEP are automatically avoided, even though Alice and Bob might use the same secret key for a long period of time. Since they use a 96-bit nonce for each plaintext that they encrypt and they use a counter for each 64 byte portion of the keystream for encrypting a particular plaintext, with very high probability ChaCha20 keystream is never reused by Alice and Bob.

2.6 Block Ciphers

A *block cipher* is a SKES that breaks up the plaintext into blocks of a fixed length (e.g. 128 bits), and encrypts the blocks one at a time.

In contrast, a stream cipher encrypts the plaintext one character (usually a bit) at a time.

Historically important example of a block cipher: The Data Encryption Standard (DES)



Key length: 56 bits; Size of key space 2^{56} ; Block length: 64 bits.

2.6.1 Brief History of Block Ciphers

Late 1960's: Feistel network and LUCIFER designed at IBM.

1972: NBS (now NIST: National Institute of Standards and Technology) solicits proposals for encryption algorithms for the protection of computer data.

1973-1974: IBM develops DES.

1975: NSA (National Security Agency) "fixes" DES. Reduces the key length from 64 bits to 56 bits. (NSA tried for 48 bits; compromised at 56 bits.)

To understand NSA's motivation for doing this, let's take a quick look at the NSA organization and its mission.

NSA

The National Security Agency: Founded in 1952. Budget is Classified (estimated: \$10.8 billion/year), as well as its number of employees (30,000–40,000).

NSA has two main divisions: *Signals Intelligence* (SIGINT) which produces foreign intelligence information and *Information Assurance* (IA) which protects all classified and sensitive information that is stored or sent through US government equipment.

The NSA is very influential in setting US government export policy for cryptographic products (especially encryption). Canadian counterpart: Communications Security Establishment (**CSE**).

Most countries around the world have organizations analogous to the NSA. Here is the logo of Ministry of State Security (China):



Picture from wiki: https://commons.wikimedia.org/wiki/File:Ministry_of_State_Security_of_the_People%27s_Republic_of_China.svg

NSA Shenanigans

Edward Snowden. CIA/NSA contractor who disclosed in 2013 up to 200,000 NSA classified documents to the press. Currently in Russia under temporary asylum. Documents revealed the enormous capabilities of NSA and its partners (including CSE and GCHQ):

- “Groundbreaking cryptanalytic capabilities”
- Collects vast amounts of internet communications and automatically analyzes the data
- Directly attacks routers, switches, firewalls, etc.
- Installs malware on individual computers
- Breaks into individual computers

Bruce Schneier, a well-known security expert, “The NSA is subverting the Internet and turning it into a massive surveillance tool.”

Block Cipher history continued...

1977: DES adopted as US Federal Information Processing Standard (FIPS 46).

1981: DES adopted as a US banking standard (ANSI X3.92).

1997: NIST begins the AES (Advanced Encryption Standard) competition.

1999: 5 finalists for AES announced.

2001: Rijndael adopted for AES (FIPS 197). AES has three key lengths: 128, 192 and 256 bits.

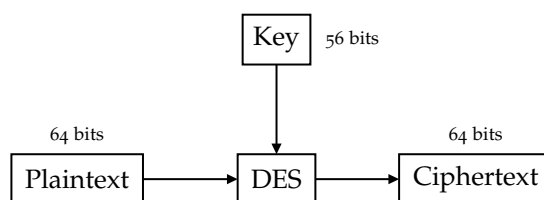
2021: No significant weaknesses found with AES (as yet). AES uptake is rapidly increasing. However, (Triple-)DES is still deployed.

2.6.2 Some Desirable Properties of Block Ciphers

Design principles described by Claude Shannon in 1949:

- Security:
 - *Diffusion*: each ciphertext bit should depend on all plaintext bits.
 - *Confusion*: the relationship between key and ciphertext bits should be complicated.
 - *Key length*: should be small, but large enough to preclude exhaustive key search.
- Efficiency:
 - Simplicity (easier to implement and analyze).
 - High encryption and decryption rate.
 - Suitability for hardware or software.

2.6.3 The Data Encryption Standard (DES)



Key length: 56 bits; Size of key space 2^{56} ; Block length: 64 bits.

The design principles of DES were classified by the NSA in the mid-1970s and they remain classified today, making analysis difficult. Nonetheless, hundreds of research papers have been written on the security of DES since the mid-1970s. (Triple-)DES is still deployed in practice, although its use is rapidly decreasing.

DES Problem 1: Small Key Size

Exhaustive search on key space takes 256 steps and can be easily parallelized.

In the mid-1990s, the RSA security company set some DES challenges. In the challenge, you were given three known plaintext-ciphertext pairs generated with a secret key. The plaintext were the ASCII messages “the unknown message is”, each block being eight ASCII characters or 64 bits. You were also given a fourth ciphertext and finding the corresponding plaintext was the challenge.

T	h	e		u	n	k	n	o	w	n		m	e	s	s	a	g	e		i	s	:		?	?	?	?	?	?	?	?
---	---	---	--	---	---	---	---	---	---	---	--	---	---	---	---	---	---	---	--	---	---	---	--	---	---	---	---	---	---	---	---

The DES challenge was first solved in June of 1997. This was done in three months by Internet search. In 1999, it was broken in 56 hours by DeepCrack machine (1800 chips; \$250,000). In 2006, broken in 153 hours by COPACOBANA machine (\$10,000). In 2012, broken in 11.5 hours by crack.sh (\$200). These experiments show us that the cost and time to perform exhaustive key search decrease over time as computers get faster and cheaper.

The RSA security company also set some 64-bit challenges using the RC5 block cipher which has a key length of 64 bits. In July 2002, broken in 1757 days. Participation by 331,252 individuals.

DES Problem 2: Small Block Size

If plaintext blocks are distributed “uniformly at random”, then the expected number of ciphertext blocks observed before a collision occurs is $\approx 2^{32}$ (by the birthday paradox). Hence the ciphertext reveals some information about the plaintext.

Small block length is also damaging to some authentication applications (more on this later).

To summarize, the only (substantial) weaknesses known in DES are the obvious ones: small key length and small block length.

Question How can one construct a more secure block cipher from DES? (i.e. without changing the internals of DES.)

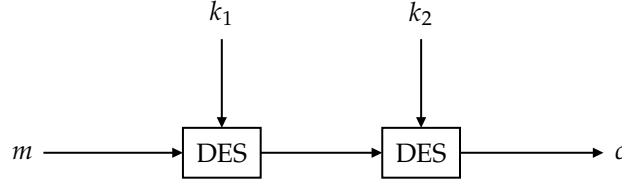
One approach is to use **multiple encryption**: Re-encrypt the ciphertext one or more times using independent keys. Hope that this increases the effective key length.

Note that multiple encryption does not always result in increased security. For example, if E_π denotes the encryption function for the simple substitution cipher with key π , then is $E_{\pi_2} \circ E_{\pi_1}$ any more secure than E_π ? No, since $E_{\pi_2} \circ E_{\pi_1} = E_{\pi_2 \circ \pi_1}$.

2.6.4 Double-DES

Key is $k = (k_1, k_2)$, $k_1, k_2 \in_R \{0, 1\}^{56}$. Here $k \in_R K$ means that k is chosen uniformly and independently at random from K .

Encryption: $c = E_{k_2}(E_{k_1}(m))$. (E = DES encryption, E^{-1} = DES decryption)



Decryption: $m = E_{k_1}^{-1}(E_{k_2}^{-1}(c))$.

The Double-DES key length is $\ell = 112$, so exhaustive key search takes 2^{112} steps (infeasible). Note also that the block length of Double DES is the same as that of DES, namely 64 bits.

Main idea $c = E_{k_2}(E_{k_1}(m))$ if and only if $E_{k_2}^{-1}(c) = E_{k_1}(m)$. Suppose now that (k_1, k_2) is Alice and Bob's secret key. Suppose also that we are given three plaintext-ciphertext pairs $(m_1, c_1), (m_2, c_2)$ and (m_3, c_3) that were generated by Alice or Bob using their secret key (k_1, k_2) . In the meet-in-the-middle attack, we'll search for keys (h_1, h_2) which also encrypt $m_i \rightarrow c_i$ for $i = 1, 2, 3$. We do this by searching all possible candidate keys (h_1, h_2) . If we have found such key (h_1, h_2) , we'll conclude that (h_1, h_2) equals (k_1, k_2) with high probability.

Algorithm 6: Meet-In-The-Middle Attack on Double-DES

Input: 3 known PT/CT pairs $(m_1, c_1), (m_2, c_2), (m_3, c_3)$

Output: The secret key (k_1, k_2)

```

1 foreach  $h_2 \in \{0, 1\}^{56}$  do
2   | Compute  $E_{h_2}^{-1}(c_1)$ , and store  $[E_{h_2}^{-1}(c_1), h_2]$  in a table sorted by first component.
3 foreach  $h_1 \in \{0, 1\}^{56}$  do
4   | Compute  $E_{h_1}(m_1)$ .
5   | Search for  $E_{h_1}(m_1)$  in the table.
   | // We say that  $E_{h_1}(m_1)$  matches table entry  $[E_{h_2}^{-1}(c_1), h_2]$  if  $E_{h_2}^{-1}(c_1) = E_{h_1}(m_1)$ .
6   | foreach match  $[E_{h_2}^{-1}(c_1), h_2]$  in the table do
7     |   if  $E_{h_2}(E_{h_1}(m_2)) = c_2$  then
8       |     if  $E_{h_2}(E_{h_1}(m_3)) = c_3$  then
9         |       | Output  $(h_1, h_2)$  and STOP.
  
```

We need to justify the *correctness* and analyze the *running time* of the attack.

Correctness

We would like to find the number of known plaintext/ciphertext pairs needed for unique key determination.

Let E be a block cipher with key space $K = \{0, 1\}^\ell$, and plaintext and ciphertext space $\{0, 1\}^L$.

Let $k' \in K$ be the secret key chosen by Alice and Bob, and let $(m_i, c_i), 1 \leq i \leq t$, be known plaintext/-ciphertext pairs, where the plaintext m_i are distinct. (Note that $c_i = E_{k'}(m_i)$ for all $1 \leq i \leq t$)

Question Then how large should t be to ensure (with probability very close to 1) that there is only one key $k \in K$ such that $E_k(m_i) = c_i$ for all $1 \leq i \leq t$?

Answer Select t so that $FK \approx 0$ where FK is so-called “false keys”.

For each $k \in K$, then encryption function $E_k : \{0,1\}^L \rightarrow \{0,1\}^L$ is a permutation. This is because the encryption function E_k is an invertible function.

We make the *heuristic assumption* that for each $k \in K$, E_k is a random function (i.e., a randomly selected function). This assumption is certainly false since E_k is *not* random, and because a random function is almost certainly not a permutation. Nonetheless, it turns out that the assumption is good enough for our analysis.

Now, fix $k \in K, k \neq k'$. The probability that $E_k(m_i) = c_i$ for all $1 \leq i \leq t$ is

$$\underbrace{\frac{1}{2^L} \cdot \frac{1}{2^L} \cdots \frac{1}{2^L}}_t = \frac{1}{2^{Lt}}.$$

Here the first probability, $\frac{1}{2^L}$, is the probability that E_k encrypts m_1 to c_1 . This is because there are 2^L possible ciphertext blocks. Similarly for $i = 2, \dots, t$. Then we multiply these probabilities together.

Thus the expected number of false keys $k \in K$ (not including k') for which $E_k(m_i) = c_i$ for all $1 \leq i \leq t$ is

$$FK = \frac{2^\ell - 1}{2^{Lt}}.$$

This expected number is the probability that a single key is a false key times the number of possible false keys, which is the total number of keys $2^\ell - 1$ because we aren't considering k' as a false key. And so the expected number of false keys is given by this expression.

Now let's return to the correctness of the meet-in-the-middle attack. Let E be the DES encryption function, so Double-DES encryption is $c = E_{k_2}(E_{k_1}(m))$.

1. If $\ell = 112, L = 64, t = 3$, then $FK \approx 1/2^{80} \approx 0$. Thus if a Double-DES key (h_1, h_2) is found for which $E_{h_2}(E_{h_1}(m_i)) = c_i$ for $i = 1, 2, 3$, then with very high probability we have $(h_1, h_2) = (k_1, k_2)$.
2. If $\ell = 112, L = 64, t = 1$, then $FK \approx 2^{48}$. Thus the expected number of Double-DES keys (h_1, h_2) for which $E_{h_2}(E_{h_1}(m_1)) = c_1$ is $\approx 2^{48}$.
3. If $\ell = 112, L = 64, t = 2$, then $FK \approx 1/2^{16}$. Thus the expected number of Double-DES keys (h_1, h_2) for which $E_{h_2}(E_{h_1}(m_1)) = c_1$ and $E_{h_2}(E_{h_1}(m_2)) = c_2$ is $\approx 1/2^{16}$.

This justifies correctness of the meet-in-the-middle attack on Double DES.

Runtime Analysis

In the analysis, we'll take one operation to be either a single DES encryption operation E , or a single DES decryption operation E^{-1} .

Now let's take a closer look at the algorithm. The running time of line 2 is 2^{56} operations because we do one single decryption for each key h_2 in the single DES key space.⁴ Similarly, the total running time of line 4 is 2^{56} operations. Line 7 is executed every time there is a match in line 6.

The number of matches equals the number of keys h_1 for which this equation holds, in other words (h_1, h_2) encrypts m_1 to c_1 . We need to count the number of keys (h_1, h_2) that encrypt m_1 to c_1 . That is given by the number of double DES false keys when $t = 1$. By the formula for false keys, we see that

⁴More precisely, the running time here means this particular line would do one operation 2^{56} times: this line has been executed 2^{56} times, and this line consists of 1 operation.

this is roughly 2^{48} , so the attack will execute line 7 a total of 2^{48} times. Thus the cost of line 7 is 2×2^{48} as this line involves 2 DES operations.

Finally, the attack will execute line 8 for each false key (h_1, h_2) that encrypts $m_1 \rightarrow c_1, m_2 \rightarrow c_2$. By the formula for false keys, the expected number of double DES keys when $t = 2$ is roughly $1/2^{16}$, and so we expect that line 8 is executed $1/2^{16}$ times for false keys (h_1, h_2) , and once for the actual secret key (k_1, k_2) . Therefore, the expected running time of line 8 is $2 \times (1 + \frac{1}{2^{16}})$.

To summarize, the number of DES operations is $\approx 2^{56} + 2^{56} + 2 \cdot 2^{48} \approx 2^{57}$. (We are not counting the time to do the sorting and searching)

Note that this is a lot less than the expected running time of exhaustive key search for double DES, which is roughly 2^{112} . Note also that this is a feasible amount of computation and so we might conclude that the meet-in-the-middle attack on Double DES demonstrates that Double DES offers no more security than single DES.

Space requirements

In line 2, the table entry has length $64 + 56$ bits, and we have 2^{56} table entries. Therefore, the space of the table in line 2 is $2^{56}(64 + 56)$ bits $\approx 1,080,863$ Tbytes.

To put this in context, we consider storage units conversion:

- 10^3 bytes = 1 Kbyte (kilo) $\approx 2^{10}$ bytes
- 10^3 Kbytes = 1 Mbyte (mega) $\approx 2^{20}$ bytes
- 10^3 Mbytes = 1 Gbyte (giga) $\approx 2^{30}$ bytes
- 10^3 Gbytes = 1 Tbyte (tera) $\approx 2^{40}$ bytes
- 10^3 Tbytes = 1 Pbyte (peta) $\approx 2^{50}$ bytes
- 10^3 Pbytes = 1 Ebyte (exa) $\approx 2^{60}$ bytes

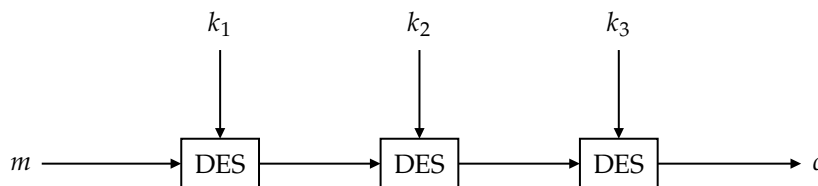
Accessing such a large storage would slow down the attack. Whether the meet-in-the-middle attack on Double DES is indeed cost effective, and in particular is much better than exhaustive key search on single DES? To address this concern, consider time-memory tradeoff. The attack can be modified to decrease the storage requirements at the expense of time: for $s \in [1, 55]$, we have time 2^{56+s} steps and memory 2^{56-s} units.

Conclusions: Double-DES has the same effective key length as DES. Double-DES is not much more secure than DES.

2.6.5 Triple-DES

Triple-DES. Key is $k = (k_1, k_2, k_3)$, $k_1, k_2, k_3 \in_R \{0, 1\}^{56}$.

Encryption: $c = E_{k_3}(E_{k_2}(E_{k_1}(m)))$ where E = DES encryption, E^{-1} = DES decryption



Decryption: $m = E_{k_1}^{-1}(E_{k_2}^{-1}(E_{k_3}^{-1}(c)))$.

Key length of Triple-DES is $\ell = 168$, so exhaustive key search takes 2^{168} steps (infeasible)

One can verify meet-in-the-middle attack takes $\approx 2^{112}$ steps. So, the effective key length of Triple-DES against exhaustive key search is ≈ 112 bits, still infeasible. However, no proof that Triple-DES is more secure than DES.

Note that block length is unchanged.

Triple-DES is still deployed in practice, especially by some financial institutions, where its use can be expected to continue until 2030.

2.6.6 The Advanced Encryption Standard (AES)

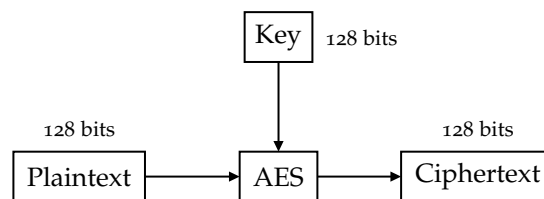
In this section, we will discuss AES, the block cipher that is the most widely deployed symmetric-key encryption scheme.

<http://www.nist.gov/aes>

In 1997, NIST began the competition for selecting the Advanced Encryption Standard or AES to replace the aging Data Encryption Standard.

Four main requirements:

- Key lengths: 128, 192 or 256 bits.
- Block length: 128 bits.
- Efficient on both hardware and software platforms.
- Availability on a worldwide, non-exclusive, royalty-free basis.



In this lecture, we'll describe AES encryption for the case where the secret key as length 128 bits.

The AES Process: 1998: 15 submissions in Round 1. 1999: 5 finalists selected by NIST: MARS, RC6, Rijndael, Serpent, Twofish. 1999: NSA performed a hardware efficiency comparison. **2000: Rijndael was selected.** 2001: The AES standard is officially adopted (FIPS 197). Rijndael is an example of a substitution-permutation network. 2021: No attacks have been found on AES that are (significantly) faster than exhaustive key search.

Substitution-Permutation Networks

A **substitution-permutation network** (SPN) is an iterated block cipher where a round consists of a substitution operation followed by a permutation operation.

The components of an SPN cipher are the following:

- n : the block length.
- ℓ : the key length.
- h : the number of rounds.
- A fixed invertible function $S : \{0,1\}^b \rightarrow \{0,1\}^b$, where b is a divisor of n .
- A fixed permutation P on $\{1,2,\dots,n\}$.
- A key scheduling algorithm that determines subkeys $k_1, k_2, \dots, k_h, k_{h+1}$ from a key k .

Note that n, ℓ, h, S, P and the key scheduling algorithm are public. The only secret in AES is the key k that is selected.

Here is a description of encryption:

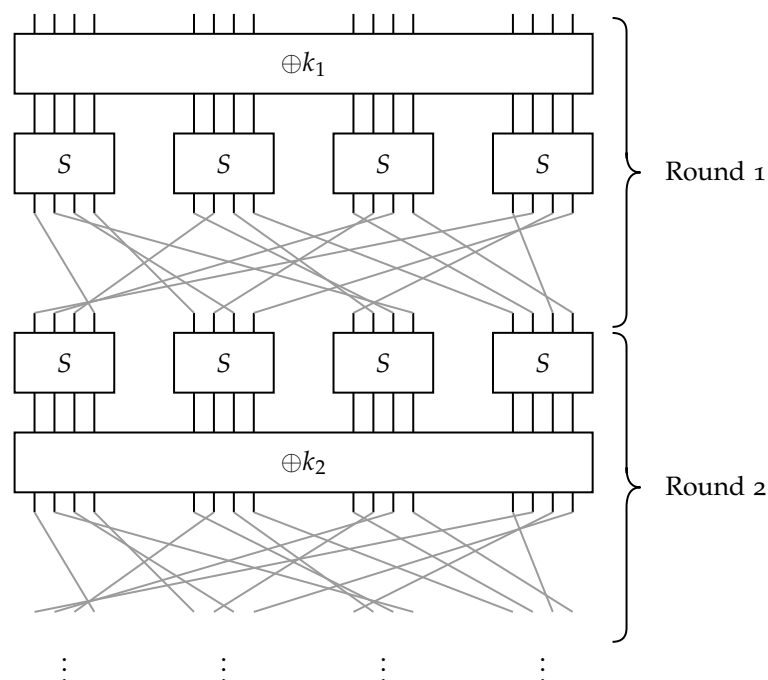
Algorithm 7: Encryption of Substitution-Permutation Networks

```

1  $A \leftarrow \text{plaintext}$ 
2 for  $i = 1 \dots h$  do
3    $A \leftarrow A \oplus k_i$  // XOR
4    $A \leftarrow S(A)$  // Substitution
5    $A \leftarrow P(A)$  // Permutation
6  $A \leftarrow A \oplus k_{h+1}$ 
7 ciphertext  $\leftarrow A$ 

```

The substitution provides confusion. The permutation provides diffusion. Decryption is just the reverse of encryption.



AES is an SPN, where the permutation operation is replaced by two invertible linear transformations. All operations are byte oriented (e.g., $b = 8$ so the S-box maps 8-bits to 8-bits). This allows AES to be efficiently implemented on software platforms. The block length of AES is $n = 128$ bits. Each subkey is 128 bits. AES accepts three different key lengths. The number of rounds h depends on the key length:

cipher	key length ℓ	h
AES-128	128	10
AES-192	192	12
AES-256	256	14

AES Round Operations

Each round updates a variable called State which consists of a 4×4 array of bytes (note: $4 \cdot 4 \cdot 8 = 128$, the block length). State is initialized with the plaintext:

$a_{0,0}$	$a_{0,1}$	$a_{0,2}$	$a_{0,3}$
$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$
$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$
$a_{3,0}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$

\leftarrow plaintext

After h rounds are completed, a final subkey is XOR-ed with State, the result being the ciphertext.

The AES round function uses four invertible operations:

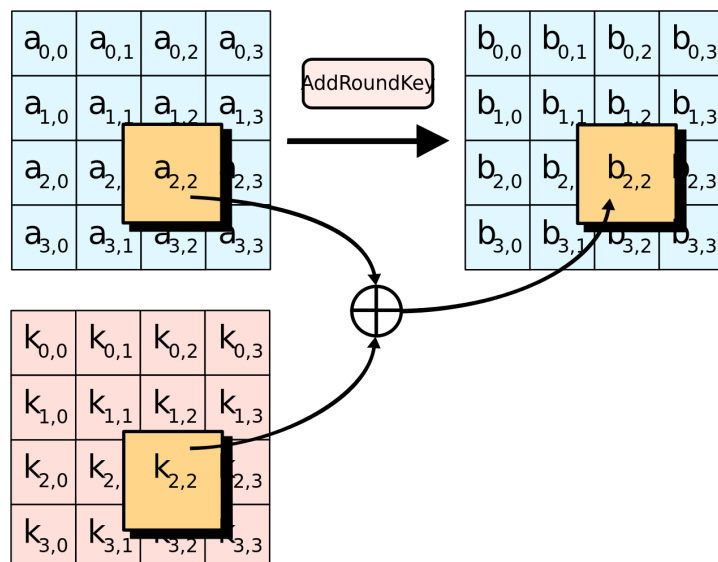
1. AddRoundKey (key mixing)
2. SubBytes (S-box)
3. ShiftRows (permutation)
4. MixColumns (linear transformation).

The pictures below are from wiki:

- <https://commons.wikimedia.org/wiki/File:AES-AddRoundKey.svg>
- <https://commons.wikimedia.org/wiki/File:AES-SubBytes.svg>
- <https://commons.wikimedia.org/wiki/File:AES-ShiftRows.svg>
- <https://commons.wikimedia.org/wiki/File:AES-MixColumns.svg>

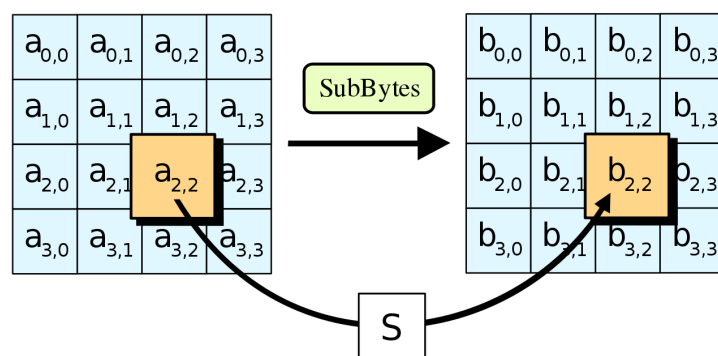
Add Round Key

Bitwise-XOR each byte of State with the corresponding byte of the subkey.



Substitute Bytes

Take each byte in State and replace it with the output of the S-box.



$S : \{0,1\}^8 \rightarrow \{0,1\}^8$ is a fixed, public, invertible function. The S-box is a non-linear function.

The Finite Field $GF(2^8)$

The elements of the finite field $GF(2^8)$ are the polynomials of degree at most 7 in $\mathbb{Z}_2[y]$, with addition and multiplication performed modulo the irreducible polynomial $f(y) = y^8 + y^4 + y^3 + y + 1$.

Interpret an 8-bit string $a = a_7a_6a_5a_4a_3a_2a_1a_0$ as coefficients of the polynomial $a(y) = a_7y^7 + a_6y^6 + a_5y^5 + \dots + a_1y + a_0$ and vice versa.

Example:

Let $a = 11101100 = \text{ec}$ and $b = 00111011 = 3\text{b}$, so $a(y) = y^7 + y^6 + y^5 + y^3 + y^2$ and similarly $b(y) = y^5 + y^4 + y^3 + y + 1$.

1. Addition: $a(y) + b(y) = y^7 + y^6 + y^5 + y^3 + y^2 + y^5 + y^4 + y^3 + y + 1$, so $\text{ec} + 3\text{b} = \text{d7}$.
2. Multiplication: $a(y) \cdot b(y) = y^{12} + y^{10} + y^8 + y^4 + y^2$, which leaves a remainder of $r(y) = y^7 + y^6 + y^3$ upon division by $f(y)$. Hence $a \cdot b = 11001000$ in $GF(2^8)$, or $\text{ec} \cdot 3\text{b} = \text{c8}$.
3. Inversion: $\text{ec}^{-1} = 5\text{f}$ since $\text{ec} \cdot 5\text{d} = 01$.

S-box Definition

Let $p \in \{0,1\}^8$, and consider p as an element of $GF(2^8)$.

1. Let $q = p^{-1}$ if $p \neq 0$, and $q = p$ if $p = 0$.
2. Define $q = (q_7q_6q_5q_4q_3q_2q_1q_0)$.
3. Compute

$$\begin{pmatrix} r_0 \\ r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \\ r_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \pmod{2}$$

Then $S(p) = r = (r_7r_6r_5r_4r_3r_2r_1r_0)$.

So S maps the byte p to the byte r . This is a nonlinear function since inversion in the field $GF(2^8)$ is a non-linear function.

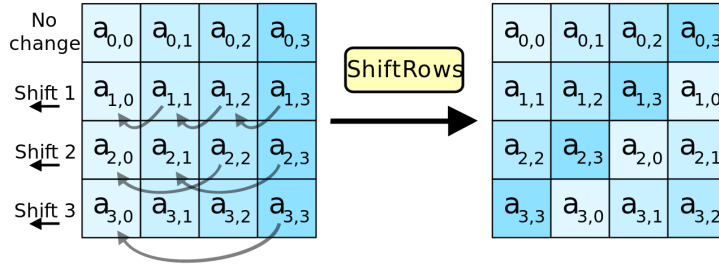
Instead of the algebraic description of the S-box, I can define the S function by this table.

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	63	7c	77	7b	f2	6b	6f	c5	30	01	67	2b	fe	d7	ab	76
1	ca	82	c9	7d	fa	59	47	fo	ad	d4	a2	af	9c	a4	72	co
2	b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
3	04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
4	09	83	2c	1a	1b	6e	5a	ao	52	3b	d6	b3	29	e3	2f	84
5	53	d1	00	ed	20	fc	b1	5b	6a	cb	be	39	4a	4c	58	cf
6	do	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
7	51	a3	40	8f	92	9d	38	f5	bc	b6	da	21	10	ff	f3	d2
8	cd	oc	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
9	60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	ob	db
a	eo	32	3a	oa	49	06	24	5c	c2	d3	ac	62	91	95	e4	79
b	e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
c	ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
d	70	3e	b5	66	48	03	f6	oe	61	35	57	b9	86	c1	1d	9e
e	e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
f	8c	a1	89	od	bf	e6	42	68	41	99	2d	of	bo	54	bb	16

For example, $S(a8) = c2$.

Shift Rows

Permute the bytes of State by applying a *cyclic shift* to each row.



Mix Columns

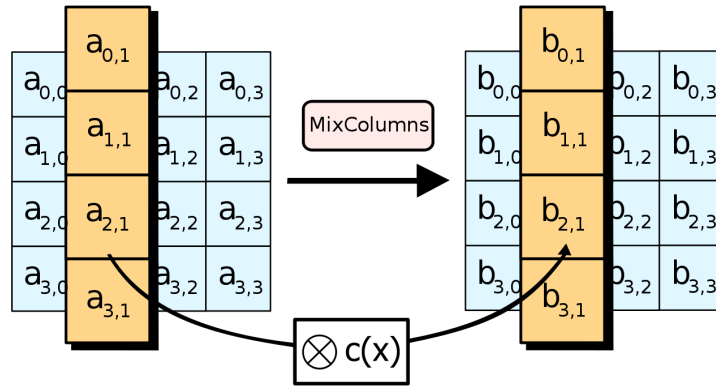
The purpose of MixColumns is to mix up the bits in the four bytes of each column of the State.

1. Read column i of State as a polynomial:

$$(a_{0,i}, a_{1,i}, a_{2,i}, a_{3,i}) = a_{0,i} + a_{1,i}x + a_{2,i}x^2 + a_{3,i}x^3$$

(interpret the coefficients as elements of the finite field $GF(2^8)$).

2. Multiply this polynomial with the constant polynomial $c(x) = 03 \cdot x^3 + 01 \cdot x^2 + 01 \cdot x + 02$ and reduce modulo $x^4 - 1$. This gives a new polynomial: $b_{0,i} + b_{1,i}x + b_{2,i}x^2 + b_{3,i}x^3$



The $\otimes c(x)$ Operation

Let $a(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, where each $a_i \in GF(2^8)$.

Let $c(x) = 02 + 01x + 01x^2 + 03x^3$, where 01, 02, 03 are elements in $GF(2^8)$ (in hexadecimal).

To compute $a(x) \otimes c(x)$:

- Compute $d(x) = a(x) \times c(x)$ (polynomial multiplication, where coefficient arithmetic is in $GF(2^8)$).
- Divide by $x^4 - 1$ to find the remainder polynomial $r(x)$ (equivalently, replace x^4 by 1, x^5 by x , and x^6 by x^2).
- Then $a(x) \otimes c(x) = r(x)$.

Let $a(x) = d0f112bb = d0 + f1x + 12x^2 + bbx^3$. $a(x) \otimes c(x) = 1a + a4x + d3x^2 + e5x^3 = 1aa4d3e5$.

AES Encryption

From the key k derive $h + 1$ subkeys k_0, k_1, \dots, k_h . Below is a description of encryption:

Algorithm 8: AES Encryption

```

1 State  $\leftarrow$  plaintext
2 State  $\leftarrow$  State  $\oplus k_0$ 
3 for  $i = 1 \dots h - 1$  do
4   State  $\leftarrow$  SubBytes(State)
5   State  $\leftarrow$  ShiftRows(State)
6   State  $\leftarrow$  MixColumns(State)
7   State  $\leftarrow$  State  $\oplus k_i$ 
8 State  $\leftarrow$  SubBytes(State)
9 State  $\leftarrow$  ShiftRows(State)
10 State  $\leftarrow$  State  $\oplus k_h$ 
11 ciphertext  $\leftarrow$  State

```

Note that in the final round, **MixColumns** is not applied. This helps make decryption more similar to encryption, and thus is useful when implementing AES. [Details omitted]

AES Decryption

From the key k derive $h + 1$ subkeys k_0, k_1, \dots, k_h . Below is a description of decryption:

Algorithm 9: AES Decryption

```

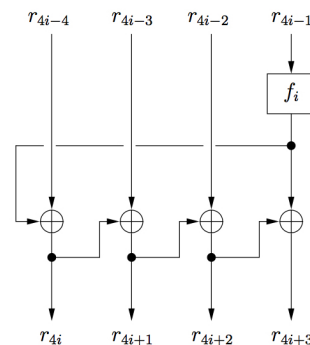
1 State  $\leftarrow$  ciphertext
2 State  $\leftarrow$  State  $\oplus k_h$ 
3 State  $\leftarrow$  InvShiftRows(State)
4 State  $\leftarrow$  InvSubBytes(State)
5 for  $i = h - 1 \dots 1$  do
6   State  $\leftarrow$  State  $\oplus k_i$ 
7   State  $\leftarrow$  InvMixColumns(State)
8   State  $\leftarrow$  InvShiftRows(State)
9   State  $\leftarrow$  InvSubBytes(State)
10 State  $\leftarrow$  State  $\oplus k_0$ 
11 plaintext  $\leftarrow$  State

```

InvMixColumns is multiplication by $d(x) = 0e + 09x + 0dx^2 + 0bx^3$ modulo $x^4 - 1$.

To complete the description of AES encryption, let's see *AES key schedule* for 128-bit keys. For 128-bit keys, AES has 10 rounds, so we need 11 subkeys.

- The first subkey $k_0 = (r_0, r_1, r_2, r_3)$ which is the actual AES key.
- The second subkey is $k_1 = (r_4, r_5, r_6, r_7)$.
- The third subkey is $k_2 = (r_8, r_9, r_9, r_{10})$.
- ...
- The 11th subkey is $k_{10} = (r_{40}, r_{41}, r_{42}, r_{43})$.



Finally, here is a description of the *key schedule functions* f_i . The functions $f_i : \{0,1\}^{32} \rightarrow \{0,1\}^{32}$ are defined as follows:

1. The input is divided into 4 bytes: (a, b, c, d) .
2. Left-rotate the bytes: (b, c, d, a) .
3. Apply the AES S-box to each byte: $(S(b), S(c), S(d), S(a))$.
4. XOR the leftmost byte with the constant ℓ_i , and output the result: $(S(b) \oplus \ell_i, S(c), S(d), S(a))$.

The constants ℓ_i :

i	ℓ_i	i	ℓ_i	i	ℓ_i	i	ℓ_i	i	ℓ_i
1	0x01	2	0x02	3	0x04	4	0x08	5	0x10
6	0x20	7	0x40	8	0x80	9	0x1b	10	0x36

The AES encryption scheme is superior to Triple DES in many ways:

1. AES has three key lengths 128, 192 and 256 bits, whereas Triple DES has only one key length, namely 168 bits. Although its effective key length is 112 bits.
2. AES has block length 128 bits whereas the block length of Triple DES is only 64 bits.
3. AES is byte-oriented and so has very fast software implementations. On the other hand, Triple DES is bit-oriented and is significantly slower in software.

2.6.7 Block Cipher Modes of Operation

From https://en.wikipedia.org/wiki/Block_cipher_mode_of_operation:

In cryptography, a block cipher mode of operation is an algorithm that uses a block cipher to provide information security such as confidentiality or authenticity. A block cipher by itself is only suitable for the secure cryptographic transformation (encryption or decryption) of one fixed-length group of bits called a block. A mode of operation describes how to repeatedly apply a cipher's single-block operation to securely transform amounts of data larger than a block.

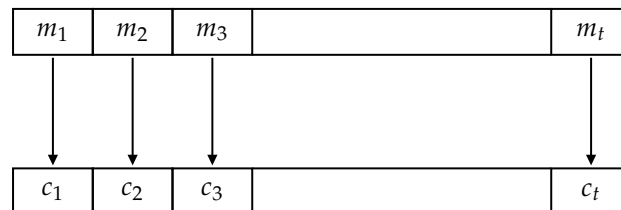
In practice, one might need to encrypt a large quantity of data. Plaintext message is $m = m_1, m_2, \dots, m_t$, where each m_i is an L -bit block.

Question How should we use a block cipher $E_k : \{0,1\}^L \rightarrow \{0,1\}^L$ to encrypt m ?

One way to encrypt the multi-block message m is to use electronic codebook or ECB mode.

Electronic Codebook (ECB) Mode

Encrypt blocks independently, one at a time: $c = c_1, c_2, \dots, c_t$, where $c_i = E_k(m_i)$.



Decryption: $m_i = E_k^{-1}(c_i)$, $i = 1, 2, \dots, t$.

Drawback: Identical plaintexts result in identical ciphertexts (under the same key), and thus ECB

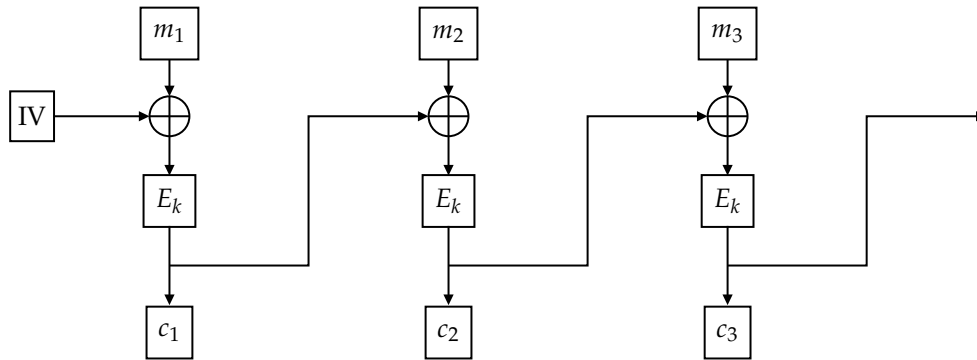
encryption is not (semantically) secure against chosen-plaintext attacks.

Suppose that the adversary has a challenge ciphertext c of length t blocks. The adversary would like to learn something about the plaintext m that corresponds to c . The adversary can select a plaintext m' of block length t , and obtain from Alice the encryption c' of m' . Now if c' equals c , then the adversary can conclude that m equals m' . On the other hand, if c' is not equal to c , then the adversary can conclude that m is not equal to m' . In either case, the adversary has learned something about the unknown plaintext m , and thus has broken semantic security.

The main problem with ECB mode is that the encryption process is *deterministic*. And so, for security one must include randomization into the encryption process. One way to do that is using *Cipher Block Chaining* or CBC mode.

Cipher Block Chaining (CBC) Mode

Encryption: Select $c_0 \in_R \{0,1\}^L$ (c_0 is a random non-secret IV). Then compute $c_i = E_k(m_i \oplus c_{i-1})$, $i = 1, 2, \dots, t$.



Ciphertext is $(c_0, c_1, c_2, \dots, c_t)$.

Decryption: $m_i = E_k^{-1}(c_i) \oplus c_{i-1}$, $i = 1, 2, \dots, t$.

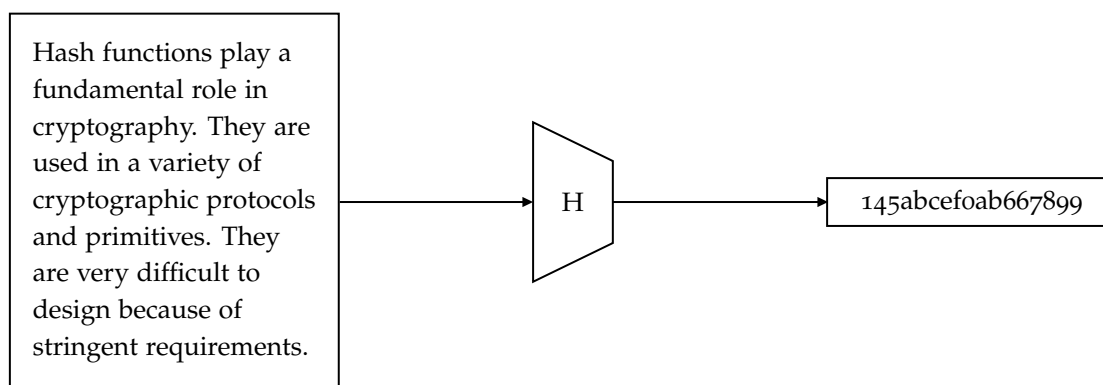
Note that Alice selects a new random IV c_0 for every plaintext m that she encrypts. Identical plaintexts with different IVs result in different ciphertexts and for this reason CBC encryption is (semantically) secure against chosen-plaintext attacks (for a well chosen block cipher E).

Hash Functions

3.1 Introduction

Hash functions play a fundamental role in cryptography. They are used in a variety of cryptographic primitives and protocols. They are very difficult to design because of very stringent security and performance requirements. Over the years, many hash functions have been broken. The main candidates available today are: SHA-1; SHA-2 family: SHA-224, *SHA-256*, SHA-384, SHA-512

What is a Hash Function? Informally speaking, a hash function is a fixed public function that takes as input messages of any length and outputs a hash value of a fixed length. In some sense the output serves as a fingerprint of the much longer input.



See these websites

- <http://www.xorbin.com/tools/md5-hash-calculator>
- <http://www.xorbin.com/tools/sha1-hash-calculator>
- <http://www.xorbin.com/tools/sha256-hash-calculator>

for calculators for the hash functions MD5, SHA-1 and SHA-256.

hash function

A **hash function** is a mapping H such that:

1. H maps binary messages of arbitrary lengths $\leq L$ to outputs of a fixed length n :

$$H : \{0,1\}^{\leq L} \rightarrow \{0,1\}^n.$$

(L is large, e.g., $L = 2^{64}$; n is small, e.g., $n = 256$)

2. $H(x)$ can be efficiently computed for all $x \in \{0,1\}^{\leq L}$.

H is called an n -bit hash function. $H(x)$ is called the hash or message digest of x .

Note:

The description of a hash function is public and fixed. There are no secret keys.

For simplicity, we will usually write $\{0,1\}^*$ instead of $\{0,1\}^{\leq L}$.

Toy Hash Function

$$H : \{0,1\}^{\leq 4} \rightarrow \{0,1\}^2$$

x	$H(x)$	x	$H(x)$	x	$H(x)$	x	$H(x)$
0	00	1	01				
00	11	01	01	10	01	11	00
000	00	001	10	010	11	011	11
100	11	101	01	110	01	111	10
0000	00	0001	11	0010	11	0011	00
0100	01	0101	10	0110	10	0111	01
1000	11	1001	01	1010	00	1011	01
1100	10	1101	00	1110	00	1111	11

(00, 1000) is a *collision* because these two messages have the same hash value, namely 11.

1001 is a *preimage* of 01 since the hash of 1001 is 01.

10 is a *second preimage* of 1011 since the hash of 1011 is 01 and hash of 10 is also 01.

Example: SHA-256

$$\text{SHA-256: } \{0,1\}^* \rightarrow \{0,1\}^{256}$$

$$\text{SHA-256}(\text{"Hello there"}) =$$

0x4e47826698bb4630fb4451010062fadbf85d61427cbdfaed7ad0f23f239bed89

$$\text{SHA-256}(\text{"Hello There"}) =$$

0xabf5dacd019d2229174f1daa9e62852554ab1b955fe6ae6bbbb214bab611f6f5

$$\text{SHA-256}(\text{"Welcome to CO 487"}) =$$

0x819ba4b1e568e516738b48d15b568952d4a35ea73f801c873907d3ae1f5546fb

$$\text{SHA-256}(\text{"Welcome to CO 687"}) =$$

0x404fb0ee527b8f9f01c337e915e8beb6e03983cfd9544296b8cf0e09c9d8753d

So the idea behind hash functions is that for each message, the hash function outputs a hash value that appears random and has no apparent resemblance to the original message.

3.1.1 Hash Functions from Block Ciphers

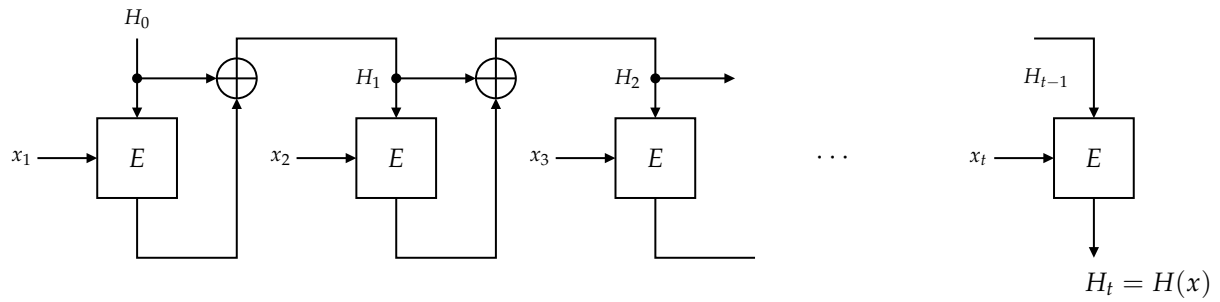
Davies-Meyer hash function

Let E_k be an m -bit block cipher with n -bit key k . Let IV be a fixed m -bit initializing value.

To compute $H(x)$, do:

Algorithm 10: Davies-Meyer hash function

- 1 Break up $x \parallel 1$ into n -bit blocks: $\bar{x} = x_1, x_2, \dots, x_t$, padding out the last block with 0 bits if necessary.
 - 2 $H_0 := IV$
 - 3 **for** $i = 1, \dots, t$ **do**
 - 4 \lfloor Compute $H_i = E_{x_i}(H_{i-1}) \oplus H_{i-1}$
 - 5 $H(x) := H_t$
-



Note that hashing x does not use any secret keying information. The blocks of the message x are used as the keys in the hashing process, and so anyone who knows the message x can compute its hash.

Hash functions are the Swiss Army knife of cryptography. Hash functions are used for all kinds of applications they were not designed for, and for this reason it's sometimes difficult to assess security of these applications. One reason for this widespread use of hash functions is *speed*.

3.1.2 Desirable security properties for hash functions

Preimage Resistance (PR)

$$H : \{0, 1\}^* \rightarrow \{0, 1\}^n.$$

preimage resistance hash function

Given a hash value $y \in_R \{0, 1\}^n$, it is computationally infeasible to find (with non-negligible probability of success) *any* $x \in \{0, 1\}^*$ such that $H(x) = y$.

Here x is called *a* preimage of y .

Note that $x \in_R S$ means that x is chosen independently and uniformly at random from S .

Preimage resistance is required if hash functions are used for *password protection* on a multi-user computer system: Server stores $(\text{userid}, H(\text{password}))$ in a password file. Thus, if an attacker gets a copy of the password file, she does not learn any passwords. She only learns the hashes of the passwords.

2nd Preimage Resistance (2PR)

$$H : \{0,1\}^* \rightarrow \{0,1\}^n.$$

second-preimage resistance hash function

Given an input $x \in_R \{0,1\}^*$, it is computationally infeasible to find (with non-negligible probability of success) a second input $x' \in \{0,1\}^*$, $x' \neq x$, such that $H(x') = H(x)$.

2PR is needed if hash functions are used as *Modification Detection Codes* (MDCs): To ensure that a message m is not modified by unauthorized means, one computes $H(m)$ and protects $H(m)$ from unauthorized modification. Protecting $H(m)$ might be easier than protecting m itself because $H(m)$ is in general much smaller than m .

Modification detection codes can be used for *virus protection*. A user Alice wishes to download a piece of software x from a website that might be untrusted. Suppose also that Alice has an authentic copy of the hash of x , which might have been obtained from the print copy of a trade magazine. So Alice will download x from the website, compute its hash, and compare the hash value with the hash from the trade magazine. If the two hash values are equal, then Alice installs the software on her computer. An adversary who wants Alice to install a piece of malware needs to format the malware x' so that the hash of x' equals the hash of x . In other words the adversary needs to find a second preimage for the message x , and so for this application we need H to be second-preimage resistant.

Collision Resistance (CR)

$$H : \{0,1\}^* \rightarrow \{0,1\}^n.$$

collision-resistant hash function

It is computationally infeasible to find (with non-negligible probability of success) two distinct inputs $x, x' \in \{0,1\}^*$ such that $H(x) = H(x')$.

The pair (x, x') is called a *collision* for H .

Note that H has many collisions since the domain is much larger in general than its codomain, and so by the pigeonhole principle there are many pairs of messages with the same hash value. So the issue isn't whether collisions exist, but whether one can find a collision *efficiently*.

Collision resistance is required when hash functions are used to compute *message digests* or *digital signature schemes*: For reasons of efficiency, instead of signing a (long) message, the (much shorter) message digest is signed. This action requires preimage-resistance, 2nd preimage resistance, and collision resistance. (More on this later)

To see why collision resistance is required: Suppose that the legitimate signer Alice can find two messages x_1 and x_2 , with $x_1 \neq x_2$ and $H(x_1) = H(x_2)$. Alice can sign x_1 and later claim to have signed x_2 . In this way, Alice might be able to repudiate her signature on x_1 .

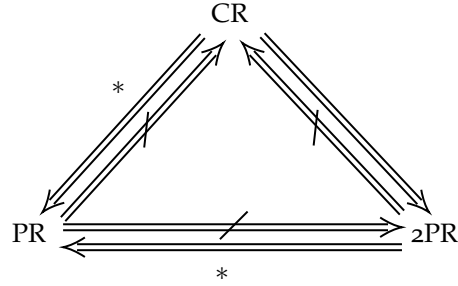
Later on in the course, we'll see many *other applications* of hash functions:

1. Message Authentication Codes (HMAC). [More on this later]
2. Pseudorandom bit generation: Distilling random bits s from several "pseudorandom" sources x_1, \dots, x_t . Output $s = H(x_1, \dots, x_t)$.
3. Key derivation function (KDF): Deriving a cryptographic key from a shared secret. [More later]

4. Proof-of-work in cryptocurrencies (Bitcoin): [More on this later]
5. Quantum-safe signature schemes: [More on this later]

3.1.3 Relationships between PR, 2PR, CR

Let $H : \{0,1\}^* \rightarrow \{0,1\}^n$ be a hash function.



* for somewhat uniform hash functions

1. CR implies 2PR

Proof:

Suppose that H is not 2PR.

Select $x \in_R \{0,1\}^*$. Since H is not 2PR, we can efficiently find $x' \in \{0,1\}^*$, $x' \neq x$, with the same hash value $H(x') = H(x)$. Thus we have efficiently found a collision (x, x') for H .

Hence H is not CR. □

Note that this proof establishes the contrapositive statement.

2. 2PR does not guarantee CR

Proof:

Suppose that $H : \{0,1\}^* \rightarrow \{0,1\}^n$ is 2PR.

Consider $\bar{H} : \{0,1\}^* \rightarrow \{0,1\}^n$ defined by:

$$\bar{H}(x) = \begin{cases} H(0), & \text{if } x = 1, \\ H(x), & \text{if } x \neq 1. \end{cases}$$

Thus we have $\bar{H}(1) = H(0) = \bar{H}(0)$.

Suppose that \bar{H} is not 2PR. So, given $x \in_R \{0,1\}^*$, we can efficiently find $x' \in \{0,1\}^*$, $x' \neq x$, with $\bar{H}(x') = \bar{H}(x)$. With probability essentially 1, we can assume that $x \neq 0,1$. Hence $\bar{H}(x) = H(x)$.

Now, if $x' \neq 1$, then $\bar{H}(x') = H(x') = H(x)$; while if $x' = 1$, then $\bar{H}(x') = \bar{H}(1) = H(0) = H(x)$. In either case, we have found a second preimage for x with respect to H . This contradicts the assumption that H is second preimage resistant. So, \bar{H} must be 2PR.

Also, \bar{H} is not CR, since $(0,1)$ is a collision for \bar{H} .

Hence \bar{H} is second preimage resistant but not collision resistant. □

3. CR does not guarantee PR

Proof:

Suppose that $H : \{0,1\}^* \rightarrow \{0,1\}^n$ is CR.

Consider $\bar{H} : \{0,1\}^* \rightarrow \{0,1\}^{n+1}$ defined by:

$$\bar{H}(x) = \begin{cases} 1 \parallel x, & \text{if } x \in \{0,1\}^n, \\ 0 \parallel H(x), & \text{if } x \notin \{0,1\}^n. \end{cases}$$

Then we claim that \bar{H} is CR.

First note that hash values that begin with the 1 have unique preimages, and hence there does not exist a collision for \bar{H} where the colliding messages have hash values that begin with a 1.

Suppose next that we can efficiently find a collision for \bar{H} where the colliding messages have hash values that begin with a 0. But then this collision for \bar{H} is also a collision for H that we have officially found. This contradicts the assumption that H is collision resistant, and so we conclude that \bar{H} is also collision resistant.

And, \bar{H} is not PR since preimages can be efficiently found for at least half of all $y \in \{0,1\}^{n+1}$. \square

However, if H is “somewhat uniform” (i.e., all hash values have roughly the same number of preimages), then CR does imply PR.

Proof:

Suppose H is a somewhat uniform hash function that is not PR. Select $x \in_R \{0,1\}^*$ and compute $y = H(x)$. Since H is somewhat uniform, y is also a randomly selected hash value.

Since H is not PR, we can efficiently find a preimage x' of y . Since H is somewhat uniform, we expect that y has many preimages, and thus $x' \neq x$ with very high probability. Thus, (x, x') is a collision for H that we have efficiently found, so H is not CR. \square

Note that we shall henceforth assume that hash functions are somewhat uniform.

4. PR does not guarantee 2PR

Proof:

Suppose $H : \{0,1\}^* \rightarrow \{0,1\}^n$ is PR.

Define $\bar{H} : \{0,1\}^* \rightarrow \{0,1\}^n$ by

$$\bar{H}(x_1, x_2, \dots, x_t) = H(0, x_2, \dots, x_t).$$

Then we first claim \bar{H} is PR.

Suppose that \bar{H} is not preimage resistant. Then, given a random hash value y , we can efficiently find a message x such that $\bar{H}(x)$ equals y . But then $H(x') = y$ where x' is the same as x except that its first bit is now zero. And so we can efficiently find preimages for H . This contradicts the assumption that H is preimage resistant, and so we conclude that \bar{H} is also preimage resistant.

Then let's prove \bar{H} is not 2PR.

Suppose that we're given a random message x . Then we can easily find a second preimage for x with respect to \bar{H} , namely by flipping the first bit of x to get the message x' . And this shows that \bar{H} is not second preimage resistant.

So \bar{H} is a hash function that is preimage resistant but is not second preimage resistant. \square

5. 2PR implies PR (for somewhat uniform H)

Proof:

Exercise. □

Note that CR is the hardest property to satisfy because if a hash function is CR then it is also PR and also 2PR. On the other hand, PR does not imply CR and 2PR does not imply CR.

3.2 Generic Attacks

generic attack

A **generic attack** on a hash function $H : \{0,1\}^* \rightarrow \{0,1\}^n$ does not exploit any properties the specific hash function may have.

In the analysis of a generic attack, we view H as a *random selected function* in the sense that for each $x \in \{0,1\}^*$, we assume that the value $y = H(x)$ was chosen by selecting $y \in_R \{0,1\}^n$.

From a security point of view, a random function is an *ideal* hash function. However, random functions are not suitable for practical applications because they cannot be compactly stored.

3.2.1 Find Preimages

Here's a *generic attack for finding preimages* for an n -bit hash function $H : \{0,1\}^* \rightarrow \{0,1\}^n$.

Given $y \in_R \{0,1\}^n$, repeatedly select arbitrary $x \in \{0,1\}^*$ until $H(x) = y$.

For each message x , the probability that $H(x)$ equals the specific y is $1/2^n$ because the total number of hash values is 2^n and we're assuming that H is a randomly selected function. Hence the *expected number of steps* is $\approx 2^n$. (Here, a 'step' is a hash function evaluation.)

This attack is infeasible if $n \geq 128$.

Note:

It has been proven that this generic attack for finding preimages is optimal, i.e., no faster *generic* attack exists.

Of course, for a specific hash function, there might exist a preimage finding algorithm that is faster than the generic attack.

3.2.2 Find Collisions

Select arbitrary $x \in \{0,1\}^*$ and store $(H(x), x)$ in a table sorted by first entry. Continue until a collision is found.

Since there are 2^n possible hash values, by the birthday paradox we conclude that the *expected number of steps* is $\sqrt{\pi 2^n / 2} \approx \sqrt{2^n}$. (Here, a "step" is a hash function evaluation.)

This attack is infeasible if $n \geq 256$.

It has been proven that this generic attack for finding collisions is optimal in terms of the number of hash function evaluations.

A major draw back of this algorithm is the *expected space required*: $\sqrt{\pi 2^n / 2} \approx \sqrt{2^n}$.

If $n = 128$, then the expected running time is 2^{64} steps (feasible). However, the expected space required is 7×10^8 Tbytes (infeasible).

3.2.3 VW Parallel Collision Search

VW: Paul van Oorschot, Michael Wiener (1993)

Expected number of steps is $\sqrt{2^n}$. However, expected space required is negligible.

It is easy to parallelize: m -fold speedup with m processors.

The VW attack finds a random pair of n -bit messages with the same hash value. It might appear that such a collision is useless in practice. However, the VW attack can easily be modified to find “meaningful” collisions. See the details in [Parallel Collision Search with Cryptanalytic Applications](#).

Conclusion: If collision resistance is desired, then use an n -bit hash function with $n \geq 256$

Parallel collision search (VW Method)

Problem: Find a collision for $H : \{0, 1\}^* \rightarrow \{0, 1\}^n$.

Assumption: H is a random function.

Notation: Let $N = 2^n$.

Define a sequence $\{x_i\}_{i \geq 0}$ by

$$x_0 \in_R \{0, 1\}^n, \quad x_i = H(x_{i-1}) \text{ for } i \geq 1.$$



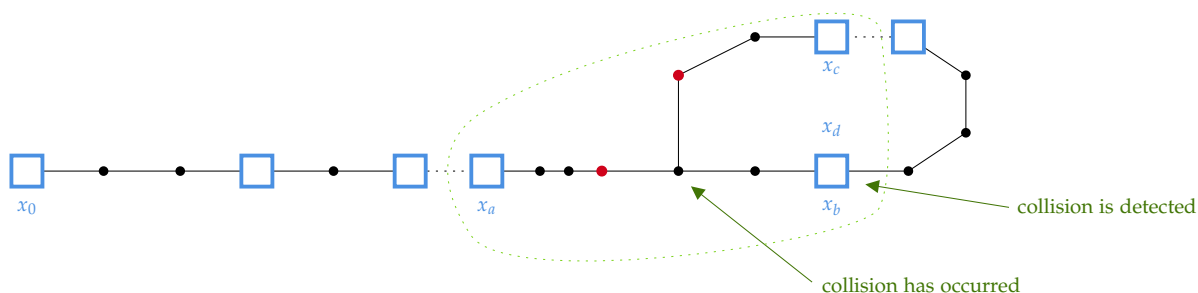
This sequence cannot be distinct forever, because the terms are drawn from a finite set, and so the sequence must eventually collide with itself. Let j be the smallest index for which $x_j = x_i$ for some $i < j$; such a j must exist. Then $x_{j+\ell} = x_{i+\ell}$ for all $\ell \geq 1$.

By the birthday paradox, $E[j] \approx \sqrt{\frac{\pi N}{2}} \approx \sqrt{N}$. In fact, $E[i] \approx \frac{1}{2}\sqrt{N}$ where i is the length of the tail of the sequence. And $E[j - i] \approx \frac{1}{2}\sqrt{N}$ where $j - i$ is the length of the cycle.

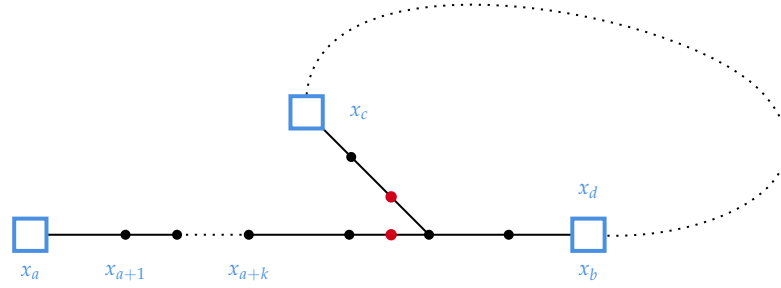
Now, $i \neq 0$ with overwhelming probability; in that event, (x_{i-1}, x_{j-1}) (red points) is a collision for H .

One way to find this collision is to compute the terms x_0, x_1, x_2, \dots and store them in a table. However the table would be too large, namely of size roughly \sqrt{N} . How to find (x_{i-1}, x_{j-1}) without using much storage? VW method's idea is to store only *distinguished points*.

Select an easily testable distinguishing property for elements of $\{0, 1\}^n$ (e.g., leading 32 bits are all 0). Let θ be the proportion of elements of $\{0, 1\}^n$ that are distinguished. In a nutshell, VW algorithm is to compute x_0, x_1, x_2, \dots and only store the points in the sequence that are distinguished.



We can expand the green portion of the diagram as follows:



Algorithm 11: VW collision finding

```

// Stage 1 (detecting a collision)
1 Select  $x_0 \in_R \{0,1\}^n$ 
2 Store  $(x_0, 0, -)$  in a sorted table
3  $LP \leftarrow x_0$  //  $LP =$  last point stored in the table
4 for  $d = 1, 2, 3, 4, \dots$  do
5   Compute  $x_d = H(x_{d-1})$ 
6   if  $x_d$  is distinguished then
7     if  $x_d$  is already in table then
8       // Say  $x_d = x_b$  where  $b < d$ 
9       Go to Stage 2.
10    else
11      Store  $(x_d, d, LP)$  in a table.
12       $LP \leftarrow x_d$ 

// Stage 2 (finding a collision)
12  $\ell_1 \leftarrow b - d, \ell_2 \leftarrow d - c$ 
13 // WLOG suppose  $\ell_1 \geq \ell_2$ 
14  $k \leftarrow \ell_1 - \ell_2$ 
15 Compute  $x_{a+1}, x_{a+2}, \dots, x_{a+k}$ .
16 for  $m = 1, 2, 3, \dots$  do
17   Compute  $(x_{a+k+m}, x_{c+m})$ 
18 until  $x_{a+k+m} = x_{c+m}$ 
19 The collision is  $(x_{a+k+m-1}, x_{c+m-1})$ 
  
```

Let's analyze the VW attack.

- Stage 1: Expected # of H -evaluations is

$$\underbrace{\sqrt{\frac{\pi N}{2}}}_{\text{collision occurs}} + \underbrace{\frac{1}{\theta}}_{\text{collision is detected}} \approx \sqrt{N} + \frac{1}{\theta}$$

- Stage 2: Expected # of H -evaluations is $\leq \frac{3}{\theta}$ (see optional readings)

Overall expected running time: $\boxed{\sqrt{N} + \frac{4}{\theta}}$

Storage: $\approx \theta\sqrt{N}(3n)$ bits (each table entry has bitlength $3n$)

Let's revisit the example when $n = 128$. Take $\theta = \frac{1}{2^{32}}$. Then the expected time of VW collision search

is 2^{64} H -evaluations (feasible), and the expected storage is 241 Gbytes (free).

Parallelizing VW collision search

Suppose we had m processors available to us. Run a copy of VW on each of the m processors. Each processor would report distinguished points to a central server, and the central server would check for repeated distinguish points.

Note that each processor begins with its own random starting point, and so each processor computes a different sequence. Each sequence can now collide either with itself or with another sequence as shown here. When two sequences from different processors collide then we have a collision of n -bit strings. This collision is detected by the central server because this distinguished point will be computed twice, once by this processor and once by this processor. When a collision is detected then the collision can be found as before using Stage 2 of the VW attack.

Expected time $\approx \sqrt{N}/m + 4/\theta$. Expected storage $\approx \theta\sqrt{N}(3n)$ bits

Note that we have factor- m speedup. No communication between processors. Processors only occasionally communicate with central server. So VW parallelizes very well.

3.3 Iterated Hash Functions (Merkle Meta-Method)

Iterated hash functions has two components:

- Fixed initializing value $IV \in \{0,1\}^n$,
- Compression function $f : \{0,1\}^{n+r} \rightarrow \{0,1\}^n$ (efficiently computable).

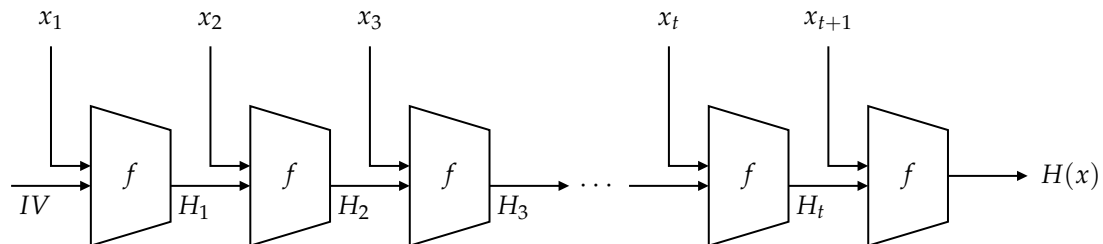
To compute $H(x)$ where x has bitlength $b < 2^r$ do:

Algorithm 12: Merkle's hash function

- 1 Break up x into r -bit blocks: $\bar{x} = x_1, x_2, \dots, x_t$, padding out the last block with 0 bits if necessary.
 - 2 Define x_{t+1} , the *length-block*, to hold the right-justified binary representation of b .
 - 3 $H_0 := IV$
 - 4 **for** $i = 1, 2, \dots, t+1$ **do**
 - 5 Compute $H_i = f(H_{i-1}, x_i)$
 - 6 $H(x) := H_{t+1}$
-

The H_i 's are called *chaining variables*.

Here is a diagram that illustrates Merkle's hash function construction.



3.3.1 Collision Resistance of Iterated Hash Functions

Merkle's Theorem

If the compression function f is collision resistant, then the hash function H is also collision resistant.

Merkle's theorem reduces the problem of designing collision-resistant hash functions to that of designing collision-resistant compression functions.

Proof:

Suppose H is not CR. Then we can *efficiently* find $x, x' \in \{0, 1\}^*$, $x \neq x'$, with $H(x) = H(x')$.

Let

$$\begin{aligned}\bar{x} &= x_1, x_2, \dots, x_t, & b &= \text{bitlength}(x), & x_{t+1} &= \text{length block} \\ \bar{x}' &= x'_1, x'_2, \dots, x'_{t'}, & b' &= \text{bitlength}(x'), & x'_{t'+1} &= \text{length block}.\end{aligned}$$

We efficiently compute:

$$\begin{array}{ll} H_0 = IV & H_0 = IV \\ H_1 = f(H_0, x_1) & H'_1 = f(H_0, x'_1) \\ H_2 = f(H_1, x_2) & H'_2 = f(H'_1, x'_2) \\ \vdots & \vdots \\ H_{t-1} = f(H_{t-2}, x_{t-1}) & H'_{t'-1} = f(H'_{t'-2}, x'_{t'-1}) \\ H_t = f(H_{t-1}, x_t) & H'_{t'} = f(H'_{t'-1}, x'_{t'}) \\ H_{t+1} = f(H_t, x_{t+1}) & H'_{t'+1} = f(H'_{t'}, x'_{t'+1}) \end{array}$$

Since $H(x) = H(x')$, we have $H_{t+1} = H'_{t'+1}$.

Now if $b \neq b'$, then $x_{t+1} \neq x'_{t'+1}$. Thus $(H_t, x_{t+1}), (H'_{t'}, x'_{t'+1})$ is a collision for f that we have efficiently found. Hence f is not collision resistant.

Suppose $b = b'$. Then $t = t'$ and $x_{t+1} = x'_{t'+1}$.

Let i be the largest index, $i \in [0, t]$, for which $(H_i, x_{i+1}) \neq (H'_i, x'_{i+1})$. Such i must exist since $x \neq x'$. Then

$$H_{i+1} = f(H_i, x_{i+1}) = f(H'_i, x'_{i+1}) = H'_{i+1},$$

so $(H_i, x_{i+1}), (H'_i, x'_{i+1})$ is a collision for f that we have efficiently found. Hence f is not CR. \square

3.3.2 Provable Security

A major theme of cryptographic research is to formulate precise security definitions and assumptions, and then *prove* that a cryptographic protocol is secure. A *proof of security* is certainly desirable since it rules out the possibility of attacks being discovered in the future. However, it isn't always easy to assess the practical security assurances (if any) that a security proof provides.

- The assumptions might be unrealistic, or false, or circular.
- The security proof might be fallacious.
- The security model might not account for certain kinds of realistic attacks.

- The security proof might be asymptotic.
- The security proof might have a large tightness gap.

Optional reading: <http://anotherlook.ca>

3.4 MDx-Family of Hash Functions

MDx is a family of iterated hash functions.

MD4

MD4 was proposed by Ron Rivest in 1990. MD4 has 128-bit outputs. (VW finds an MD4 collision in time 2^{64}) Wang et al. (2004) found collisions for MD4 by hand. So this showed that MD4 was completely insecure from the point of view of collision resistance. Leurent (2008) discovered an algorithm for finding MD4 preimages in 2^{102} steps.

MD5

MD5 is a strengthened version of MD4. Designed by Ron Rivest in 1991. MD5 has 128-bit outputs.

Wang and Yu (2004) found MD5 collisions in 2^{39} steps. MD5 collisions can now be found in 2^{24} steps (in a few seconds on a laptop computer). Sasaki & Aoki (2009) discovered a method for finding preimages for MD5 in $2^{123.4}$ steps.

Therefore, MD5 should not be used if collision resistance is required, but is probably okay as a preimage-resistant hash function.

MD5 is still used today. The reason is that MD5 was widely deployed in the 1990s and the 2000s and it's very expensive and cumbersome to replace all these implementations. For example, in 2006, MD5 is implemented more than 850 times in Microsoft Windows source code. In 2014, Microsoft issued a patch that restricts the use of certificates with MD5 in Windows: <http://tinyurl.com/MicrosoftMD5>. The reason for the patch was the discovery of the Flame malware.

Flame Malware

See [https://en.wikipedia.org/wiki/Flame_\(malware\)](https://en.wikipedia.org/wiki/Flame_(malware))

Discovered in May 2012. Highly sophisticated espionage tool. Targeted computers in Iran and the Middle East. Suspected to originate from the US and/or Israeli government; US government has denied all responsibility. Contains a forged Microsoft certificate for Windows code signing. Forged certificate used a new, "zero-day MD5 chosen-prefix" collision attack. Microsoft no longer allows the use of MD5 for code signing.

3.5 SHA

3.5.1 SHA-1

Secure Hash Algorithm (SHA) was designed by NSA and published by NIST in 1993 (FIPS 180). 160-bit iterated hash function, based on MD4. Slightly modified to SHA-1 (FIPS 180-1) in 1994 in order to fix an (undisclosed) security weakness. Wang et al. (2005) found collisions for SHA in 239 steps. Wang et al. (2005) discovered a collision-finding algorithm for SHA-1 that takes 263 steps. The first SHA-1 collision was found on February 23, 2017. No preimage or 2nd preimage attacks (that are faster than the generic attacks) are known for SHA-1.

Microsoft's SHA-1 Plan

See <http://tinyurl.com/MicrosoftSHA1>. As of May 9, 2017: TLS server-authentication certificates that use SHA-1 will be considered invalid. However, Microsoft still permits the use of SHA-1 in Code signature file hashes, Code signing certificates, Timestamp signature hashes, Timestamping certificates, etc.

3.5.2 SHA-2 Family

In 2001, NSA proposed variable output-length versions of SHA-1.

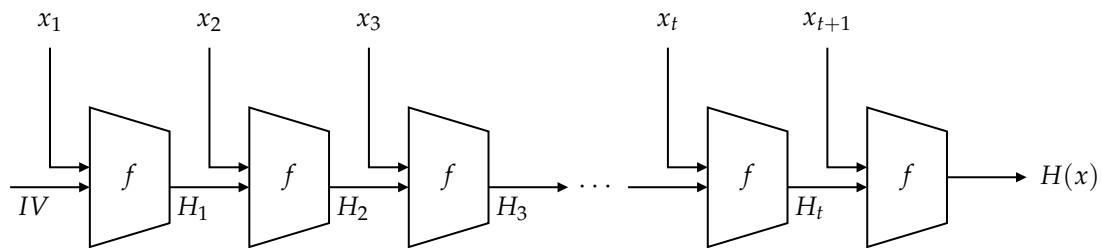
Output lengths are 224 bits (SHA-224 and SHA-512/224), 256 bits (SHA-256 and SHA-512/256), 384 bits (SHA-384) and 512 bits (SHA-512).

2020: No weaknesses in any of these hash functions have been found.

Note: The security levels of these hash functions against collision-finding attacks is the same as the security levels of Triple-DES, AES-128, AES-192 and AES-256 against exhaustive key search attacks. For example, the VW attack on SHA-224 can find a collision in 2^{112} operations, whereas the meet-in-the-middle attack on Triple-DES finds a secret key in 2^{112} steps.

The SHA-2 hash functions are standardized in FIPS 180-2.

3.5.3 Description of SHA-256



SHA-256 is an iterated hash function (Merkle meta-method).

$n = 256, r = 512$.

Compression function is $f : \{0, 1\}^{256+512} \rightarrow \{0, 1\}^{256}$.

Input: bitstring x of arbitrary bitlength $b \geq 0$.

Output: 256-bit hash value $H(x)$ of x .

SHA-256 Notation

A, B, C, D, E, F, G, H	32-bit words.
$+$	addition modulo 2^{32} .
\bar{A}	bitwise complement.
$A \gg s$	shift A right through s positions.
$A \hookrightarrow s$	rotate A right through s positions.
AB	bitwise AND.
$A \oplus B$	bitwise exclusive-OR.
$f(A, B, C)$	$AB \oplus \bar{A}C$.
$g(A, B, C)$	$AB \oplus AC \oplus BC$.
$r_1(A)$	$(A \hookrightarrow 2) \oplus (A \hookrightarrow 13) \oplus (A \hookrightarrow 22)$.
$r_2(A)$	$(A \hookrightarrow 6) \oplus (A \hookrightarrow 11) \oplus (A \hookrightarrow 25)$.
$r_3(A)$	$(A \hookrightarrow 7) \oplus (A \hookrightarrow 18) \oplus (A \gg 3)$.
$r_4(A)$	$(A \hookrightarrow 17) \oplus (A \hookrightarrow 19) \oplus (A \gg 10)$.

SHA-256 Constants

32-bit initial chaining values (IVs):

$$h_1 = 0x6a09e667, \quad h_2 = 0xbb67ae85, \quad h_3 = 0x3c6ef372, \quad h_4 = 0xa54ff53a, \\ h_5 = 0x510e527f, \quad h_6 = 0x9b05688c, \quad h_7 = 0x1f83d9ab, \quad h_8 = 0x5be0cd19.$$

These words were obtained by taking the first 32 bits of the fractional parts of the square roots of the first 8 prime numbers.

Per-round integer additive constants:

$$y_0 = 0x428a2f98, \quad y_1 = 0x71374491, \quad y_2 = 0xb5c0fbcf, \quad y_3 = 0xe9b5dba5, \\ \dots \quad \dots \quad y_{62} = 0xbef9a3f7, \quad y_{63} = 0xc67178f2.$$

These words were obtained by taking the first 32 bits of the fractional parts of the cube roots of the first 64 prime numbers.

Algorithm 13: SHA-256 Preprocessing

-
- 1 Pad x (with 1 followed by as few 0's as possible) so that its bitlength is 64 less than a multiple of 512.
 - 2 Append a 64-bit representation of $b \bmod 2^{64}$.
 - 3 The formatted input is $x_0, x_1, \dots, x_{16m-1}$, where each x_i is a 32-bit word.
 - 4 Initialize chaining variables: $(H_1, H_2, \dots, H_6, H_7, H_8) \leftarrow (h_1, h_2, \dots, h_6, h_7, h_8)$.
 - 5 **foreach** $i = 0 \dots m - 1$ **do**
 - 6 Copy the i -th block of sixteen 32-bit words into temporary storage: $X_j \leftarrow x_{16i+j}, 0 \leq j \leq 15$.
 // Expand the 16-word block into a 64-word block:
 - 7 **for** $j = 16 \dots 63$ **do**
 - 8 $X_j \leftarrow r_4(X_{j-2}) + X_{j-7} + r_3(X_{j-15}) + X_{j-16}$.
 // Initialize working variables
 - 9 $(A, B, \dots, F, G, H) \leftarrow (H_1, H_2, \dots, H_6, H_7, H_8)$.
 - 10 **for** $j = 0 \dots 63$ **do**
 - 11 $T_1 \leftarrow H + r_2(E) + f(E, F, G) + y_j + X_j$
 - 12 $T_2 \leftarrow r_1(A) + g(A, B, C)$
 - 13 $H \leftarrow G, \quad G \leftarrow F, \quad F \leftarrow E, \quad E \leftarrow D + T_1$
 - 14 $D \leftarrow C, \quad C \leftarrow B, \quad B \leftarrow A, \quad A \leftarrow T_1 + T_2$.
 - 15 Update chaining values: $(H_1, H_2, \dots, H_7, H_8) \leftarrow (H_1 + A, H_2 + B, \dots, H_7 + G, H_8 + H)$.
-

Output: $\text{SHA-256}(x) = H_1 \parallel H_2 \parallel H_3 \parallel H_4 \parallel H_5 \parallel H_5 \parallel H_6 \parallel H_7 \parallel H_8$.

Performance: Speed benchmarks (2017) for software implementations on an Intel Core i9 2.9 GHz 6-core Coffee Lake (8950HK) using OpenSSL 1.1.1d. SHA-256 is quite fast in software. It's not as fast as MD5 or SHA-1, but that's probably just as well since MD5 and SHA-1 are considered to be insecure as far as collision resistance is concerned.

3.5.4 SHA-3

The SHA-2 design is similar to SHA-1, and thus there are lingering concerns that the SHA-1 weaknesses could eventually extend to SHA-2.

SHA-3: NIST hash function competition. 64 candidates submitted by Oct 31 2008 deadline.

2012: Keccak was selected as the winner. Keccak uses the “sponge construction” and *not* the Merkle iterated hash design. SHA-3 is being used in practice, but is not as widely deployed as SHA-2.

3.5.5 NIST's Policy on Hash Functions

August 5, 2015. See: <http://csrc.nist.gov/groups/ST/hash/policy.html>

Should stop using SHA-1 for digital signatures and other applications that require collision resistance. May still use SHA-1 for HMAC, KDFs, and random number generators. May use SHA-2 for all applications that employ secure hash algorithms. SHA-3 may also be used, but this is not required.

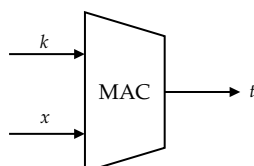
Message authentication code schemes

4.1 Introduction

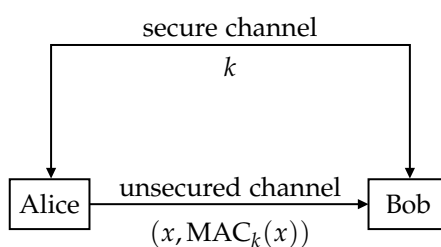
message authentication code

A **message authentication code** scheme is a family of functions $\text{MAC}_k : \{0,1\}^* \rightarrow \{0,1\}^n$ parameterized by an ℓ -bit key k , where each function MAC_k can be efficiently computed.

$t = \text{MAC}_k(x)$ is called the *MAC* or *tag* of x with key k .



MAC schemes are used for providing (symmetric-key) data integrity and data origin authentication.



To provide data integrity and data origin authentication:

1. Alice and Bob establish a secret key $k \in \{0,1\}^\ell$.
2. Alice computes tag $t = \text{MAC}_k(x)$ and sends (x, t) to Bob.
3. Bob verifies that $t = \text{MAC}_k(x)$.

Note:

No confidentiality since the message x is sent in the clear. If confidentiality is required, then one has to properly combine an encryption scheme with a MAC scheme.

Non-repudiation. This is because when Bob receives a tag message from Alice and the tag is valid,

then Bob is convinced that the message came from Alice. However, Bob cannot convince a third party that the message came from Alice because Bob could have generated the message and its tag himself. This is the consequence of Alice and Bob sharing the same secret key.

An adversary who captures the message and tag can *replay* this tag message to Bob at a future point in time and Bob will accept the tagged message. To avoid replay, add a timestamp or sequence number.

4.1.1 Security Definition

Let K be the secret key shared by Alice and Bob.

The adversary does not know k , but is allowed to obtain (from Alice or Bob) tags for messages of her choosing. The adversary's goal is to obtain the tag of *any* new message, i.e., a message whose tag she did not already obtain from Alice or Bob.

secure MAC scheme

A MAC scheme is **secure** if given some tags $\text{MAC}_k(x_i)$ for x_i 's of one's own choosing, it is computationally infeasible (with non-negligible probability of success) a message-tag pair $(x, \text{MAC}_k(x))$ for any new message x .

More concisely, a MAC scheme is secure if it is existentially unforgeable against chosen-message attack.

Note: A secure MAC scheme can be used to provide *data integrity* and *data origin authentication*.

An *ideal* MAC scheme has the following property:

For each key $k \in \{0,1\}^\ell$, the function $\text{MAC}_k : \{0,1\}^* \rightarrow \{0,1\}^n$ is a random function.

Ideal MAC schemes are useless in practice. However, when analyzing a generic attack on a MAC scheme, it is reasonable to assume that the MAC scheme is ideal.

4.2 Generic Attacks on MAC schemes

Guessing the MAC of a message $x \in \{0,1\}^*$:

Select $y \in_R \{0,1\}^n$ and guess that $\text{MAC}_k(x) = y$.

Assuming that MAC_k is random function, the probability of success is $1/2^n$. Note that the adversary cannot check her guess directly because the adversary does not know the key k . MAC guessing is infeasible if $n \geq 128$.

Exhaustive search on the key space:

Given r known message-tag pairs $(x_1, t_1), \dots, (x_r, t_r)$, one can check whether a guess k of the key is correct by verifying that $\text{MAC}_k(x_i) = t_i$, for $i = 1, 2, \dots, r$.

Assume that MAC_k 's are random functions, the expected number of keys for which the tags verify is

$$1 + FK = 1 + (2^\ell - 1)/2^{nr}$$

For example, if $\ell = 128$, $n = 128$, $r = 2$, then $FK \approx 1/2^{128}$.

Expected number of steps $\approx 2^\ell$. Exhaustive search is infeasible if $\ell \geq 128$.

4.3 MACs Based on Block Ciphers

CBC-MAC

Let E be an n -bit block cipher with key space $\{0,1\}^\ell$.

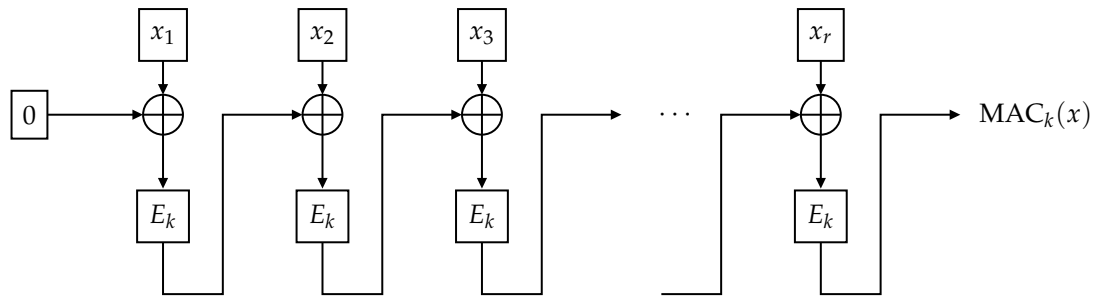
Assumption: Suppose that plaintext messages all have lengths that are multiplies of n , if not then we would add some padding to the message.

To compute $\text{MAC}_k(x)$:

1. Divide x into n -bit blocks x_1, x_2, \dots, x_r .
2. Compute $H_1 = E_k(x_1)$.
3. For $2 \leq i \leq r$, compute $H_i = E_k(H_{i-1} \oplus x_i)$.
4. Then $\text{MAC}_k(x) = H_r$.

Algorithm 14: CBC-MAC

- 1 Divide x into n -bit blocks x_1, x_2, \dots, x_r .
 - 2 Compute $H_1 = E_k(x_1)$.
 - 3 **for** $i = 2, \dots, r$ **do**
 - 4 | Compute $H_i = E_k(H_{i-1} \oplus x_i)$.
 - 5 $\text{MAC}_k(x) := H_r$
-



4.3.1 Security of CBC-MAC

CBC MAC comes with a rigorous security analysis from 1994 [Bellare, Kilian & Rogaway 1994]. Here's an informal statement of a Theorem:

Informal statement of a Theorem

Suppose that E is an "ideal" encryption scheme. (That is, for each $k \in \{0,1\}^\ell$, $E_k : \{0,1\}^n \rightarrow \{0,1\}^n$ is a 'random' permutation.) Then CBC-MAC *with fixed-length inputs* is a secure MAC scheme.

CBC-MAC (as described above without additional measures) is not secure if *variable length* messages are allowed. Here is a *chosen-message attack* on CBC-MAC:

1. Select an arbitrary n -bit block x_1 .
2. Obtain the tag t_1 of the one-block message x_1 (so $t_1 = E_k(x_1)$).
3. Obtain the tag t_2 of the one-block message t_1 (so $t_2 = E_k(t_1)$).
4. Then t_2 is the tag of the 2-block message $(x_1, 0)$ since $t_2 = E_k(0 \oplus E_k(x_1)) = E_k(E_k(x_1)) = E_k(t_1)$.

This attack might not be very realistic, but nonetheless it illustrates that CBC MAC does not meet our very strong notion of security, namely existential unforgeability against chosen message attack.

4.3.2 Encrypted CBC-MAC (EMAC)

One countermeasure for variable-length messages is **Encrypted CBC-MAC**:



Encrypt the last block under a second key s : $\text{EMAC}_{k,s}(x) = E_s(H_r)$, where $H_r = \text{CBC-MAC}_k(x)$.

EMAC has a rigorous security analysis from the year 2000. [Petrack & Rackoff 2000]:

Informal statement of a Theorem

Suppose that E is an “ideal” encryption scheme. Then EMAC is a secure MAC scheme (for inputs of any length).

4.4 MACs Based on Hash Functions

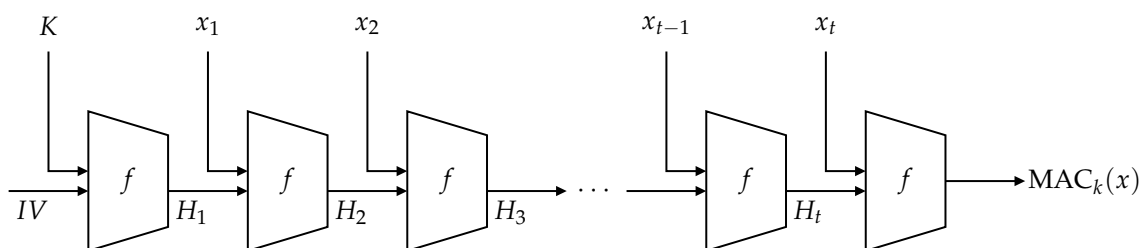
Hash functions were not originally designed for message authentication; in particular they are not “keyed” primitives.

Question How to use them to construct secure MACs?

Let H be an iterated n -bit hash function (without the length-block). Let $n + r$ be the input blocklength of the compression function $f : \{0, 1\}^{n+r} \rightarrow \{0, 1\}^n$. For example, for SHA-256, $n = 256, r = 512$. Let $k \in \{0, 1\}^n$. Let K denote k padded with $(r - n)$ 0's. So K has bitlength r . Thus $K = \underbrace{\leftarrow 0 \rightarrow}_{r-n} \underbrace{\leftarrow k \rightarrow}_n$.

4.4.1 Secret Prefix Method

MAC definition: $\text{MAC}_k(x) = H(K, x)$.



This is *insecure*. Here is a *length extension attack*: suppose $(x, \text{MAC}_k(x))$ is known. Suppose the bitlength of x is a multiple of r . Then $\text{MAC}_k(x \parallel y)$ can be computed for any y (without knowledge of k).

Also insecure if a length block is postpended to $K \parallel x$ prior to application of H .

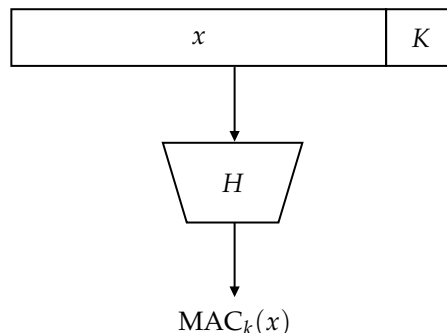
4.4.2 Secret Suffix Method

MAC definition: $\text{MAC}_k(x) = H(x, K)$.

The attack on the secret prefix method does not work here.

Suppose that a collision (x_1, x_2) can be found for H (i.e., $H(x_1) = H(x_2)$). We assume that x_1 and x_2 both have bitlengths that are multiples of r .

Thus $H(x_1, K) = H(x_2, K)$, and so $\text{MAC}_k(x_1) = \text{MAC}_k(x_2)$. Then the MAC for x_1 can be requested, giving the MAC for x_2 . Hence if H is not CR, then the secret suffix method MAC is *insecure*.

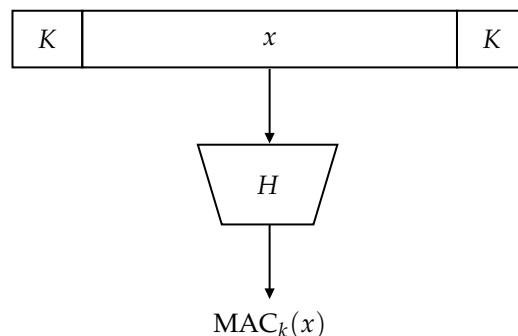


4.4.3 Envelope Method

MAC definition: $\text{MAC}_k(x) = H(K, x, K)$.

The MAC key is used both at the start and end of the MAC computation. Thus attacks on the secret prefix method and the secret suffix method do not work here because the key block is appended to the left and to the right of the message x .

The envelope method appears to be secure (i.e., no serious attacks have been found).



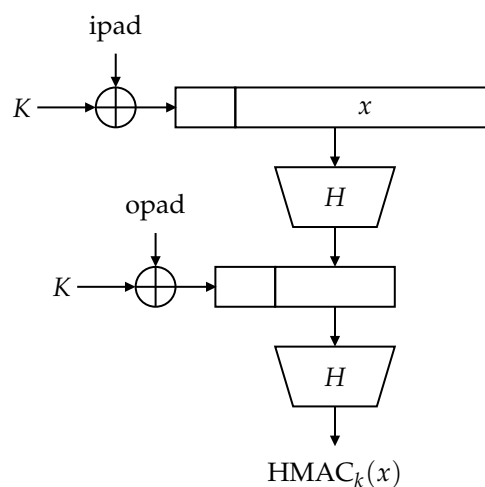
4.4.4 HMAC

MAC definition: $\text{HMAC}_k(x) = H(K \oplus \text{opad}, H(K \oplus \text{ipad}, x))$.

The most commonly used MAC scheme today is HMAC. It is “Hash-based” MAC and was designed by Bellare, Canetti & Krawczyk (1996).

Define two r -bit strings (in hexadecimal notation): $\text{ipad} = 0x36$, $\text{opad} = 0x5C$; each repeated $r/8$ times. These bytes $0x36$ and $0x5C$ are arbitrarily chosen. The important thing is that they be fixed and different from each other.

The main part of the HMAC algorithm is the secret prefix method, which is highly insecure. However since the key block is appended to the left of the resulting hash value and then this two-block message is hashed once more, the length extension attack is prevented.



HMAC comes with a rigorous security analysis.

Informal statement of a Theorem

Suppose that the compression function f used in H is a secure MAC with fixed length messages and a secret IV as the key. Then HMAC is a secure MAC algorithm.

In practice, one should use HMAC with the SHA-256 hash function. HMAC is specified in IETF¹ RFC 2104 and FIPS 198. HMAC is used in IPsec (Internet Protocol Security) and TLS.

Key Derivation Function

HMAC is commonly used as a *key derivation function* (KDF).

Suppose that Alice has a secret key k , and wishes to derive several *session keys* (e.g., to encrypt data in different communication sessions).

Alice computes $sk_1 = \text{HMAC}_k(1), sk_2 = \text{HMAC}_k(2), sk_3 = \text{HMAC}_k(3), \dots$

Rationale: Without knowledge of k , an adversary is unable to learn anything about any particular session key sk_j , even though it may have learnt some other session keys.

This works because HMAC is existentially unforgeable against chosen-message attack.

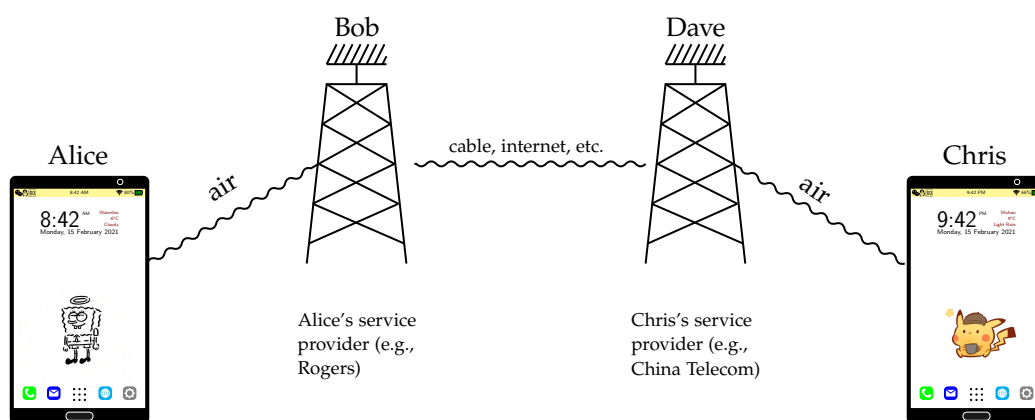
4.5 Case study: GSM

Global standards for mobile communications:

- 2G, 2.5G: GSM (Global System for Mobile Communication)
- 3G: UMTS (Universal Mobile Telecommunications System)
- 4G: LTE (Long Term Evolution)
- 5G: in the process of gradually moving on to 5G...

We will sketch the basic security mechanisms in GSM. GSM security is notable since it uses only symmetric-key primitives. UMTS and LTE security improves upon GSM security in several ways, but will not be discussed here.

Here is the basic GSM setup².



The objective of GSM security is to secure the communications between Alice and Bob, and between Dave and Chris, i.e., between a cell phone and its nearest base station. GSM does not provide security

¹Internet Engineering Task Force

²phone drawings are from <https://tex.stackexchange.com/a/479893>

for the communications between base stations – this communications could be secured using a variety of means depending on the service provider, location, country, and so on. And it's even possible that no security is used at all.

4.5.1 GSM Security

Cryptographic ingredients:

- Enc: A symmetric-key encryption scheme.
- MAC: A symmetric-key MAC scheme.
- KDF: A key derivation function.

How do Alice and Bob agree upon a shared secret key? The GSM solution is to use *SIM cards*. A SIM card manufacturer randomly selects a secret key k , and installs it in a SIM card. A copy of k is given to the cell phone service provider. When a user purchases cell phone service, she gets the SIM card which she installs in her phone. Note that A different key k is chosen for each user.

The security objectives of GSM are

1. *Entity authentication*: Cell phone service provider needs to be assured that entities accessing its service are legitimate subscribers.
2. *Confidentiality*: Users need the assurance that their cell phone communications are private.

Here is the basic GSM security protocol. Alice: cell phone user, Bob: cell phone service provider.

Algorithm 15: Basic GSM security protocol

- 1 Alice sends an authentication request to Bob.
 - 2 Bob selects a challenge $r \in_R \{0, 1\}^{128}$, and sends r to Alice.
 - 3 Alice's SIM card uses k to compute the response $t = \text{MAC}_k(r)$. Alice sends t to Bob.
 - 4 Bob retrieves Alice's key k from its database, and verifies that $t = \text{MAC}_k(r)$.
 - 5 Alice and Bob compute an encryption key $K_E = \text{KDF}_k(r)$, and thereafter use the encryption algorithm Enc_{K_E} to encrypt and decrypt messages for each other for the remainder of the session.
-

One drawback with using only symmetric-key crypto is that the SIM card manufacturer and the cell phone service providers have to securely maintain a large database of SIM keys k .

In 2015, the Snowden leaks revealed that NSA and GCHQ had stolen SIM keys from Gemalto, which manufactures about 2 billion SIM cards each year. See <http://tinyurl.com/NSASIM>

Authentic Encryption

A symmetric-key encryption scheme E provides confidentiality, e.g., $E = \text{AES}$, but not authentication. On the other hand, a MAC scheme provides authentication (data origin authentication and data integrity), e.g., $\text{MAC} = \text{HMAC}$, but not confidentiality. What if confidentiality and authentication are *both* required?

First Method: Encrypt-and-MAC

- Alice sends $(c, t) = (E_{k_1}(m), \text{MAC}_{k_2}(m))$ to Bob, where m is the plaintext and k_1, k_2 are secret keys she shares with Bob.
- Bob decrypts c to obtain $m = E_{k_1}^{-1}(c)$ and then verifies that $t = \text{MAC}_{k_2}(m)$.

Intuitively, this provides authenticated encryption because the ciphertext hides the plaintext and the tag authenticates the plaintext. However, this generic method might have some security vulnerabilities: it isn't clear that the tag also hides the plaintext completely. Indeed, it's possible that the tag of a message does leak some bits about the plaintext. This is because MAC schemes were not designed for confidentiality, but rather for authentication.

Second Method: Encrypt-then-MAC

- Alice sends $(c, t) = (E_{k_1}(m), \text{MAC}_{k_2}(E_{k_1}(m)))$ to Bob, where m is the plaintext and k_1, k_2 are secret keys she shares with Bob.
- Bob first verifies that $t = \text{MAC}_{k_2}(c)$ and then decrypts c to obtain $m = E_{k_1}^{-1}(c)$.

This method has been deemed to be secure, provided of course that the encryption scheme E and the MAC scheme employed are secure.

Special-Purpose AE Schemes

Many specialized authenticated encryption schemes have been developed, the most popular of these being **Galois/Counter Mode (GCM)**.

These modes can be faster than generic Encrypt-then-MAC, and also allow for the authentication (but not encryption) of “header” data. This is a useful feature because in many communications protocols the data to be transmitted is preceded by some header data which is specified by the communications protocol. This header data should not be encrypted, but it's usually desirable to protect its authenticity.

5.1 AES-GCM

This is an authenticated encryption scheme proposed by David McGrew and John Viega in 2004. It was adopted as a NIST standard in 2007. AES-GCM uses the CTR mode of encryption and a custom-designed MAC scheme.

5.1.1 CTR: CounTeR Mode of Encryption

Let $k \in_R \{0, 1\}^{128}$ be the secret key.

Let $M = (M_1, M_2, \dots, M_u)$ be a plaintext message, where each M_i is a 128-bit block, $u \leq 2^{32} - 2$.

To encrypt M , Alice does the following:

Algorithm 16: CTR: Encryption

```

1 Select  $IV \in_R \{0, 1\}^{96}$ .
2  $J_0 := IV \parallel 0^{31} \parallel 1$ .
3 for  $i = 1 \dots u$  do
4    $J_i \leftarrow J_{i-1} + 1$  // increment the counter
5   Compute  $C_i = \text{AES}_k(J_i) \oplus M_i$ .
6 Send  $(IV, C_1, C_2, \dots, C_u)$  to Bob.
```

So just like ChaCha20, the counter mode of encryption uses an IV and a counter. This prevents against the accidental reuse of keystream.

To decrypt, Bob does the following:

Algorithm 17: CTR: Decryption

```

1  $J_0 := IV \parallel 0^{31} \parallel 1$ .
2 for  $i = 1 \dots u$  do
3    $J_i \leftarrow J_{i-1} + 1$  // increment the counter
4   Compute  $M_i = \text{AES}_k(J_i) \oplus C_i$ .
```

Note:

1. CTR mode of encryption can be viewed as a stream cipher.
2. As was the case with CBC encryption, identical plaintexts with different IVs result in different ciphertexts.
3. It is critical that the IV should not be repeated; this can be difficult to achieve in practice.
4. Unlike CBC encryption, CTR encryption is parallelizable.
5. Note that AES^{-1} is not used.

Multiplying Blocks

Let $a = a_0a_1a_2 \dots a_{127}$ be a 128-bit block.

We associate the binary polynomial $a(x) = a_0 + a_1x + a_2x^2 + \dots + a_{127}x^{127} \in \mathbb{Z}_2[x]$ with a .

Let $f(x) = 1 + x + x^2 + x^7 + x^{128}$.

If a and b are 128-bit blocks then define $c = a \bullet b$ to be the block corresponding to the polynomial $c(x) = a(x) \cdot b(x) \bmod f(x)$ in $\mathbb{Z}_2[x]$. In other words, that is, $c(x)$ is the remainder upon dividing

$a(x) \cdot b(x)$ by $f(x)$, where coefficient arithmetic is performed modulo 2. This is multiplication in the Galois Field $GF(2^{128})$.

Example: 6 bit-blocks, $GF(2^6)$

Let $f(x) = 1 + x + x^6$. Let $a = 001101$ and $b = 100111$.

Then $a(x) = x^2 + x^3 + x^5$, $b(x) = 1 + x^3 + x^4 + x^5$. Thus, $c(x) = x^2 + x^5$ and $c = a \bullet b = 001001$.

Now let's describe AES-GCM authenticated encryption.

Input:

- Data to be authenticated (but not encrypted) $A = (A_1, A_2, \dots, A_v)$. This might be some header data.
- Data to be encrypted and authenticated: $M = (M_1, M_2, \dots, M_u)$.
- Secret key $k \in_R \{0, 1\}^{128}$.

Output: (IV, A, C, t) , where

- IV is a 96-bit initialization vector.
- $A = (A_1, A_2, \dots, A_v)$ is authenticated data.
- $C = (C_1, C_2, \dots, C_u)$ is the encrypted/authenticated data.
- t is a 128-bit authentication tag.

5.1.2 AES-GCM Encryption, Decryption/Authentication Procedure

Alice does the following:

Algorithm 18: AES-GCM encryption/authentication

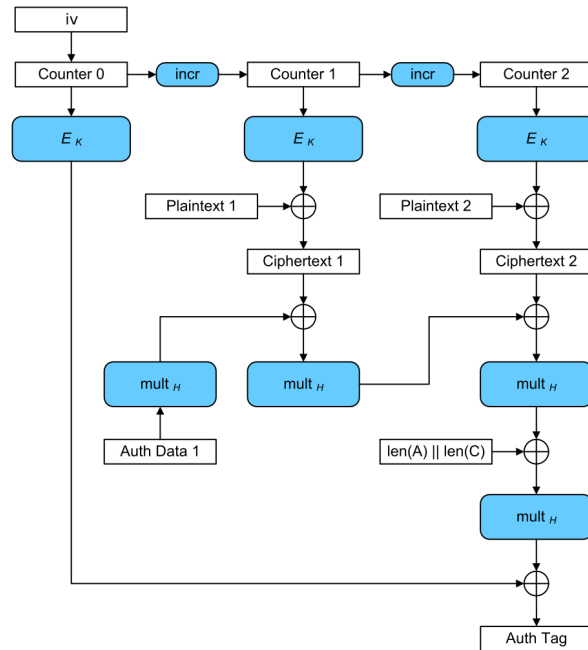
```

1  $L := L_A \parallel L_M$ , where  $L_A, L_M$  are the bitlengths of  $A, M$  expressed as 64-bit integers. ( $L$  is the
   length block.)
2 Select  $IV \in_R \{0, 1\}^{96}$  and let  $J_0 = IV \parallel 0^{31} \parallel 1$ .
   // Encryption
3 for  $i = 1 \dots u$  do
4    $J_i \leftarrow J_{i-1} + 1$ 
5    $C_i \leftarrow \text{AES}_k(J_i) \oplus M_i$ 
   // Authentication
6  $T := 0^{128}$ 
7  $H := \text{AES}_k(0^{128})$ .
8 for  $i = 1 \dots v$  do
9    $T \leftarrow (T \oplus A_i) \bullet H$ .
10 for  $i = 1 \dots u$  do
11    $T \leftarrow (T \oplus C_i) \bullet H$ .
12  $T \leftarrow (T \oplus L) \bullet H$ .
13  $t := \text{AES}_k(J_0) \oplus T$ .
14 Output:  $(IV, A, C, t)$ .
```

Note:

A secret key should be used to encrypt at most 2^{32} messages in order to minimize the possibility of keystream reuse.

Here is a depiction of AES-GCM when the authentication data A is one-block long ($v = 1$) and the plaintext data M is two-blocks long ($u = 2$). Picture from https://commons.wikimedia.org/wiki/File:GCM-Galois_Counter_Mode_with_IV.svg



Upon receiving (IV, A, C, t) , Bob does the following:

Algorithm 19: AES-GCM decryption/authentication

```

1  $L := L_A \parallel L_C$ , where  $L_A, L_C$  are the bitlengths of  $A, C$  expressed as 64-bit integers.
   // Authentication
2  $T := 0^{128}$ 
3  $H := \text{AES}_k(0^{128})$ .
4 for  $i = 1 \dots v$  do
5    $T \leftarrow (T \oplus A_i) \bullet H$ .
6 for  $i = 1 \dots u$  do
7    $T \leftarrow (T \oplus C_i) \bullet H$ .
8  $T \leftarrow (T \oplus L) \bullet H$ .
9  $t' := \text{AES}_k(J_0) \oplus T$ .
10 if  $t' = t$  then
11   Proceed to decryption;
12 else
13   Reject.
   // Decryption
14  $J_0 := IV \parallel 0^{31} \parallel 1$ .
15 for  $i = 1 \dots u$  do
16    $J_i \leftarrow J_{i-1} + 1$ 
17    $M_i \leftarrow \text{AES}_k(J_i) \oplus C_i$ 
18 Accept and output  $(A, M)$ .
```

5.1.3 Insights into Authentication Mechanism

The purpose of GCM authentication is to mix up the bits of the secret key k , the bits of the message A , the bits of the ciphertext C , and the bits of the length block L , and the IV which is contained in the counter block J_0 , to produce the authentication tag t . This is done by these **four lines of codes**.

- $T \leftarrow 0^{128}, H \leftarrow \text{AES}_k(0^{128})$
- **For i from 1 to v do:** $T \leftarrow (T + A_i)H$

$$\begin{aligned} T &= (((0 + A_1)H + A_2)H + A_3)H + \dots + A_v)H \\ &= A_1H^v + A_2H^{v-1} + A_3H^{v-2} + \dots + A_{v-1}H^2 + A_vH \end{aligned}$$

- **For i from 1 to u do:** $T \leftarrow (T + C_i)H$.
- $T \leftarrow (T + L)H$

$$T = A_1H^{u+v+1} + A_2H^{u+v} + \dots + A_vH^{u+2} + C_1H^{u+1} + \dots + C_uH^2 + LH = f_{A,M}(H),$$

where

$$f_{A,M}(x) = A_1x^{u+v+1} + \dots + A_vx^{u+2} + C_1x^{u+1} + \dots + C_u x^2 + Lx \in GF(2^{128})[x]$$

- Hence, $t = \text{AES}_k(J_0) \oplus f_{A,M}(H)$.

To justify the security of the AES-GCM authentication mechanism, we'll consider the case where AES-GCM is used for authentication only.

Consider AES-GCM with no M , so $u = 0, L_M = 0$. The message to be authenticated is $A = (A_1, A_2, \dots, A_v)$, where $v \leq \ell$. The tag is (IV, t) , where $IV \in_R \{0, 1\}^{96}$ and $t = \text{AES}_k(J_0) + f_A(H)$. Again, $J_0 = IV \parallel 0^{31} \parallel 1, H = \text{AES}_k(0)$.

Attack goal: this is the one in the security definition for MAC schemes. Eve has message-tag pairs (for messages of her choosing): (A^j, IV^j, t^j) , $1 \leq j \leq r$. Her goal is to produce a message-tag forgery (A^*, IV^*, t^*) , where $A^* \notin \{A^1, A^2, \dots, A^r\}$.

We can assume that no two IV's in this list are the same. This is because the IV is produced by the MACing oracle are 96-bit strings and were randomly generated. We can also assume that Eve does not know k or H .

Security Argument (Informal)

Now, suppose that Eve outputs a forgery (A^*, IV^*, t^*) . The security argument is divided into two cases.

1. If $IV^* \notin \{IV^1, \dots, IV^r\}$, then $J_0^* \notin \{J_0^1, \dots, J_0^r\}$, and so Eve doesn't know $\text{AES}_k(J_0^*)$, which serve as a one-time pad for $f_{A^*}(H)$. Thus, the probability that Eve can output a valid tag (i.e., $t^* = \text{AES}_k(J_0^*) + f_{A^*}(H)$) is only $1/2^{128}$.
2. Suppose that $IV^* = IV^j$, for some $1 \leq j \leq r$, so $J_0^* = J_0^j$. Then

$$t^* - t^j = \text{AES}_k(J_0^*) + f_{A^*}(H) - \text{AES}_k(J_0^j) - f_{A^j}(H) = f_{A^*}(H) - f_{A^j}(H)$$

So, Eve has produced A^*, A^j and α such that $\alpha = f_{A^*}(H) - f_{A^j}(H)$, without knowledge of H . But this can only be done with negligible probability, as the following lemma shows.

Lemma

For all distinct $A, B \in (\{0, 1\}^{128})^{\leq \ell}$ and $\alpha \in \{0, 1\}^{128}$,

$$\Pr[f_A(H) - f_B(H) = \alpha] \leq (\ell + 1)/2^{128}$$

(which is negligible), where the probability is assessed over random choices of $H \in \{0, 1\}^{128}$.

Proof:

Let $A, B \in (\{0, 1\}^{128})^{\leq \ell}$ with $A \neq B$, and let $\alpha \in \{0, 1\}^{128}$. Suppose $A \in \{0, 1\}^{128r}$ and $B \in \{0, 1\}^{128w}$. Then

$$\begin{aligned} f_A(H) - f_B(H) - \alpha &= (A_1 H^{v+1} + A_2 H^v + \cdots + A_v H^2 + L_A H) \\ &\quad - (B_1 H^{w+1} + B_2 H^w + \cdots + B_w H^2 + L_B H) - \alpha, \end{aligned}$$

which is a polynomial in H of degree $\leq \max(v, w) + 1 \leq \ell + 1$.

Since a nonzero polynomial of degree $\leq \ell + 1$ can have at most $\ell + 1$ roots, there are at most $\ell + 1$ $H \in \{0, 1\}^{128}$ satisfying $f_A(H) - f_B(H) - \alpha = 0$. \square

5.1.4 Some Features of AES-GCM

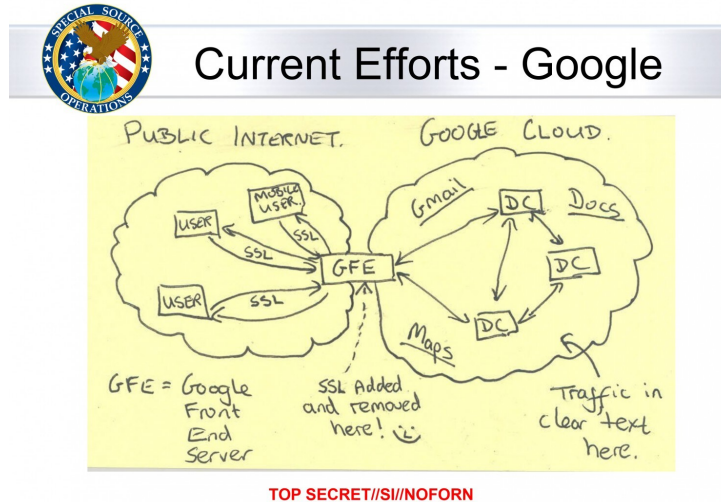
1. Performs authentication and encryption.
2. Supports authentication only (by using empty M).
3. Very fast implementations on Intel and AMD processors because of special AES-NI and PCLMULQDQ instructions for the AES and \bullet operations.
4. Encryption and decryption can be parallelized.
5. AES-GCM can be used in streaming mode¹.
6. Security is justified by a security proof: Original McGrew-Viega security proof (2004) was wrong. The proof was fixed in 2012 by Iwata-Ohashi-Minematsu.

AES-GCM is widely used today.

¹Streaming mode means that both the encryption and the authentication operations can be performed on the blocks of m as they are received, one at a time. This is useful in streaming applications.

5.2 Google Encryption

TOP SECRET//SI//NOFORN



TOP SECRET//SI//NOFORN

This is from NSA classified documents that were leaked by Snowden in 2013. On the left we see a bunch of users on the internet accessing Google's services such as gmail and search, and they do this by communicating with a Google front-end server GFE. The communications between the computers and Google's front-end servers are protected using TLS, also known as SSL. The data transmitted by the users are decrypted by the Google front-end servers, and then transmitted through Google's private network to its data centers situated around the world. Communications between the front-end servers and data centers and between data centers is through these private links. the slide had two comments, one noting that SSL is added and removed here, and the other is that traffic in the clear text is transmitted in the cloud.

GCHQ/NSA MUSCULAR program

Disclosed by Edward Snowden on October 30, 2013. Surveillance program conducted by GCHQ (British spy agency) in partnership with NSA. An unnamed telecommunications operator provided GCHQ with secret access to its fibre optic cables that transported data between Google and Yahoo! data centres. Millions of records collected each day, so that they couldn't possibly store all of it, so they had to process and discard a lot of it on the fly. Google was not encrypting user data when it transported it between data centers. In November 2013, Google and Yahoo! announced they were encrypting all traffic between their data centres.

5.2.1 Google Data

Google has 21 data centres around the world. A data centre (DC) contains tens of thousands of servers. It has lots of physical security (cameras, biometric identification, metal detectors, vehicle barriers, etc.) Communication between these servers and the outside world is all done via Google Front End (GFE) servers. Servers within a data centre communicate via a LAN (Local Area Network). Servers in different data centres communicate via a WLAN (Wide Local Area Network).

Broadly speaking, there are three kinds of data to protect:

1. Data communicated between individual users (browsers) and Google (GFEs).
 - Data is encrypted and authenticated using TLS.
 - TLS uses symmetric-key enc. (e.g., AES), symmetric-key authentication (e.g., HMAC); au-

thenticated enc. (e.g., AES-GCM); key establishment (e.g., RSA public-key enc., ECDH); public key certificates (RSA signatures).

- TLS is used by all web servers and browsers (not Google specific).
2. Data communicated between Google servers (perhaps in different data centres).

Data is secured using Google's version of TLS (Application Layer Transport Security, ALTS). ALTS handles roughly 10^{10} remote procedure calls (RPCs) per second.

3. Data stored at data centres.

In the remainder of this section, we'll just discuss Google's methods for storing data at its data centers.

5.2.2 Key Management Service (KMS)

All data stored within Google data centres is encrypted with AES256-GCM (GCM = Galois Counter Mode). AES128-CTR + HMAC-SHA256 (Encrypt-then-MAC) is used in some legacy applications. However, Google is in the process of moving the encryption for these legacy applications to AES-256 GCM.

What secret keys does Google use to encrypt all this data? At one extreme, Google could use one secret key to encrypt all the data in its backend. However, this is clearly risky when the adversary gets his hands on this secret key. Google's solution is at the other extreme: a unique session key to encrypt each piece of data. This raised a lot of practical questions, which are addressed by Google's KMS.

The KMS manages the many AES secret keys that are used to encrypt/decrypt data by the many storage services within data centres. So the KMS has some very stringent requirements. Some of the KMS requirements:

- Availability: $> 99.9995\%$ requests are served.
- Latency: 99% of requests are served in < 10 ms.
- Scalability: Handle $\approx 10^7$ requests/second.
- Security: Effortless and foolproof key rotation (3 months).
- Efficiency: To minimize number of machines needed.

On January 24, 2014, a KMS configuration file was truncated by error. As a result, the KMS did not know the secret keys used to decrypt stored data. Gmail, Calendar, Docs, etc. crashed for 25+ minutes. "This got a lot of attention within Google." Subsequently, Google made many changes to its KMS. $\gg 99.9999\%$ of KMS requests are served. 99.9% of requests are served in $< 200\mu\text{s}$.

5.2.3 Google's Key Hierarchy

1. Storage systems (millions of processes)
 - Encrypts data with DEKs (Data Encryption Keys).
2. KMS (tens of thousands)
 - Encrypts DEKs with KEKs (Key Encryption Keys).
3. Root KMS (hundreds)
 - Encrypts KEKs with KMS Master Keys.
4. Root KMS Master Key Distributor (hundreds)

Encrypts KMS Master Keys with the Root KMS Master Key.

5. Physical safes (two)

The Root KMS Master Key is backed up on hardware devices.

1. Storage Systems

Suppose that a storage system wishes to encrypt some data item m . The storage system does the following.

1. Break up m into chunks, m_1, \dots, m_ℓ .

Each chunk can be up to several Gigabytes in size.

2. Generate ℓ Data Encryption Keys (DEKs), k_1, \dots, k_ℓ .

Multiple sources of entropy are sampled (e.g., Intel's RDRAND instruction; inter-packet arrival times; measurement of disk seeks) The samples are then combined and hashed using a key derivation function (KDF) to produce a 256-bit secret key.

3. Encrypt with AES256-GCM: $c_1 = \text{AES}_{k_1}(m_1), \dots, c_\ell = \text{AES}_{k_\ell}(m_\ell)$,

4. Send the DEKs k_1, \dots, k_ℓ to a KMS.

This transmission is protected with ALTS, Google's internal version of TLS.

5. Receive the wrapped (encrypted) keys w_1, \dots, w_ℓ from the KMS.

The KMS encrypts the DEKs with its Key Encryption Keys (KEKs).

6. Store $(c_1, w_1), \dots, (c_\ell, w_\ell)$.

These encrypted chunks are replicated and distributed across Google's storage systems.

When an application needs a piece of data, the stored system needs to decrypt the corresponding chunk, (c_j, w_j) :

1. The storage system sends the wrapped key w_j to the KMS.
2. The KMS decrypts w_j using the appropriate KEK, and sends the DEK k_j to the storage system.
3. The storage system decrypts c_j using k_j .

Note:

1. Each data chunk has a unique identifier.
2. The KMS maintains an Access Control List (ACL) to ensure that a data chunk can only be decrypted by the authorized storage system.
3. Note that each chunk of data is encrypted using a different DEK. This ensures that if a DEK is compromised, then only one chunk of data is potentially compromised.
4. The storage system does not store the DEKs k_1, \dots, k_ℓ but rather the wrapped data encryption keys.
5. If a chunk of data m_j is updated, it is re-encrypted with a new DEK k'_j rather than using the old DEK k_j . So each data encryption key is only used once.

2. KMS

Key Management Services (KMSs) generate the AES256-GCM Key Encryption Keys (KEKs), and maintain the Access Control Lists (one ACL list for each KEK). The KMS encrypts/decrypts DEKs using the KEKs, in accordance with the ACL. The KEKs never leave the KMS. The KMS also maintains an audit trail of when a KEK was used. KEKs are rotated (i.e., changed). The standard rotation frequency is once every 90 days.

3, 4. Root KMS

The Root KMS wraps KEKs with AES256-GCM KMS Master Keys. There are about a dozen KMS Master Keys. The Root KMSs are run on dedicated secured machines in Google's data centres.

The KMS Master Keys are wrapped with the AES256-GCM Root KMS Master Key. The Root KMS Master Key is stored in RAM on the Root KMS machines. The Root KMS Master Key Distributor ensures that all Root KMSs always have the same version of the Root KMS Master Key.

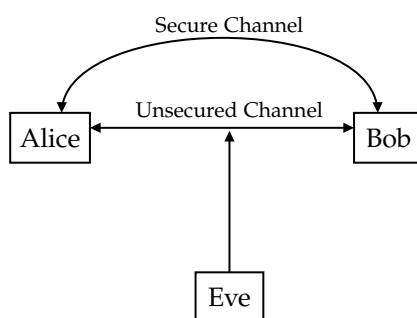
5. Physical Safes

The Root KMS Master Key is backed up on two hardware devices stored in physical safes in highly secured areas in two physically separated Google locations. Fewer than 20 Google employees have access to these safes. The backups will be used if Google ever has to do a complete reboot.

Introduction to Public-Key Cryptography

6.1 Drawbacks with Symmetric-Key Cryptography

Symmetric-key cryptography: Communicating parties a priori share some secret keying information.

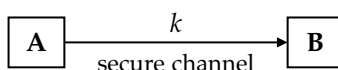


The shared secret keys can then be used to achieve confidentiality (e.g., using AES), or authentication (e.g., using HMAC), or both (e.g., using AES-GCM).

6.1.1 Key Establishment Problem

How do Alice and Bob establish the secret key k ?

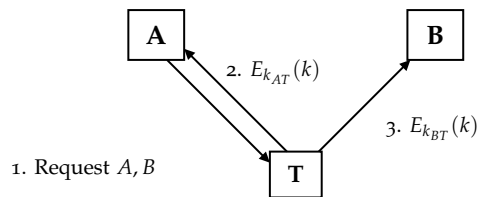
Method 1 *Point-to-point key distribution.* (Alice selects the key, sends it to Bob over a secure channel)



The secure channel could be: A trusted courier, a face-to-face meeting; installation of an authentication key in a SIM card. This is generally not practical for large-scale applications.

Method 2 *Use a Trusted Third Party (TTP) T .*

Each user A shares a secret key k_{AT} with T for a symmetric-key encryption scheme E . To establish this key, A must visit T once. T serves as a **key distribution centre** (KDC):



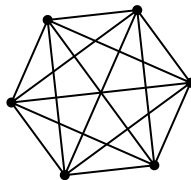
1. A sends T a request for a key to share with B .
2. T selects a session key k , and encrypts it for A using k_{AT} .
3. T encrypts k for B using k_{BT} .

Drawbacks of using a KDC:

1. The TTP must be unconditionally trusted.
2. The TTP is an attractive target.
3. The TTP must be on-line. Potential bottleneck. Critical reliability point.

6.1.2 Key Management Problem

In a network of n users, each user has to share a different key with every other user.



Each user thus has to store $n - 1$ different secret keys. The total number of secret keys is $\binom{n}{2} \approx n^2/2$.

6.1.3 Non-Repudiation is Difficult to Achieve

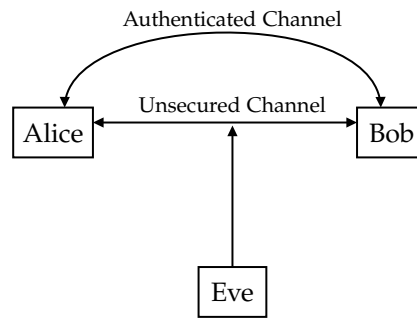
Recall non-repudiation is preventing an entity from denying previous actions or commitments, denying being the source of a message.

Strictly speaking, symmetric-key techniques cannot be used to achieve non-repudiation. This is because Alice and Bob share a secret key, so Bob can do with the secret key whatever Alice can do with that secret key. If Alice uses the secret key to send Bob a message, perhaps say using HMAC, Bob knows that the message came from Alice but Bob could not prove to a judge that the message came from Alice because Bob could have generated the message and its authentication tag himself.

6.2 Public-Key Cryptography

public-key cryptography

Communicating parties a priori share some *authenticated* (but non-secret) information.



Invented by Ralph Merkle, Whitfield Diffie, Martin Hellman in 1975.

Excerpts from Merkle's CS 244 project proposal (Computer Security, UC Berkeley, Fall 1974)

"Secure communications are made possible because of knowledge, known to both people, which is not known to anyone else. Usually, both people know this knowledge because they were able to hold a private conversation with each other before they began to send encrypted messages over an unsecure channel."

"It might seem intuitively obvious that if two people have never had opportunity to prearrange an encryption method, then they will be unable to communicate securely over an insecure channel. While this might seem intuitively obvious, I believe it is false. I believe that it is possible for two people to communicate securely without having made any prior arrangements that are not completely public."

6.2.1 Merkle Puzzles

This method is meant to be proof-of-concept. It's not meant to be practical. The goal for Alice and Bob is to establish a secret session key by communicating over an authenticated (but non-secret) channel.

1. Alice creates N puzzles P_i , $1 \leq i \leq N$ (e.g., $N = 10^9$). Each puzzle takes t hours to solve (e.g., $t = 5$). The solution to P_i reveals a 128-bit session key sk_i and a randomly-selected 128-bit serial number n_i (which Alice selected and stored).
2. Alice sends P_1, P_2, \dots, P_N to Bob.
3. Bob selects j at random from $[1, N]$ and solves puzzle P_j to obtain sk_j and n_j .
4. Bob sends n_j to Alice.
5. The secret session key is sk_j .

An eavesdropper has to solve 500,000,000 puzzles on average to determine the puzzle index j (and thus sk_j).

Example:

$P_i = \text{AES-CBC}_{k_i}(sk_i, n_i, n_i)$, where $k_i = (r_i \parallel 0^{88})$ and r_i is a randomly selected 40-bit string. P_i can be solved in 2^{40} steps by exhaustive key search.

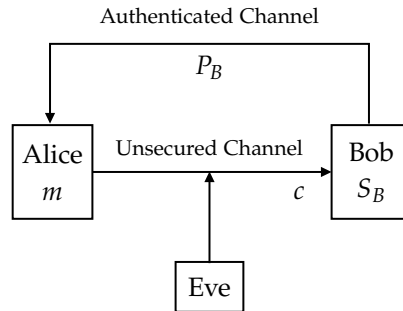
Key Pair Generation for Public-Key Cryptography

Each entity A does the following:

1. Generate a key pair (P_A, S_A) .
2. S_A is A 's secret key.
3. P_A is A 's public key.

Security requirement: It should be infeasible for an adversary to recover S_A from P_A . For example, $S_A = (p, q)$ where p and q are randomly-selected prime numbers; $P_A = p \cdot q$.

6.2.2 Public-Key Encryption



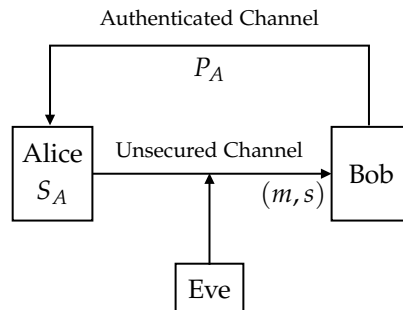
To encrypt a secret message m for Bob, Alice does:

1. Obtain an authenticated copy of Bob's public key P_B .
2. Compute $c = E(P_B, m)$; E is the encryption function.
3. Send c to Bob.

To decrypt c , Bob does:

1. Compute $m = D(S_B, c)$; D is the decryption function.

6.2.3 Digital Signatures



To sign a message m , Alice does:

1. Compute $s = \text{Sign}(S_A, m)$.
2. Send m and s to Bob.

To verify Alice's signature s on m , Bob does:

1. Obtain an authenticated copy of Alice's public key P_A .
2. Accept if $\text{Verify}(P_A, m, s) = \text{"Accept"}$.

Suppose that Alice generates a signed message (m, s) . Then *anyone* who has an authentic copy of Alice's public key P_A can verify the authenticity of the signed message. This authentication property cannot be achieved with a symmetric-key MAC scheme. Digital signatures are widely used to sign software updates which are then broadcast to computers around the world.

6.2.4 Hybrid Schemes

Advantages of public-key cryptography:

- No requirement for a secured channel.
- Each user has only 1 key pair, which simplifies key management.
- A signed message can be verified by anyone.
- Facilitates the provision of non-repudiation services (with digital signatures).

Disadvantages of public-key cryptography:

- Public-key schemes are slower than their symmetric-key counterparts.

Therefore, in practice, symmetric-key and public-key schemes are used together. Here is an example:

To encrypt a secret signed message m , Alice does:

Algorithm 20: Hybrid scheme: encryption of signed message

- 1 $s \leftarrow \text{Sign}(S_A, m)$
 - 2 Select a secret key k for a symmetric-key encryption scheme such as AES.
 - 3 Obtain an authentic copy of Bob's public key P_B .
 - 4 Send $c_1 = E(P_B, k)$ and $c_2 = \text{AES}_k(m, s)$.
-

To recover m and verify its authenticity, Bob does:

Algorithm 21: Hybrid scheme: authenticity verification

- 1 Decrypt $c_1 : k = D(S_B, c_1)$
 - 2 Decrypt c_2 using k to obtain (m, s) .
 - 3 Obtain an authentic copy of Alice's public key P_A .
 - 4 Check that $\text{Verify}(P_A, m, s) = \text{"Accept"}$.
-

6.3 Algorithmic Number Theory

Fundamental Theorem of Arithmetic

Every integer $n \geq 2$ has a unique prime factorization (up to ordering of factors).

Note that this is an example of unique factorization domain. Check [PMATH 347](#) for more details.

We then have some interesting questions:

- Given an integer $n \geq 2$, how do we find its prime factorization efficiently?
- How do we efficiently verify an alleged prime factorization of an integer $n \geq 2$?
- Given an integer $n \geq 2$, how do we efficiently decide whether n is prime or composite?

6.3.1 Complexity Theory Review

It's a review because the reader is assumed to be familiar with [CS 341](#) and [CS 466](#).

algorithm

An **algorithm** is a “well-defined computational procedure” (e.g., a Turing machine) that takes a variable input and eventually halts with some output.

For an integer factorization algorithm, the input is a positive integer n , and the output is the prime factorization of n .

efficiency

The **efficiency** of an algorithm is measured by the scarce resources it consumes (e.g. time, space, number of processors, number of chosen plaintext-ciphertext pairs).

input size

The **input size** is the number of bits required to write down the input using a reasonable encoding.

Example:

The size of a positive integer n is $\lfloor \log_2 n \rfloor + 1$ bits.

running time

The **running time** of an algorithm is an upper bound as a function of the input size, of the worst case number of basic steps the algorithm takes over all inputs of a fixed size.

polynomial-time

An algorithm is a **polynomial-time** (efficient) algorithm if its (expected) running time is $O(k^c)$, where c is a fixed positive integer, and k is the input size.

6.3.2 Basic Integer Operations

Input: Two k -bit positive integers a and b

Input size: $O(k)$ bits.

Operation	Running time of naive alg (in bit ops)
Addition: $a + b$	$O(k)$
Subtraction: $a - b$	$O(k)$
Multiplication: $a \cdot b$	$O(k^2)$
Division: $a = qb + r$	$O(k^2)$
GCD: $\gcd(a, b)$	$O(k^2)$

In 2019, it was proven that k -bit numbers can be multiplied in $O(k \log k)$ bit operations. This algorithm is really complicated and highly inefficient in practice because of the very large constants that are suppressed by the big O notation. The big open problem is whether k -bit numbers can be multiplied in $O(k)$ bit operations.

GCD: Can be efficiently computed using the Euclidean Algorithm.

6.3.3 Basic Modular Operations

Input: A k -bit integer n , and integers $a, b, m \in [0, n - 1]$.

Input size: $O(k)$ bits.

Operation	Running time of naive alg (in bit ops)
Addition: $a + b \bmod n$	$O(k)$
Subtraction: $a - b \bmod n$	$O(k)$
Multiplication: $a \cdot b \bmod n$	$O(k^2)$
Inversion: $a^{-1} \bmod n$	$O(k^2)$
Exponentiation: $a^m \bmod n$	$O(k^3)$

Note:

$x = a^{-1} \bmod n$ means $ax \equiv 1 \bmod n$ and $1 \leq x \leq n - 1$. $a^{-1} \bmod n$ exists iff $\gcd(a, n) = 1$.

$a^{-1} \bmod n$ can be efficiently computed using the Extended Euclidean Algorithm: Find $s, t \in \mathbb{Z}$ with $as + nt = 1$; then $x = s \bmod n$.

Modular Exponentiation

Input: A k -bit integer n , and integers $a, m \in [0, n - 1]$.

Output: $a^m \bmod n$.

Algorithm 22: Modular exponentiation naive algorithm 1

```

1 Compute  $d = a^m$ 
2 return  $d \bmod n$ 
```

Note the bitlength of d is (approximately) $\log_2 d = \log_2 a^m = m \cdot \log_2 a = O(2^k k)$ since $m \approx 2^k$. Hence the algorithm is not polytime.

Algorithm 23: Modular exponentiation naive algorithm 2

```

1  $A \leftarrow a$ 
2 for  $i \leftarrow 2 \dots m$  do  $A \leftarrow A \times a \bmod n$ ;
3 return  $A$ 
```

Let the binary representation of m be $m = \sum_{i=0}^{k-1} m_i 2^i$ where $m_i \in \{0, 1\}$. Then

$$a^m \equiv a^{\sum_{i=0}^{k-1} m_i 2^i} \equiv \prod_{i=0}^{k-1} a^{m_i 2^i} \equiv \prod_{\substack{0 \leq i \leq k-1 \\ m_i = 1}} a^{2^i} \pmod{n}$$

This suggests the following repeated square-and-multiply algorithm for computing $a^m \bmod n$:

Algorithm 24: Modular exponentiation: repeated square-and-multiply

```

1 Write  $m$  in binary:  $m = \sum_{i=0}^{k-1} m_i 2^i$ .
2 if  $m_0 = 1$  then  $B \leftarrow a$ ;
3 else  $B \leftarrow 1$ ;
4  $A \leftarrow a$ 
5 for  $i \leftarrow 1 \dots k - 1$  do
6    $A \leftarrow A^2 \bmod n$ 
7   if  $m_i = 1$  then  $B \leftarrow B \times A \bmod n$ ;
8 return  $B$ 
```

Analysis: At most k modular squarings and k modular multiplications, so worst-case running time is $O(k^3)$ bit operations (polytime).

RSA

7.1 Basic RSA

Ron Rivest, Adi Shamir, Len Adleman

7.1.1 RSA Encryption

Each entity Alice does the following:

Algorithm 25: RSA Key Generation

- 1 Randomly select two large, distinct primes p, q of the same bitlength.
// This can be done efficiently. In practice, p and q should have bitlength at least 1024.
 - 2 Compute $n = pq$ and $\phi = \phi(n) = (p - 1)(q - 1)$.
 - 3 Select arbitrary large e , $1 < e < \phi$, with $\gcd(e, \phi) = 1$.
 - 4 Compute the integer d , $1 < d < \phi$, with $ed \equiv 1 \pmod{\phi}$.
 - 5 Alice's public key is (n, e) ; her private key is d .
-

Note that n is called the *RSA modulus*; e is the *encryption exponent*; d is the *decryption exponent*.

To encrypt a message for Alice, Bob does:

Algorithm 26: RSA Encryption

- 1 Obtain an authentic copy of Alice's public key (n, e) .
 - 2 Represent the message as an integer $m \in [0, n - 1]$.
 - 3 Compute the ciphertext $c = m^e \pmod{n}$.
 - 4 Send c to Alice.
-

To decrypt c , Alice does

Algorithm 27: RSA Decryption

- 1 Compute $m = c^d \pmod{n}$.
-

Example: Toy example on RSA Key Generation/Encryption

Alice does the following for key generation:

1. Select primes $p = 23, q = 37$.

2. Compute $n = pq = 851$ and $\phi(n) = (p-1)(q-1) = 792$.
3. Select $e = 631$ satisfying $\gcd(631, 792) = 1$.
4. Solves $631d \equiv 1 \pmod{792}$ to get $d \equiv -305 \equiv 487 \pmod{792}$, and selects $d = 487$.
5. Alice's public key is $(n = 851, e = 631)$; her private key is $d = 487$.

To encrypt a plaintext $m = 13$ for Alice, Bob does:

1. Obtains Alice's public key $(n = 851, e = 631)$.
2. Computes $c = 13^{631} \pmod{851}$ using the repeated-square-and-multiply algorithm:

(a) Write $e = 631$ in binary: $e = 2^9 + 2^6 + 2^5 + 2^4 + 2^2 + 2^1 + 2^0$.

(b) Compute successive squarings of $m = 13 \pmod{n}$:

$$\begin{array}{ll}
 13 \equiv 13 \pmod{851} & 13^2 \equiv 169 \pmod{851} \\
 13^2 \equiv 169 \pmod{851} & 13^{2^3} \equiv 416 \pmod{851} \\
 13^{2^4} \equiv 303 \pmod{851} & 13^{2^5} \equiv 752 \pmod{851} \\
 13^{2^6} \equiv 440 \pmod{851} & 13^{2^7} \equiv 423 \pmod{851} \\
 13^{2^8} \equiv 219 \pmod{851} & 13^{2^9} \equiv 305 \pmod{851}.
 \end{array}$$

(c) Multiply together the squares 13^{2^i} for which the i th bit (where $0 \leq i < 9$) of the binary representation of 631 is 1:

$$\begin{aligned}
 13^{631} &= 13^{2^9+2^6+2^5+2^4+2^2+2^1+2^0} \\
 &= 13^{2^9} \cdot 13^{2^6} \cdot 13^{2^5} \cdot 13^{2^4} \cdot 13^{2^2} \cdot 13^{2^1} \cdot 13^{2^0} \\
 &\equiv 305 \cdot 440 \cdot 752 \cdot 303 \cdot 478 \cdot 169 \cdot 13 \pmod{851} \\
 &\equiv 616 \pmod{851}.
 \end{aligned}$$

3. Bob sends the ciphertext $c = 616$ to Alice.

To decrypt $c = 616$, Alice uses her private key $d = 487$ as follows:

1. Compute $m = 616^{487} \pmod{851}$ to get $m = 13$.

Theorem (RSA works)

For all integers m , $m^{ed} \equiv m \pmod{n}$.

Proof:

Since $ed \equiv 1 \pmod{\phi}$, we can write $ed = 1 + k(p-1)(q-1)$ for some $k \in \mathbb{Z}$. Since $ed > 1$ and $p, q \geq 2$, we have $k \geq 1$. We will prove that $m^{ed} \equiv m \pmod{p}$.

Suppose first that $p \mid m$. Then $m \equiv 0 \pmod{p}$, so $m^{ed} \equiv 0 \pmod{p}$. Hence $m^{ed} \equiv m \pmod{p}$.

Suppose now that $p \nmid m$. Then $\gcd(p, m) = 1$, so $m^{p-1} \equiv 1 \pmod{p}$ by Fermat's Little Theorem. Raising both sides to the power $k(q-1)$ and then multiply by m gives $m^{k(p-1)(q-1)+1} \equiv m \pmod{p}$. Hence $m^{ed} \equiv m \pmod{p}$.

Similarly, $m^{ed} \equiv m \pmod{q}$. As p, q are distinct primes, $\gcd(p, q) = 1$ and so $m^{ed} \equiv m \pmod{pq}$. \square

7.1.2 Basic RSA Signature Scheme

Key Generation Same as for RSA encryption: Algorithm 25.

To sign $m \in \{0,1\}^*$, Alice does:

Algorithm 28: RSA Signature Generation



- 1 Compute $M = H(m)$, where H is a hash function.
 - 2 Compute the signature $s = M^d \bmod n$. // so $s^e \equiv M^{ed} \equiv M \bmod n$
 - 3 Alice's signed message is (m, s) . // the signed message is (m, s)
-

To verify (m, s) , Bob does:

Algorithm 29: RSA Signature Verification

- 1 Obtain an authentic copy of Alice's Public key (n, e) .
 - 2 Compute $M = H(m)$
 - 3 Compute $M' = s^e \bmod n$.
 - 4 Accept (m, s) iff $M = M'$.
-

7.2 Case Study: QQ Browser

The Snowden documents suggested that the NSA (and collaborators) were exploiting vulnerabilities in the UC Browser  (a browser for mobile devices that is popular in China) to track users. This prompted Knockel, Senft and Deibert (UToronto) in 2016 to study the security of browsers used in China. They studied QQ Browser , a free web browser for Android, Windows, Mac, iOS, developed by Tencent. QQ Browser is used by hundreds of millions of cell phone users in China.

7.2.1 Version 1

When a user launches the QQ Browser (the client) on Android, the browser makes a series of WUP requests to QQ Browser's server (the server).

Via a WUP request, the browser sends personal user data to the server. This data includes: QQ username, WiFi MAC address of client, MAC addresses of all nearby WiFi access points, URL of each page visited by the browser... (UC Browser and Baidu Browser also collect similar data.) Of course, the personal user data needs to be protected as it is transmitted over the internet.

QQ Browser WUP Encryption

1. To encrypt a WUP request m , the client does the following:

- (a) Randomly generate a 128-bit AES session key k :

```
int i = 10000000 + new Random().nextInt(89999999);
int j = 10000000 + new Random().nextInt(89999999);
return (String.valueOf(i) + String.valueOf(j)).getBytes();
```

Then $k = i \parallel j$ where i, j are 8-byte ASCII strings. So the QQ browser aims to provide the 128-bit security level.

- (b) Encrypt k with the server's RSA public key (n, e) : $c_1 = k^e \bmod n$.

Here $e = 65537$ and $n = 245406417573740884710047745869965023463$ (128 bits).

- (c) Encrypt m : $c_2 = \text{AES-ECB}_k(m)$.

- (d) Send (c_1, c_2) to the server.
2. The server does the following:
 - (a) Decrypt c_1 using its RSA private key d : $k = c_1^d \bmod n$.
 - (b) Decrypt c_2 : $m = \text{AES-ECB}_k^{-1}(c_2)$.
 - (c) Encrypt the WUP response m' : $c' = \text{TEA-CBC}_{k'}(m')$,¹ where $k' = \text{sDf434o1*123+-KD}$ (in ASCII). Send c' .
3. The client decrypts $m' = \text{TEA-CBC}_{k'}^{-1}(c')$ using the hard-coded 128-bit k' .

WUPS!

1. i and j are each randomly selected integers in the interval $[10000000, 99999998]$, so the total number of keys k is $\approx 2^{26.4} \times 2^{26.4} \approx 2^{52.8}$ (not 2^{128}).
2. The server's RSA public key is easily factored: $n = 14119218591450688427 \times 17381019776996486069$.
3. The response uses the fixed key k' that is hard-coded in all QQ browsers!
4. Don't use ECB mode! One should always use CBC mode instead.

Summary: The user data is protected using extremely weak cryptography, and thus is very vulnerable to passive eavesdropping.

7.2.2 Version 2

1. To encrypt a WUP request m , the client does the following:
 - (a) Randomly generate a 128-bit AES session key k .
 - (b) Encrypt k with the server's RSA public key (n, e) : $c_1 = k^e \bmod n$. Here, $e = 65537$ and n is a 1024-bit RSA modulus.
 - (c) Encrypt m : $c_2 = \text{AES-ECB}_k(m)$.
 - (d) Send (c_1, c_2) to the server.
2. The server does the following:
 - (a) Decrypt c_1 using its RSA private key d : $K = c_1^d \bmod n$, and let k be the 128 least significant bits of K .
 - (b) Decrypt c_2 : $m = \text{AES-ECB}_k^{-1}(c_2)$.
 - (c) If m is a properly formatted WUP request, then encrypt the WUP response m' : $c' = \text{AES-ECB}_k(m')$ and send c' . If m is not properly formatted, then don't respond.
3. The client decrypts $m' = \text{AES-ECB}_k^{-1}(c')$.

In step 2(a), the integer K is represented as a 1024-bit number, of which the least significant 128-bits are taken to be k . However, the QQ server software did not check that the remaining $1024 - 128 = 896$ bits of K are all 0 (as they should be). This flaw can be exploited using a *restricted chosen-ciphertext attack*:

- (a) The adversary intercepts a ciphertext $c = (c_1, c_2)$.
- (b) She then sends to the QQ server a carefully-chosen modification $\hat{c} = (\hat{c}_1, \hat{c}_2)$ of c .

¹tiny encryption algorithm

- (c) Depending on whether the server responds or not, the adversary learns 1 bit of the secret key k .
- (d) Steps (b) and (c) are repeated to obtain all the 128 bits of k .

The attack requires 128 interactions with the QQ server and very little computation. (Note that the RSA private key d is not computed.)

Check <https://citizenlab.ca/2016/03/privacy-security-issues-qq-browser/> for more details.

7.3 Security of RSA Encryption

Security of RSA Key Generation If an adversary can factor n , then she can compute d from (n, e) . It has been proven that any efficient method for computing d from (n, e) is equivalent to factoring n .

Security of Basic RSA Encryption A basic notion of security is that it should be computationally infeasible to compute m from c . This is known as the RSA problem:

RSA Problem (RSAP)

Given an RSA public key (n, e) and $c = m^e \bmod n$, where $m \in [0, n - 1]$, compute m .

The only effective method known for solving RSAP is to factor n (and then compute d and m). Henceforth, we shall assume that RSAP is *intractable*.

Dictionary Attack Suppose that the plaintext is chosen from a relatively small (and known) set M of messages. Then, given a target ciphertext c , the adversary can encrypt each message in M until c is obtained.

Countermeasure: Append a randomly selected 128-bit string (called a *salt*) to m prior to encryption. Note that m is now encrypted to one of 2^{128} possible ciphertexts.

Chosen-Ciphertext Attack Suppose the adversary E has a target ciphertext c intended for Alice. Suppose also that E can induce Alice to decrypt *any* ciphertext for her, *except for c itself* (we say that Eve has a *decryption oracle*). Then E can decrypt c as follows:

1. Select arbitrary $x \in [2, n - 1]$ with $\gcd(x, n) = 1$.
2. Compute $\hat{c} = c \cdot x^e \bmod n$, where (n, e) is Alice's public key.

Note that $\hat{c} \neq c$, unless $\gcd(c, n) \neq 1$.²

3. Obtain the decryption \hat{m} of \hat{c} from the decryption oracle.

Note that $\hat{m} \equiv \hat{c}^d \equiv (cx^e)^d \equiv c^d x^{ed} \equiv m \cdot x \bmod n$

4. Compute $m = \hat{m} \cdot x^{-1} \bmod n$.

This attack looks very artificial. Indeed, this is a very strong attack. However, if we can design encryption schemes that resist even this very strong attack, then we're guaranteed resistance to weaker forms of attacks that might be practical such as the restricted chosen-ciphertext attack.

Countermeasure: Add some prescribed *formatting* to m prior to encryption. After decrypting the ciphertext c , if the plaintext is not properly formatted, then c is rejected (so the decryption oracle does not return a plaintext).

²If $\gcd(c, n) \neq 1$, then Eve knows a prime factor of n from $\gcd(c, n)$, then she totally breaks the RSA encryption scheme.

So, RSA encryption should incorporate **salting** and **formatting**.

secure public-key encryption scheme

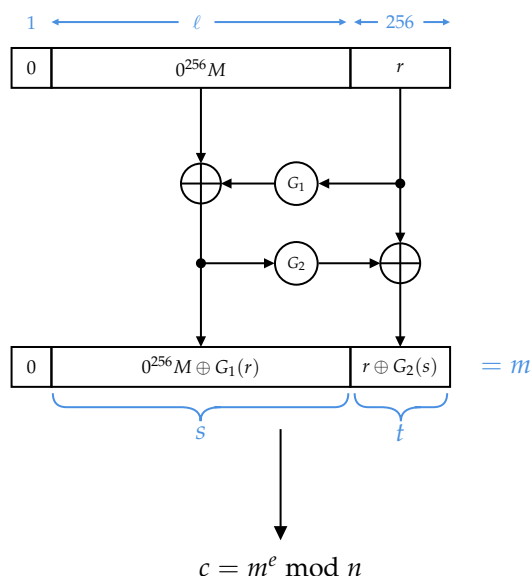
A public-key encryption scheme is **secure** if it is semantically secure against chosen-ciphertext attack by a computationally bounded adversary.

To *break* a public-key encryption scheme, E should do:

1. E is given a *challenge ciphertext* c (and the public key (n, e)).
2. E has a decryption oracle: she can present any ciphertexts for decryption except for c itself.
3. After a feasible amount of computation, E should learn *something* about the plaintext m that corresponds to c (other than its length).

RSA Optimal Asymmetric Encryption Padding (OAEP)

Encryption:



- k = bitlength of n (e.g., 3072)
- $\ell = k - 256 - 1$
- $r \in_R \{0, 1\}^{256}$ (salt)
- M plaintext ($\ell - 256$ bits)
- $G_1 : \{0, 1\}^{256} \rightarrow \{0, 1\}^\ell$
- $G_2 : \{0, 1\}^\ell \rightarrow \{0, 1\}^{256}$

G_1, G_2 : masking functions built from $H = \text{SHA256}$.

e.g., $G_1(r) = H(0, r) \parallel H(1, r) \parallel H(2, r) \parallel \dots$.

Decryption: To decrypt c , do the following:

1. Compute $m = c^d \bmod n$.
2. Parse m : $\underbrace{0}_{1} \underbrace{\leftarrow s \rightarrow}_{\ell} \underbrace{\leftarrow t \rightarrow}_{256}$
3. Compute $r = G_2(s) \oplus t$
4. Compute $G_1(r) \oplus s = \underbrace{\leftarrow a \rightarrow}_{256} \underbrace{\leftarrow b \rightarrow}_{\ell-256}$
5. If $a = 0^{256}$, then output $M = b$; else reject c .

Theorem

Suppose that RSAP is intractable. Suppose that G_1, G_2 are random functions. Then RSA-OAEP is a secure public-key encryption scheme.

7.4 Integer Factorization

7.4.1 Review from complexity theory

Big-O notation, little-o notation. E.g., $1/n = o(1)$.

polynomial-time algorithm

One whose worst-case running time function is of the form $O(n^c)$, where n is the input size and c is a constant.

exponential-time algorithm

One whose worst-case running time function is not of the form $O(n^c)$.

In this course, **fully** exponential-time functions are of the form 2^{cn} , where c is a constant, e.g., $O(2^{n/2})$.

subexponential-time algorithm

One whose worst-case running time function is of the form $2^{o(n)}$, and not of the form $O(n^c)$ for any constant c , e.g., $O(2^{\sqrt{n}})$.

Roughly speaking, “polynomial-time = efficient”, “fully exponential-time = terribly inefficient”, “subexponential-time = inefficient, but not terribly so”.

Example: Trial Division

Consider the following algorithm (trial division) for factoring RSA-moduli n . Trial divide n by the primes $2, 3, 5, 7, \dots, \lfloor \sqrt{n} \rfloor$. If any of these, say ℓ , divides n , then stop and output the factor ℓ of n .

The running time of this method is at most \sqrt{n} trial divisions, which is $O(\sqrt{n})$. Is this a polynomial-time algorithm for factoring RSA moduli?

NO! The input size is $k = O(\log n)$, and the running time is $O(\sqrt{n}) = O(2^{k/2})$, since $n = 2^k$. This is fully exponential time in k .

Let A be an algorithm whose inputs are elements of the integers modulo n , \mathbb{Z}_n , or an integer n (so the input size is $O(n \log n)$). If the expected running time of A is of the form

$$L_n[\alpha, c] = O\left(\exp\left((c + o(1))(\log_e n)^\alpha (\log_e \log_e n)^{1-\alpha}\right)\right),$$

where c is a positive constant, and α is a constant satisfying $0 < \alpha < 1$, then A is a *subexponential-time* algorithm.

Note that if $\alpha = 0$, $L_n[0, c] = O((\log n)^{c+o(1)})$, which is *polytime*. If $\alpha = 1$, $L_n[1, c] = O(n^{c+o(1)})$, which is *fully exponential time*.

7.4.2 Factoring Algorithms

Special-Purpose Factoring Algorithms

Examples: Trial division, Pollard’s $p - 1$ algorithm, Pollard’s ρ algorithm, elliptic curve factoring algorithm, special number field sieve.

Summary Factoring is believed to be a hard problem. However, we have no proof or theoretical evidence that factoring is indeed hard. However, factoring is known to be easy on a quantum computer (Shor’s algorithm). The largest number factored with Shor’s algorithm is the number 21. The big open question is whether large-scale quantum computers can ever be built. 512-bit RSA is considered insecure today. 1024-bit RSA is considered risky today (but still deployed). Applications are moving to 2048-bit and 3072-bit RSA.

7.5 RSA Signature Scheme

Recall Algorithm 25 (RSA key generation), Algorithm 28 (RSA signature generation) and Algorithm 29 (RSA signature verification).

We require that RSAP be *intractable*, since otherwise E could forge A ’s signature as follows:

1. Select arbitrary m .
2. Compute $M = H(m)$.
3. Solve $s^e \equiv M \pmod{n}$ for s .
4. Then s is A ’s signature on m .

If H is not *preimage resistant*, and the range of H is $[0, n - 1]$, E can forge signatures as follows:

1. Select $s \in_R [0, n - 1]$.
2. Compute $M = s^e \pmod{n}$.
3. Find m such that $H(m) = M$.
4. Then s is A ’s signature on m .

If H is not *2nd preimage resistant*, E could forge signatures as follows:

1. Suppose that (m, s) is a valid signed message.
2. Find an $m', m \neq m'$, such that $H(m) = H(m')$.
3. Then (m', s) is a valid signed message.

If H is not *collision resistant*, E could forge signatures as follows:

1. Select m_1, m_2 such that $H(m_1) = H(m_2)$, where m_1 and m_2 are two distinct messages.
2. Induce A to sign m_1 : $s = H(m_1)^d \pmod{n}$.
3. Then s is also A ’s signature on m_2 .

7.5.1 Attack Model

Goals of the adversary:

1. *Total break*: E recovers A ’s private key, or a method for systematically forging A ’s signatures (i.e., E can compute A ’s signature for arbitrary messages).
2. *Existential forgery*: E forges A ’s signature for a single message; E may not have any control over the content or structure of this message.

Types of attacks E can launch:

1. *Key-only attack*: The only information E has is A ’s public key.
2. *Known-message attack*: E knows some message/signature pairs.
3. *Chosen-message attack*: E has access to a signing oracle which it can use to obtain A ’s signatures on some messages of its choosing.

7.5.2 Security of a Signature Scheme

security of a signature scheme

A signature scheme is said to be **secure** if it is existentially unforgeable by a computationally bounded adversary who launches a chosen-message attack.

Note: The adversary has access to a signing oracle. Its goal is to compute a single valid message/signature pair for any message that was not previously given to the signing oracle.

Is the basic RSA signature scheme secure?

NO, if H is SHA-256;³ YES, if H is a ‘full domain’ hash function.

Full Domain Hash RSA (RSA-FDH)

Same as the basic RSA signature scheme, except that the hash function is $H : \{0,1\}^* \rightarrow [0, n-1]$. In practice, one could use: $H(m) = \text{SHA-256}(1, m) \parallel \text{SHA-256}(2, m) \parallel \dots \parallel \text{SHA-256}(t, m)$.

Theorem (Bellare & Rogaway, 1996)

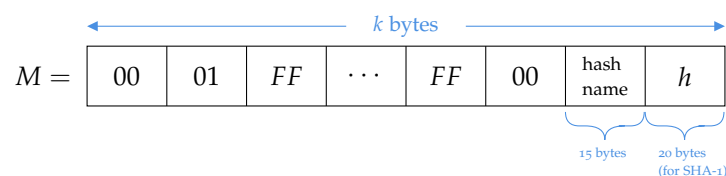
If RSAP is intractable and H is a random function, then RSA-FDH is a secure signature scheme.

7.6 RSA PKCS #1 v1.5 Signatures (1993)

To sign $m \in \{0,1\}^*$, Alice does:

Algorithm 30: RSA PKCS #1 v1.5 signature generations

- 1 Compute $h \in H(m)$, where H is a hash function from an approved list.
- 2 Format h :



// $k = \text{byte length of } n$, e.g., $k = 384$

- 3 Compute $s = M^d \bmod n$.
- 4 Send (m, s)

To verify the signature, Bob does

Algorithm 31: RSA PKCS #1 v1.5 signature verification

- 1 Obtain an authentic copy of Alice's public key (n, e) .
- 2 Compute $M = s^e \bmod n$; write M as a byte string of length k .
- 3 Check *formatting*: First byte is 00; Second byte is 01; Consecutive FF bytes, followed by 00 byte.
- 4 From the next 15 bytes, get the hash name, say SHA-1.
- 5 Let $h = \text{next 20 bytes}$.
- 6 Compute $h' = H(m)$.
- 7 Accept if $h = h'$.

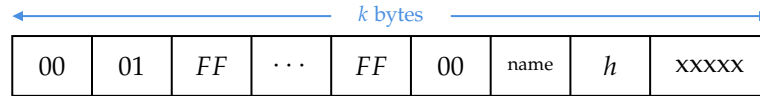
³Details not covered in this course.

7.6.1 Bleichenbacher's Attack (2006)

Breaking RSA signature by hand. OpenSSL had this flaw. Also, Firefox, Adobe Acrobat, SUN's JRE library...

Assumptions:

- $e = 3$, $H = \text{SHA-1}$ (WLOG)
- n is 3072 bits long (WLOG)
- The verifier does not check that there are no bytes after h .



Algorithm 32: Bleichenbacher's Attack

1 Select a message $m \in \{0, 1\}^*$.

2 Compute $h = H(m)$.

3 Let $D =$ bytes

4 Let $N = 2^{288} - D$, and check that $3 \mid N$ (if not, change m slightly).

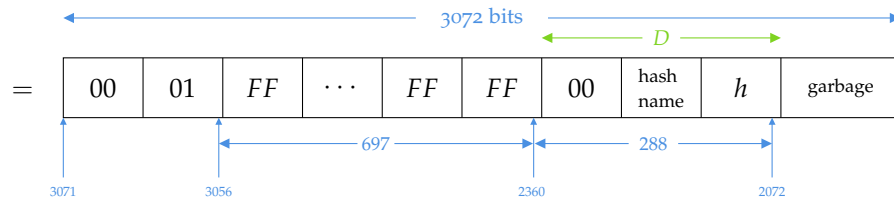
5 Let $s = 2^{1019} - (2^{34} \cdot N/3)$.

6 Output (m, s) .

We claim that (m, s) will be accepted by the verifier.

Proof:

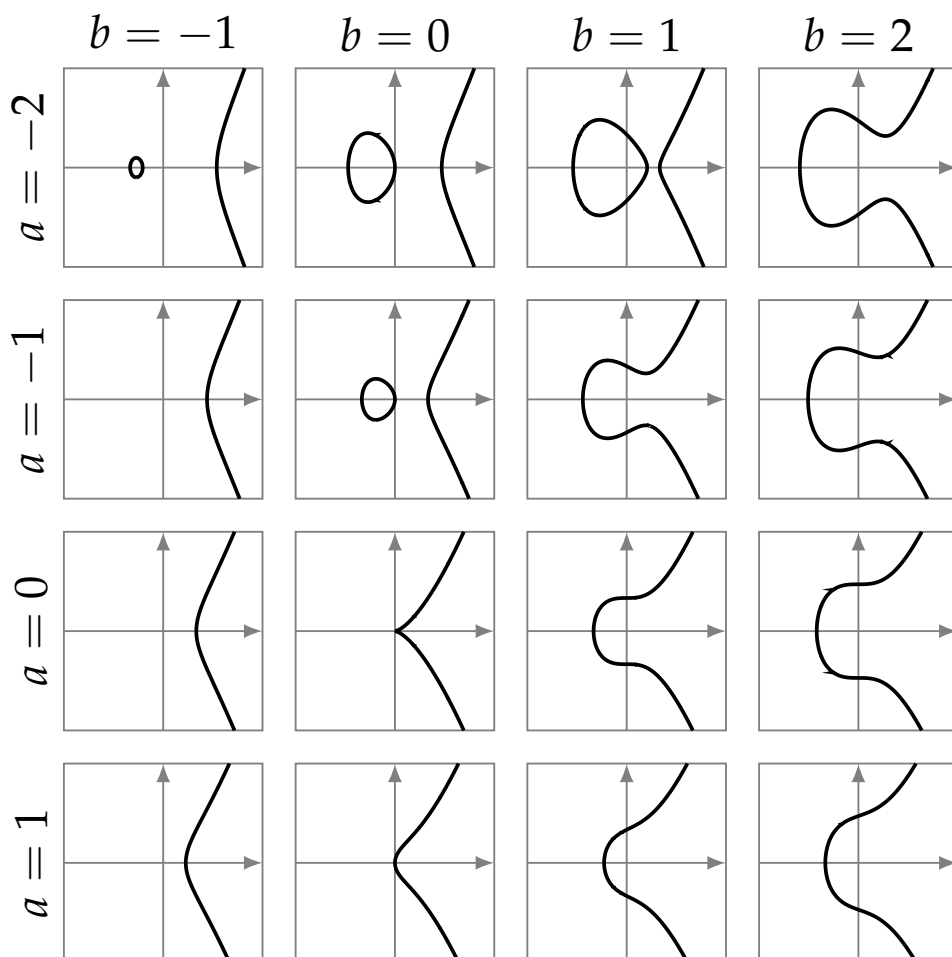
$$\begin{aligned}
 M &= s^e \mod n \\
 &= (2^{1019} - 2^{34}N/3)^3 \mod n \\
 &= 2^{3057} - 2^{2072}N + \underbrace{2^{1087}N^2/3 - (2^{34}N/3)^3}_{\text{garbage}} \mod n \\
 &= 2^{3057} - 2^{2072}(2^{288} - D) + \underbrace{\text{garbage}}_{\geq 0 \text{ and } < 2^{2072}} \mod n \\
 &= 2^{3057} - 2^{2072}(2^{288} - D) + \text{garbage} \\
 &= 2^{2360}(2^{697} - 1) + 2^{2072}D + \text{garbage}
 \end{aligned}$$



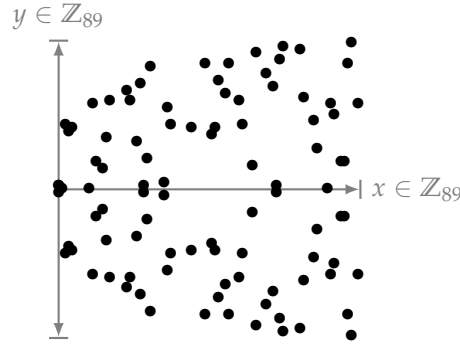
The countermeasure to Bleichenbacher's attack is to check, in step 5 of Algorithm 32, that there are no more bytes after the hash value has been extracted.

Elliptic Curve Cryptography

8.1 Elliptic Curves



$$E/\mathbb{R} = Y^2 = X^3 + aX + b$$



$$E/\mathbb{Z}_{89} : Y^2 = X^3 - 2X + 1, \#E(\mathbb{Z}_{89}) = 96$$

Picture from <https://www.iacr.org/authors/tikz/>.

Let $F = \mathbb{R}$ or $F = \mathbb{Z}_p$ (where $p \geq 5$ is prime)

elliptic curve E over F

An **elliptic curve E over F** is defined by an equation

$$E/F : Y^2 = X^3 + aX + b,$$

where $a, b \in F$ with $4a^3 + 27b^2 \neq 0$.

This is the “Weierstrass form” of an elliptic curve. The restriction on a and b ensures that the polynomial on RHS has no repeated roots.

F -rational points on E

The set of **F -rational points on E** is

$$E(F) = \{(x, y) \in F \times F : y^2 = x^3 + ax + b\} \cup \{\infty\},$$

where ∞ is a special point called the **point at infinity**.

Example:

Consider $E/\mathbb{Z}_{11} : Y^2 = X^3 + X + 6$. Then

$$E(\mathbb{Z}_{11}) = \{\infty, (2, 4), (2, 7), (3, 5), (3, 6), (5, 2), (5, 9), (7, 2), (7, 9), (8, 3), (8, 8), (10, 2), (10, 9)\},$$

so $\#E(\mathbb{Z}_{11}) = 13$ where $\#S$ = cardinality of S .

Note that $\#E(\mathbb{Z}_p)$ is finite. It's easy to see that $1 \leq \#E(\mathbb{Z}_p) \leq 2p + 1$. First, we have the point at ∞ . Then p is because there are p possible choices for x -coordinates; 2 is because for each x , there are at most 2 y -coordinates on the curve.

In fact, we have the following:

Hasse's Theorem

Let E be an elliptic curve defined over \mathbb{Z}_p . Then $(\sqrt{p} - 1)^2 \leq \#E(\mathbb{Z}_p) \leq (\sqrt{p} + 1)^2$.

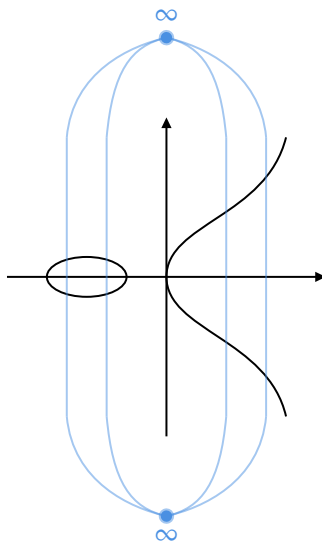
Hence, $\#E(\mathbb{Z}_p) \approx p$.

8.1.1 Point Addition

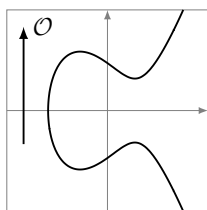
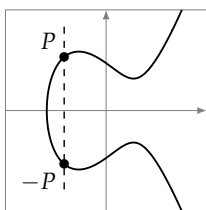
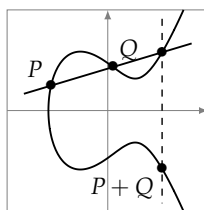
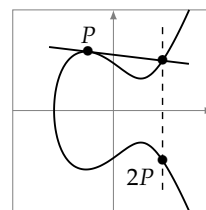
There is a 'natural' way to add two points in $E(F)$ to get a third point in $E(F)$.

Geometric description of the addition rule

Let E be an elliptic curve over \mathbb{R} . Think of ∞ as an imaginary point through which every vertical line passes (in either direction).



The **geometric rule**: Let $P, Q \in E(F)$. Let $T \in E(F)$ be the third point of intersection of the line ℓ through P and Q with the elliptic curve. Then $P + Q$ is the reflection of T in the X -axis.

Neutral element \mathcal{O} Inverse element $-P$ Addition $P + Q$
"Chord rule"Doubling $P + P$
"Tangent rule"

Algebraic description of the geometric rule

Let E be an elliptic curve defined over E .

(A1) $P + \infty = \infty + P = P$, for all $P \in E(F)$.

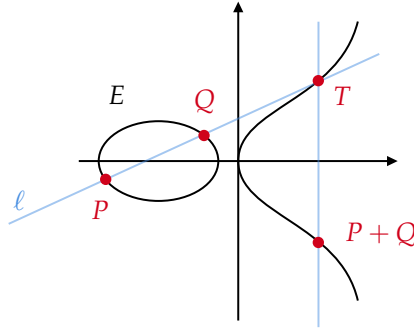
(A2) If $P = (x, y) \in E(F)$, then $-P = (x, -y)$, also $-\infty = \infty$. Also, $P + (-P) = (-P) + P = \infty$, for all $P \in E(F)$.

(A3) (Chord rule) Let $P = (x_1, y_1)$, $Q = (x_2, y_2) \in E(F)$, with $P \neq \pm Q$. Then $P + Q = (x_3, y_3)$, where

$$x_3 = \lambda^2 - x_1 - x_2, \quad y_3 = -y_1 + \lambda(x_1 - x_3), \quad \text{and } \lambda = \frac{y_2 - y_1}{x_2 - x_1}.$$

(A4) (Tangent rule) Let $P = (x_1, y_1) \in E(F)$, with $P \neq -P$. Then $P + P = (x_3, y_3)$, where

$$x_3 = \lambda^2 - 2x_1, \quad y_3 = -y_1 + \lambda(x_1 - x_3), \quad \text{and } \lambda = \frac{3x_1^2 + a}{2y_1}.$$

Derivation of the addition formula in A3

Let $P = (x_1, y_1)$, $Q = (x_2, y_2) \in E(F)$, with $P \neq \pm Q$. The equation of the line ℓ through P and Q is

$$\ell : Y = y_1 + \lambda(X - x_1), \quad (*)$$

where $\lambda = \frac{y_2 - y_1}{x_2 - x_1}$ is the slope of ℓ .

To find the third point of intersection of ℓ with E , we substitute $(*)$ into $E : Y^2 = X^3 + aX + b$ to get:

$$[y_1 + \lambda(X - x_1)]^2 = X^3 + aX + b,$$

so, $X^3 + aX + b - [y_1 + \lambda(X - x_1)]^2 = 0$. This equation has two solutions in F , namely x_1 and x_2 . Thus, it must have a third solution in F , say x_3 . Thus we have

$$X^3 + aX + b - [y_1 + \lambda(X - x_1)]^2 = (X - x_1)(X - x_2)(X - x_3).$$

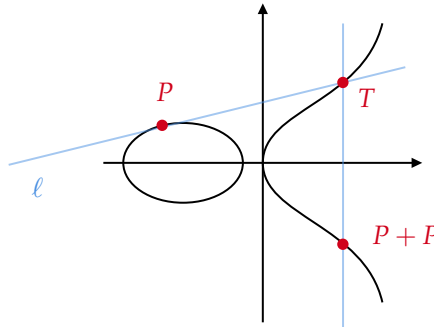
Equating coefficients of X^2 of both sides gives $-\lambda^2 = -(x_1 + x_2 + x_3)$, so

$$x_3 = \lambda^2 - x_1 - x_2$$

The coordinates of T are (x_3, y'_3) , where $y'_3 = y_1 + \lambda(x_3 - x_1)$. Hence the coordinates of $P + Q$ are $(x_3, y_3) = (x_3, -y'_3)$, so

$$y_3 = -y_1 + \lambda(x_1 - x_3)$$

Note on A4 The slope of the tangent line ℓ at $P = (x_1, y_1)$ is: $\lambda = \frac{3x_1^2 + a}{2y_1}$. $E : Y^2 = X^3 + aX + b$, $2Y \cdot \frac{dY}{dX} = 3X^2 + a$, so $\frac{dY}{dX} = \frac{3X^2 + a}{2Y}$.

**Example:**

Consider $E/\mathbb{Z}_{11} : Y^2 = X^3 + X + 6$. The set of \mathbb{Z}_{11} -rational points on E is:

$$E(\mathbb{Z}_{11}) = \{\infty, (2, 4), (2, 7), (3, 5), (3, 6), (5, 2), (5, 9), (7, 2), (7, 9), (8, 3), (8, 8), (10, 2), (10, 9)\},$$

We have

$$(2,4) + (2,7) = \infty$$

$$(2,4) + (8,3) = (5,2)$$

$$(2,4) + (2,4) = (5,9)$$

Note that in the second equation, we find that

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} = \frac{3 - 4}{8 - 2} = -\frac{1}{6} = (-1)(2) = -2 = 9$$

where $6^{-1} = 2$ since $6 \cdot 2 = 1 \pmod{11}$.

$(E(F), +)$ is an Abelian group

The addition rule satisfies the following properties:

(P1) $P + \infty = \infty + P = P$ for all $P \in E(F)$.

(P2) For each $P \in E(F)$, there exists $Q \in E(F)$ such that $P + Q = Q + P = \infty$.

(P3) $P + Q = Q + P$ for all $P, Q \in E(F)$.

(P4) $(P + Q) + R = P + (Q + R)$ for all $P, Q, R \in E(F)$.¹

8.2 Elliptic Curve Discrete Logarithm Problem (ECDLP)

Let $E : Y^2 = X^3 + aX + b$ be an elliptic curve defined over $F = \mathbb{Z}_p$. Let $n = \#E(\mathbb{Z}_p)$, and suppose that n is prime. Note that $n \approx p$ by Hasse's Theorem.

point multiplication

Let $P \in E(\mathbb{Z}_p)$ and let $k \in \mathbb{N}$. Then $kP = \underbrace{P + P + \cdots + P}_k$. Also, $0P = \infty$, and $(-k)P = -(kP)$.

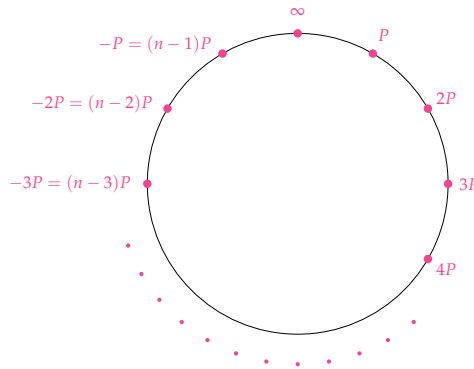
This operations is called **point multiplication**.

Theorem

Let $P \in E(\mathbb{Z}_p)$, $P \neq \infty$. Then

(i) $nP = \infty$; and

(ii) The points $\infty, P, 2P, 3P, \dots, (n-1)P$ are distinct, and so $E(\mathbb{Z}_p) = \{\infty, P, 2P, \dots, (n-1)P\}$.
 P is called a *generator* of $E(\mathbb{Z}_p)$.



¹this property (associativity) requires proof, but we won't do this.

Example:

Let $p = 23$. Consider $E/\mathbb{Z}_{23} : Y^2 = X^3 + X + 4$. $\#E(\mathbb{Z}_{23}) = 29$. The 29 points in $E(\mathbb{Z}_{23})$ are:

(0, 2)	(0, 21)	(1, 11)	(1, 12)	(4, 7)	(4, 16)	(7, 3)
(7, 20)	(8, 8)	(8, 15)	(9, 11)	(9, 12)	(10, 5)	(10, 18)
(11, 9)	(11, 14)	(13, 11)	(13, 12)	(14, 5)	(14, 18)	(15, 6)
(15, 17)	(17, 9)	(17, 14)	(18, 9)	(18, 14)	(22, 5)	(22, 18)
∞						

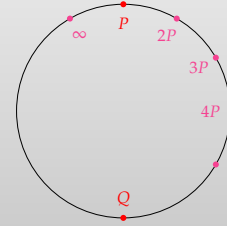
From the theorem, every point except for ∞ can serve as a generator. Let's take $(0, 2)$ as a generator:

$1P = (0, 2)$	$11P = (10, 5)$	$21P = (14, 18)$
$2P = (13, 12)$	$12P = (17, 9)$	$22P = (15, 17)$
$3P = (11, 9)$	$13P = (8, 15)$	$23P = (9, 12)$
$4P = (1, 12)$	$14P = (18, 9)$	$24P = (7, 3)$
$5P = (7, 20)$	$15P = (18, 14)$	$25P = (1, 11)$
$6P = (9, 11)$	$16P = (8, 8)$	$26P = (11, 14)$
$7P = (15, 6)$	$17P = (17, 14)$	$27P = (13, 11)$
$8P = (14, 5)$	$18P = (10, 18)$	$28P = (0, 21)$
$9P = (4, 7)$	$19P = (22, 18)$	$29P = \infty$
$10P = (22, 5)$	$20P = (4, 16)$	

Elliptic curve discrete logarithm problem

The elliptic curve discrete logarithm problem (ECDLP) is the following:

Given $E, p, n, P \in E(\mathbb{Z}_p)$ (with $P \neq \infty$), and $Q \in E(\mathbb{Z}_p)$, find the integer $\ell \in [0, n-1]$ such that $Q = \ell P$. The integer ℓ is called the *discrete logarithm* of Q to the base P , written $\ell = \log_P Q$.

**Example:**

ECDLP instance to the last example: Given $P = (0, 2)$ and $Q = (10, 18)$, find $\ell \in [0, 28]$ such that $Q = \ell P$. Solution is $\ell = 18$.

1. Brute force

The ECDLP can be solved by computing $P, 2P, 3P, \dots$, until Q is encountered. This attack takes time $O(n)$ point additions, or $O(p)$ point additions (recall that $n \approx p$). This is *fully exponential* since an ECDLP instance has size $O(\log p)$ bits.

2. Shank's Algorithm (for solving ECDLP)

Idea Let $M = \lceil \sqrt{n} \rceil$. By the division algorithm, there exist *unique* integers q, r with $\ell = qm + r$, where $0 \leq r < m$ and $0 \leq q < m$. Hence $\ell - qm = r$. Multiplying both sides by P gives $\ell P - qmP = rP$, so

$$Q - q(mP) = rP$$

This equation suggests the following algorithm for finding $\ell = \log_p Q$:

Algorithm 33: Shank's algorithm

```

1 foreach  $r \in [0, m - 1]$  do
2   Compute  $rP$ .
3   Store  $(rP, r)$  in a sorted table.
4 Compute  $M = mP$ .
5 foreach  $q \in [0, m - 1]$  do
6   Compute  $R = Q - qM$  and look it up in the table.
7   if  $R = rP$  then
8     Output " $\ell = qm + r$ " and STOP.
  
```

Runtime: $O(m) = O(\sqrt{n}) = O(\sqrt{p})$ point additions. *Storage:* $O(\sqrt{p})$ points.

3. Pollard's Algorithm

The ECDLP can be solved in time $O(\sqrt{p})$ point additions and negligible storage. Moreover VW collisions search can be used to perfectly parallelize Pollard's algorithm.

Note:

Pollard's algorithm is fastest method known for solving the ECDLP (except for some very special elliptic curves that can easily avoided in practice).

4. Shor's Algorithm

The ECDLP can be solved in polynomial time on a quantum computer.

8.3 Elliptic Curve Cryptography

ECC was invented independently by Neal Koblitz, Victor Miller in 1985.

RSA is a perfectly good public encryption scheme and signature scheme. *Why ECC?*

RSA Security is based on intractability of integer factorization. Fastest known attacks take subexponential time.

ECC Security is based on intractability of ECDLP. Fastest known attacks take fully exponential time.

Hence, elliptic curve cryptographic can use *smaller parameters* than their RSA counterparts, while achieving the same security level.

Example:

Security level	Bitlength of RSA n	Bitlength of ECC p
80	1,024	160
112	2,048	224
128	3,072	256
192	7,680	384
256	15,360	512

Smaller parameters \rightarrow faster and smaller

8.3.1 Elliptic Curves in Practice

Deployments of ECC generally use an elliptic curve that has been standardized (e.g., by NIST). Examples include: P-256, Curve25519, P-384.

P-256 Elliptic Curve

P-256 is an elliptic curve chosen by the National Security Agency in 1998 for U.S. government use. P-256 should be used for applications that require the 128-bit security level.

$p = 2^{256} - 2^{224} + 2^{192} + 2^{96} - 1$ is prime.

P-256 is the elliptic curve $E : Y^2 = X^3 - 3X + b$ over \mathbb{Z}_p , where

$$b = 41058363725152142129326129780047268409114441015993725554835256314039467401291.$$

We have $n = \#E(\mathbb{Z}_p)$ is prime, where

$$n = 115792089210356248762697446949407573529996955224135760342422259061068512044369.$$

Curve25519

Selected by Dan Bernstein in 2005, and developed with Tanja Lange and others. Curve25519 should be used for applications that require the 128-bit security level.

$p = 2^{255} - 19$ is prime.

Curve25519 is the elliptic curve $E : Y^2 = X^3 + 486662X^2 + X$ over \mathbb{Z}_p . Note: The curve is in ‘Montgomery form’.

$\#E(\mathbb{Z}_p) = 8n$, where n is the following 253-bit prime:

$$2^{252} + 27742317777372353535851937790883648493.$$

Reference implementation in C: <https://cr.yp.to/ecdh.html>

P-384 Elliptic Curve

P-384 is an elliptic curve chosen by the National Security Agency in 1998 for U.S. government use. P-384 should be used for applications that require the 192-bit security level.

$p = 2^{384} - 2^{128} - 2^{96} + 2^{32} - 1$ is prime.

P-384 is the curve $E : y^2 = x^3 - 3x + b$ over \mathbb{Z}_p , where

$$b = 27580193559959705877849011840389048093056905856361568521428707301988689241309860865136260764883745107765439761230575.$$

We have $n = \#E(\mathbb{Z}_p)$ is prime, where

$$n = 39402006196394479212279040100143613805079739270465446667946905279627659399113263569398956308152294913554433653942643.$$

8.3.2 Modular Reduction

The primes $p = 2^{256} - 2^{224} + 2^{192} + 2^{96} - 1$ and $p = 2^{384} - 2^{128} - 2^{96} + 2^{32} - 1$ used in P-256 and P-384 were chosen because the operation of reduction modulo p can be easily and efficiently implemented on a 32-bit machine without doing long division. Long division is bit-oriented, so it’s very tedious to implement in software.

To illustrate this technique, let’s consider reduction modulo the prime $p = 2^{192} - 2^{64} - 1$ on a 64-bit machine. ²

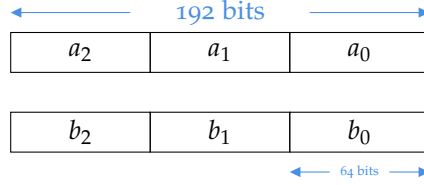
²By a 64-bit machine, prof means a computer with built-in instructions for addition, subtraction, multiplication of 64-bit integers.

Now $p = 2^{192} - 2^{64} - 1$. Suppose that $a, b \in [0, p - 1]$. We wish to compute $c = a \cdot b \bmod p$.

We have

$$a = a_2 2^{128} + a_1 2^{64} + a_0, \quad \text{and} \quad b = b_2 2^{128} + b_1 2^{64} + b_0,$$

where $a_0, a_1, a_2, b_0, b_1, b_2$ are 64-bit integers.



We first compute $d = a \cdot b$ using ordinary 2^{64} -base integer arithmetic. Let

$$d = d_5 2^{320} + d_4 2^{256} + d_3 2^{192} + d_2 2^{128} + d_1 2^{64} + d_0,$$

where each d_i is a 64-bit integer.



We now need to reduce d modulo p .

We have

$$\begin{aligned} 2^{192} &\equiv 2^{64} + 1 \pmod{p}, \\ 2^{256} &\equiv 2^{128} + 2^{64} \pmod{p}, \\ 2^{320} &\equiv 2^{192} + 2^{128} \equiv 2^{128} + 2^{64} + 1 \pmod{p}. \end{aligned}$$

Hence

$$\begin{aligned} c &\equiv d \pmod{p} \\ &= d_5 2^{320} + d_4 2^{256} + d_3 2^{192} + d_2 2^{128} + d_1 2^{64} + d_0 \pmod{p} \\ &= d_5 (2^{128} + 2^{64} + 1) + d_4 (2^{128} + 2^{64}) + d_3 (2^{64} + 1) + d_2 2^{128} + d_1 2^{64} + d_0 \pmod{p} \\ &= \boxed{d_5 \mid d_5 \mid d_5} + \boxed{d_4 \mid d_4 \mid 0} + \boxed{0 \mid d_3 \mid d_3} + \boxed{d_2 \mid d_1 \mid d_0} \pmod{p} \end{aligned}$$

This suggests the following algorithm for computing $c = a \cdot b \bmod p$:

Algorithm 34: $c = a \cdot b \bmod p$

Input: $a, b \in [0, p - 1]$

Output: $c = a \cdot b \bmod p$ where $p = 2^{192} - 2^{64} - 1$

- 1 Compute the 384-bit number $d = a \cdot b = (d_5, d_4, d_3, d_2, d_1, d_0)$; where each d_i is a 64-bit integer.
- 2 Define the 192-bit integers:

$$t_1 = (d_2, d_1, d_0), \quad t_2 = (0, d_3, d_3), \quad t_3 = (d_4, d_4, 0), \quad t_4 = (d_5, d_5, d_5).$$

- 3 Compute $c = t_1 + t_2 + t_3 + t_4$.
 - 4 **if** $c \geq p$ **then** repeatedly subtract p from c until $c \in [0, p - 1]$;
 - 5 **return** c
-

Note:

No long division is done.

In step 3, we have $0 \leq c < 4p$, so at most *three* subtractions by p are required in step 4.

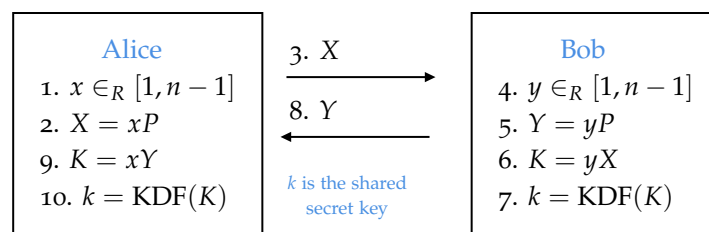
8.4 Elliptic Curve Diffie–Hellman (ECDH)

ECDH is the elliptic curve analogue of a key agreement protocol first proposed by Diffie and Hellman in 1975. The *objective* is for two communicating parties agree upon a shared secret key k . They can use k in a symmetric-key scheme such as HMAC or AES-GCM.

ECDH Domain Parameters (these parameters are public)

- An elliptic curve $E = Y^2 = X^3 + aX + b$ defined over \mathbb{Z}_p . (e.g., P-256)
- $n = \#E(\mathbb{Z}_p)$, where n is prime.
- A generator $P \in E(\mathbb{Z}_p)$.
- A key derivation function KDF (such as a hash function).

8.4.1 Unauthenticated ECDH



Alice and Bob compute the same k since

$$xY = x(yP) = (xy)P = (yx)P = (yx)P = y(xP) = yX.$$

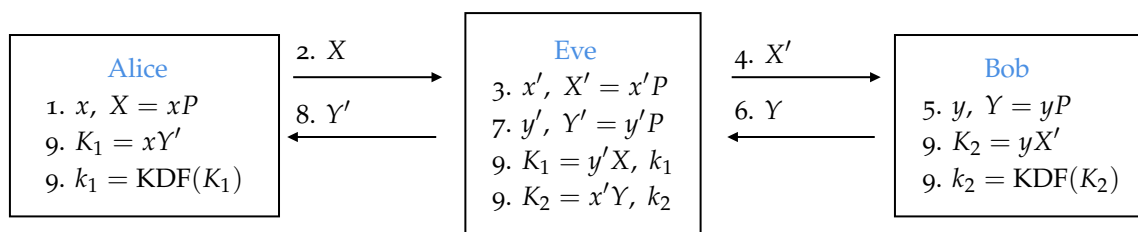
An eavesdropper sees X, Y (and the domain parameters). Her goal is to compute $K = xY = yX$; this is called the **ECDH problem**. The fastest way known to solve this problem is to first compute x or y (ECDLP). Therefore, assuming ECDLP is intractable, ECDH problem is also intractable.

This basic version of ECDH is authenticated because X and Y are not authenticated. When Alice receives Y from Bob, she isn't sure that Y was indeed by Bob.

Malicious-Intruder-in-the-Middle (MITM) Attack

This attack is more commonly called a “Man-in-the-Middle” attack, but the instructor sees no reason why one has to be a man to run the attack. So “Malicious-Intruder-in-the-Middle” is more descriptive.

Since X and Y are *not authenticated*, unauthenticated ECDH is vulnerable to a MITM attack.



Suppose now that Alice sends $c = E_{k_1}(m)$ to Bob, where $E = \text{AES-GCM}$. Eve intercepts c , computes $m = E_{k_1}^{-1}(c)$, $c' = E_{k_2}(m)$ and sends c' to Bob. Bob computes $m = E_{k_2}^{-1}(c')$. Therefore, Bob and Eve can both decrypt the plaintext m that Alice sends. Thus Eve has compromised the confidentiality of Alice's transmissions. Unbeknown to Alice and Bob, Eve has also computed m .

8.4.2 Authenticated ECDH

Alice sends Bob X , her RSA/ECDSA signature on X , and a certificate³ for her RSA/ECDSA public key. Bob verifies the certificate, and then uses Alice's public key to verify the signature on X . Similarly, Bob signs Y .

Note that MITM attack is thwarted: when Eve replaces Bob's Y with Y' , Eve is unable to compute the correct signature on Y' assuming that the signature scheme is secure, so Alice would reject Y' .

8.5 Case Study: ECDH in Google

In November 2011, Google started using ECDH as its default key establishment mechanism in TLS-secured applications including Gmail and encrypted search. ECDH was chosen because it provides forward secrecy. The elliptic curve used was P-256.

8.5.1 TLS (as was commonly implemented)

When a web browser (Alice) visits a secured web page (Bob):

1. Bob sends its certificate to Alice.
2. Alice verifies the signature in the certificate using the (intermediate) Certificate Authority's RSA public key. If the certificate verifies, then Alice is assured that she has an authentic copy of Bob's RSA public key.
3. Alice selects a random session key k , and encrypts k with Bob's RSA public key. Alice sends the resulting ciphertext c to Bob.
4. Bob decrypts c using his RSA private key and obtains k .
5. The session key is used to authenticate (with HMAC) and encrypt (with AES) all data exchanged for the remainder of the session.

TLS, as described, does not provide *forward secrecy*.

That is, suppose that an eavesdropper saves a copy of c and the encrypted data. If, at a future point in time, the eavesdropper is able to break into Bob's machine and learn his RSA private key, then the eavesdropper is able to decrypt c and thus recover k and the data that was encrypted with k . Alternatively, a law enforcement agent could demand that Bob hand over his RSA private key.

A key establishment is said to provide *forward secrecy* if compromise of long-term secret keys does not compromise past session keys.

Using ECDH to establish a session key k provides forward secrecy.

8.5.2 TLS as implemented by Google

When a web browser (Alice) visits a secured web page (Bob):

1. Bob sends its certificate to Alice.
2. Alice verifies the signature in the certificate using the (intermediate) Certificate Authority's RSA public key. If the certificate verifies, then Alice is assured that she has an authentic copy of Bob's RSA public key.
3. Alice selects $x \in_R [1, n - 1]$ and sends $X = xP$ to Bob. (P is a fixed point on P-256.)

³an electronic document that asserts that a public key belongs to Alice

4. Bob selects $y \in_R [1, n - 1]$, signs $Y = yP$ with its RSA signing key, and sends Y and the signature to Alice.
5. Alice verifies the signature using Bob's RSA public key, and is thus assured that Y was sent by Bob.
6. Both Alice and Bob compute the shared secret $K = xY = yX = xyP$, and derive a session key k from K .
7. Alice deletes x . Bob deletes y .
8. The session key is used to authenticate and encrypt (with AES-GCM) all data exchanged for the remainder of the session.
9. At the end of the session, Alice and Bob delete k .

Note that forward secrecy is provided. If an eavesdropper such as law enforcement has captured the ciphertext and wishes to decrypt it, law enforcement needs the session key k . However both Alice and Bob have deleted the session key and all the data used to derive the session key.

For further information, see <http://tinyurl.com/GoogleECDH>.

RSA vs DL vs ECC Usage in TLS

Credit to [Nadia Heninger](#). Percentage of https connections that use RSA, DL or ECC as of Nov. 2016:

Key exchange		Signatures	
RSA:	39%	RSA:	99%
DH:	10%	DSA:	> 0%
ECDH:	51%	ECDSA:	1%

8.6 The Elliptic Curve Digital Signature Algorithm (ECDSA)

ECDSA Domain Parameters

- Elliptic curve E , defined over \mathbb{Z}_p . (e.g., P-256)
- $n = \#E(\mathbb{Z}_p)$ (n prime)
- A generator $P \in E(\mathbb{Z}_p)$.
- A collision-resistant hash function H . (e.g., SHA-256)

To generate keys, Alice does

Algorithm 35: ECDSA key generation

- 1 Select $a \in_R [1, n - 1]$, and compute $A = aP$.
 - 2 Alice's public key is A ; her private key is a .
-

Note that computing a from A is an instance of the ECDLP.

To sign a message $M \in \{0,1\}^*$, Alice does the following:

Algorithm 36: ECDSA signature generation

- 1 Compute $m = H(M)$, and interpret m as an integer.
 - 2 Select a per-message secret $k \in_R [1, n-1]$.
 - 3 Compute $R = kP$. Let $r = x(R) \bmod n$, and check that $r \neq 0$. // $x(R)$ is x -coord. of R
 - 4 Compute $s = k^{-1}(m + ar) \bmod n$, and check that $s \neq 0$.
 - 5 Alice's signature on M is (r, s) .
-

We need $r \neq 0$ because in step 4, a is multiplied by r . $s \neq 0$ because we need to find inverses of $s \bmod n$ in signature verification.

Note that a fresh random k should be selected each time Alice signs a message. Suppose Alice signs both M_1 and M_2 using the same k , and assume the adversary learns (r, s_1) and (r, s_2) . From the signature equation, the only unknowns are k and a , then from these two equations, the adversary can solve efficiently. Also, k should be securely destroyed after it is used. Otherwise the only unknown is a , then the adversary can efficiently compute a from k .

Final note that ECDSA is a randomized signature scheme. New per-message secret k is generated every time she signs a message m .

To verify Alice's signature (r, s) on M , Bob does:

Algorithm 37: ECDSA signature verification

- 1 Obtain an authentic copy of Alice's public key A .
 - 2 Check that $1 \leq r, s \leq n-1$.
 - 3 Compute $m = H(M)$.
 - 4 Compute $u_1 = ms^{-1} \bmod n$ and $u_2 = rs^{-1} \bmod n$.
 - 5 Compute $V = u_1P + u_2A$ and verify that $V \neq \infty$.
 - 6 Compute $v = x(V) \bmod n$.
 - 7 Accept $M, (r, s)$ iff $v = r$.
-

$$\begin{aligned}
 s &= k^{-1}(m + ar) \bmod n \\
 \iff k &\equiv s^{-1}(m + ar) \pmod{n} \\
 \iff k &\equiv u_1 + u_2a \pmod{n} \\
 \iff kP &= u_1P + u_2aP \\
 \iff kP &= u_1P + u_2A \\
 \iff kP &= V \\
 \implies x(kP) \bmod n &= x(V) \bmod n \\
 \iff r &= v
 \end{aligned}$$

Thus, if Alice generated $M, (r, s)$, then Bob will accept.

ECDSA is believed to be *secure*, assuming that ECDLP is intractable, and H is a 'secure' hash function.

ECDSA vs. RSA RSA-FDH: 3,072-bit n , ECDSA: 256-bit p .

1. In practice, RSA is used with $e = 3$ or $e = 2^{16} + 1$. So, RSA signature verification ($s^e \bmod n$) is generally faster than ECDSA ($V = u_1P + u_2A$).
2. RSA signature generation ($H(M)^d \bmod n$) is generally slower than ECDSA ($R = kP$).
3. RSA signatures are 3,072 bits in length; ECDSA signatures have bitlength 512 bits.

Benchmarks using openssl on the instructor's MacBook Air:

	# sig. gen. per sec	# sig. ver. per sec
ECDSA with P-256	1730	2295
ECDSA with P-384	807	1207
RSA with 2048-bit n	938	17015
RSA with 4096-bit n	125	4489

Bluetooth Security

9.1 Introduction

Bluetooth is an industry standard for short-range low-power wireless communication technology. It's primarily used to establish wireless personal area networks (WPANs).

Bluetooth v1.0 released in 1999. Bluetooth v5.2 released on December 31, 2019. In 2017, Bluetooth product shipments were more than 3.6 billion. Four variants: BR (basic rate), EDR (enhanced data rate), HS (high speed), LE (low energy). The Bluetooth specification is about 3,000 pages long, which makes a comprehensive security analysis very difficult.

Five basic security services are specified in the Bluetooth standard:

1. **Confidentiality:** preventing information compromise caused by eavesdropping by ensuring that only authorized devices can access and view transmitted data.
2. **Message Integrity:** verifying that a message sent between two Bluetooth devices has not been altered in transit.
3. **Authentication:** verifying the identity of communicating devices based on their Bluetooth addresses.
4. **Authorization:** allowing the control of resources by ensuring that a device is authorized to use a service before permitting it to do so.
5. **Pairing:** creating one or more shared secret keys and the storing of these keys for use in subsequent connections in order to form a trusted device pair.

To establish a long-term secure Bluetooth connection, the two devices have to “pair”.

There are three security mechanisms:

1. Legacy
2. Secure Simple Pairing
3. Secure Connections

(Secure Simple Pairing and Secure Connections are very similar)

Secure Connections

Four association models depending on the I/O capabilities of the devices:

1. **Numeric Comparison:** requires both devices to have displays for 6-digit numbers, and one of the two devices to have an “OK/Reject” confirmation button.
2. **PassKey Entry:** requires one device to have input capability (e.g., keyboard), and the other device to have display (but not input) capability.
3. **Just Works:** designed for the situation where at least one of the pairing devices has neither a display nor a keyboard (e.g. headsets).
4. **Out-of-Band** (vendor specific): designed for devices that support a common additional wireless (e.g. NFC) or wired technology for the purpose of device discovery and cryptographic value exchange.

We shall study Numeric Comparison.

9.2 The Bluetooth security protocol - Numeric Comparison

Notation:

P-256 elliptic curve	E, p, n, P
Alice	the initiating device
Bob	the responder device
A, B	Alice’s and Bob’s 48-bit Bluetooth addresses
N_A, N_B	Alice’s and Bob’s randomly-selected 128-bit nonces
C_B	Bob’s commitment (to his nonce)
V_A, V_B	Alice’s and Bob’s 6-digit verification values
E_A, E_B	Alice’s and Bob’s exchange confirmation values
LK	link key
S_A, S_B	Alice’s and Bob’s signed responses
ACO	authenticated ciphering offset
K_E	encryption key

To prevent MITM attacks, one must ensure that Alice and Bob select their nonces independently of each other. This can be achieved by “commitments”. After exchanging the points X and Y , Bob selects his nonce N_B and sends to Alice a commitment of this nonce. The commitment is a value which does not reveal anything at all about the nonce itself. Later, Bob can prove to Alice that his commitment corresponds to the nonce N_B . But Bob cannot prove to Alice that the commitment corresponds to any other nonce different from N_B . Therefore, after receiving the commitment from Bob, Alice sends her nonce to Bob, and finally Bob sends his nonce to Alice. Alice checks that the nonce indeed corresponds to the previous commitment. In this way, Alice and Bob are forced to select nonces independently of each other.

Phase 1: Public Key Exchange

Purpose: To establish an (unauthenticated) shared secret.

ECDH is used:

Algorithm 38: Bluetooth security protocol phase 1: public key exchange

- 1 Alice selects $x \in_R [1, n-1]$, computes $X = xP$ and sends X to Bob.
 - 2 Bob selects $y \in_R [1, n-1]$, computes $Y = yP$, sends Y to Alice.
Bob computes $K =$ the x -coordinate of yX .
 - 3 Alice computes $K =$ the x -coordinate of xY .
 - 4 The shared secret is K (256 bits).
-

Phase 2: Authentication Stage 1

Purpose: Provide some protection against active MITM attacks.

Algorithm 39: Bluetooth security protocol phase 2: authentication stage 1

- 1 Bob selects nonce $N_B \in_R \{0, 1\}^{128}$, computes a commitment $C_B = \text{HMAC-SHA-256}_{N_B}(Y, X)$, and sends C_B to Alice.
- 2 Alice selects nonce $N_A \in_R \{0, 1\}^{128}$ and sends N_A to Bob.
- 3 Bob sends N_B to Alice.
- 4 Alice checks if $C_B = \text{HMAC-SHA-256}_{N_B}(Y, X)$.
- 5 Alice computes verification value

$$V_A = (\text{SHA-256}(X, Y, N_A, N_B) \bmod 2^{32}) \bmod 10^6,$$

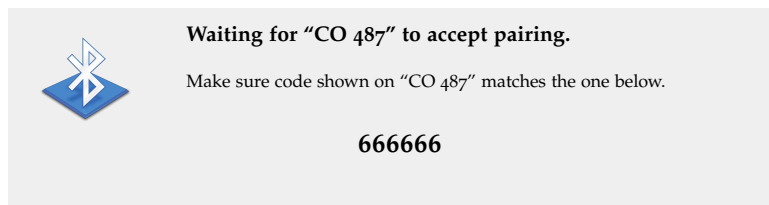
and displays this value.

- 6 Bob computes verification value

$$V_B = (\text{SHA-256}(X, Y, N_A, N_B) \bmod 2^{32}) \bmod 10^6,$$

and displays this value.

- 7 if $V_A = V_B$ then the owner presses the “OK” button.
-



Notes on Phase 2:

- The use of the commitment C_B means that Alice and Bob each selects their nonces before seeing the other party's nonce.
- Failure in step 4 indicates the presence of an attacker, or other transmission error.

- An active MITM attack will result in the two 6-digit verification values V_A, V_B being different with probability 0.999999.

Phase 3: Authentication Stage 2

Purpose: Confirm that both devices have successfully completed the exchange.

Algorithm 40: Bluetooth security protocol phase 3: authentication stage 2

- 1 Alice computes $E_A = 128$ least significant bits of $\text{HMAC-SHA-256}_K(N_A, N_B, A, B)$ and sends exchange confirmation value E_A to Bob.
 - 2 Bob verifies that the 128 least significant bits of $\text{HMAC-SHA-256}_K(N_A, N_B, A, B)$ are equal to E_A , computes exchange confirmation value $E_B = 128$ most significant bits of $\text{HMAC-SHA-256}_K(N_B, N_A, B, A)$, and sends E_B to Alice.
 - 3 Alice verifies the 128 most significant bits of $\text{HMAC-SHA-256}_K(N_B, N_A, B, A)$ are equal to E_B .
-

Phase 4: Link Key Calculation

Purpose: Compute the long-term link key LK .

Algorithm 41: Bluetooth security protocol phase 4: link key calculation

- 1 Both parties compute $LK = 128$ most significant bits of $\text{HMAC-SHA-256}_K(N_A, N_B, A, B)$.
-

Notes:

- The link key is the long-term authentication key.
- The link key is used to maintain the pairing.
- The use of nonces in protocol ensure the freshness of the link key even if long-term ECDH values are used by both sides.

Authentication and Encryption

Purpose: Check that the communicating devices hold the same link key (and thus are paired), and then derive a shared secret input to the encryption-key generation procedure.

Algorithm 42: Bluetooth security protocol: authentication and encryption

- 1 Alice generates $R_A \in_R \{0, 1\}^{128}$ and sends (A, R_A) to Bob.
- 2 Bob generates $R_B \in_R \{0, 1\}^{128}$ and sends (B, R_B) to Alice.
- 3 Alice and Bob computes:
 - (a) Device authentication key $h = 128$ most significant bits of $\text{HMAC-SHA-256}_{LK}(A, B)$.
 - (b) Device authentication confirmation values

$$t = \text{HMAC-SHA-256}_h(R_A, R_B)$$

$$S_A = \text{leftmost 32 bits of } t$$

$$S_B = \text{next 32 bits of } t$$

$$\text{ACO} = \text{next 64 bits of } t$$

- 4 Alice sends S_A to Bob (signed response)
 - 5 Bob sends S_B to Alice (signed response)
 - 6 Alice and Bob compare the signed responses they received with the values they computed.
-

The encryption key K_E derived may vary in length in single byte increments from 1 byte to 16 bytes. The length is set during a negotiation process that occurs between the two communicating devices.

For 128-bit keys: $K_E = 128$ most significant bits of $\text{HMAC-SHA-256}_{LK}(A, B, \text{ACO})$. For smaller ℓ -bit keys, take the ℓ most significant bits, and set the remaining bits to 0.

Encryption is performed using either the E0 stream cipher in Bluetooth version 4.0 and earlier, or AES-CCM (AES-CTR encryption + Cipher Block Chaining MAC) in Bluetooth version 4.1 and later.

9.3 KNOB Attack

Antonioli-Tippenhauer-Rasmussen (August 2019)

See <https://knobattack.com/>

The encryption key negotiation protocol is used by two Bluetooth devices to agree on the length of encryption keys K_E . This was introduced to cope with international encryption regulations and to facilitate security upgrades. However, it turns out that the negotiation is neither *authenticated* nor *encrypted*. So, an active adversary can cause the devices to agree to produce an 8-bit key K_E . This key can then be brute forced in real time (and thus confidentiality is lost).

The attack was implemented, and used to decrypt an encrypted file exchanged over Bluetooth between a Nexus 5 phone and a Motorola G3 phone. Most Bluetooth chips (Intel, Broadcom, Apple...) and devices were vulnerable to the attack.

Countermeasure: Set the key length to 128 bits.

Key Management

10.1 Public Key Management

Key management

A set of techniques and procedures supporting the establishment and maintenance of keying relationships between authorized parties.

We will consider management of public keys that are used for public-key encryption, Diffie-Hellman key agreement, and for verification of digital signatures. It's instructive to think about two scenarios:

1. Suppose that A wishes to use public-key encryption (or hybrid encryption) to encrypt a message for B . To do this, A needs an authentic copy of B 's public key.
 - *Example:* A wishes to send B a confidential email.
 - *Example:* A wishes to send her credit number to B while making an online purchase.
2. Suppose that A receives a message purportedly signed by B . To verify the signature on the message, A needs an authentic copy of B 's public key.
 - *Example:* A person A wishes to verify the authenticity of a software patch that was purportedly signed by Microsoft (B).
 - *Example:* The manager B of a bank branch can authorize financial transactions worth up to \$20,000.

These two scenarios raise some questions and concerns:

- Where does A get B 's public key from?
- How does A know she really has B 's public key?
- How can a bank limit use of B 's public/private key pair?
- What happens if B 's private key is compromised? Who is liable?
- How can a bank revoke B 's public key?
- How can B 's public keys be updated?

Techniques for Distributing Public Keys

1. Point-to-point delivery over a trusted channel.

Trusted courier. One-time user registration. Voice. Embedded in a browser or operating system.

2. Direct access to a trusted public file.

Digitally signed file. The file could be placed on a public website. Anyone could download the file and then use the trusted entity's public signature verification key to verify the authenticity of the file.

3. Use of an on-line trusted server.

Online Certification Status Protocol (OCSP) is a standardized protocol whereby a user can send real-time requests to check the authenticity of a public key. A trusted OCSP server would verify the public key and send a signed confirmation to the user.

4. **Off-line certification authority (CA).**

We will discuss certification authorities.

10.2 Certification Authorities (CAs)

A CA issues *certificates* that bind an entity's identity A and its public key.

A 's certificate Cert_A consists of:

1. Data part D_A : A 's identity, her public key, and other information such as validity period.
2. Signature part S_T : The CA's signature on the data part.

B obtains an authentic copy of A 's public key as follows:

1. Obtain an authentic copy of the CA's public key (e.g., shipped in browsers or in an operating system).
2. Obtain Cert_A (over an unsecured channel).
3. Verify the CA's signature S_T on D_A .

Note that the CA does not have to be trusted with users' private keys. The CA has to be trusted to not create false certificates.

10.3 Public-Key Infrastructures (PKI)

PKI A collection of technologies and processes for managing public keys, their corresponding private keys, and their use by applications.

Some components of a PKI

- Certificate format.
- The certification process.
- Certificate revocation.
- Trust models.
- Certificate distribution.

- **Certificate policy:** Details of intended use and scope of a particular certificate.
- **Certification practices statement (CPS):** Practices and policies followed by a CA.

Although conceptually very simple, there are many practical problems that are encountered when deploying PKI on a large scale. Many of these problems arise from business, legal, and useability considerations.

The problems include

- **Interoperability** (alleviated by standards and certificate formats).
- **Certificate revocation.**

It is estimated that $\gg 500,000$ TLS private keys were compromised due to the Heartbleed bug (discovered in 2014). Lots of certificates needed to be revoked.

- **Liability.**
- **Trust models.**

10.4 Case Study: TLS

SSL (Secure Sockets Layer) was designed by Netscape. TLS is an IETF version of SSL. SSL/TLS is used by web browsers to protect web transactions.

There are many versions of SSL/TLS: SSL 2.0 (1995), SSL 3.0 (1996), ¹ TLS 1.0 (1999), TLS 1.1 (2006), TLS 1.2 (2008), TLS 1.3 (2018). Google removed SSL 3.0 from Chrome only in 2014.

The main components of TLS are:

1. **Handshake protocol:** Allows the server to authenticate itself to the client, and then negotiate cryptographic keys.
2. **Record protocol:** Used to encrypt and authenticate transmitted data.

10.4.1 TLS Handshake Protocol

Phase 1 Establish security capabilities.

Negotiate protocol version, cryptographic algorithms, security levels, etc.

Phase 2 Server authentication and key exchange.

Server sends its certificate, and key exchange parameters (if any).

Phase 3 Client authentication and key exchange.

Client sends its certificate (if available) and key exchange parameters.

Phase 4 Finish.

The main key establishment schemes are:

1. **RSA key transport:** The shared secret k is selected by the client and encrypted with the server's RSA public key. (Not allowed in TLS 1.3)
2. **Elliptic Curve Diffie-Hellman (ECDH)**

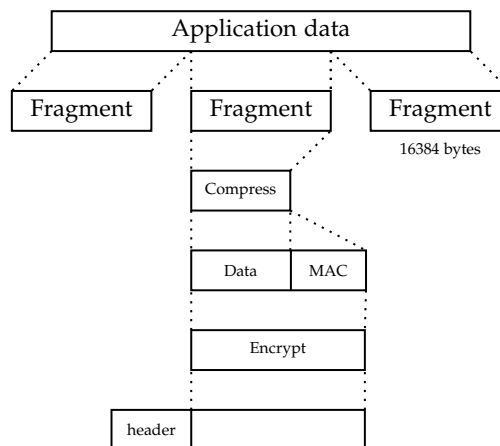
¹almost identical to SSL 2.0

- The server selects a one-time EC Diffie-Hellman public key $X = xP$ and signs it with its RSA or ECDSA signature key.
- The client selects a one-time EC Diffie-Hellman public key $Y = yP$.
- The session key is $K = \text{KDF}(xyP)$.

Note that in both RSA key transport and ECDH, the authentication is one-way: from server to client.

10.4.2 TLS 1.2 Record Protocol

Suppose that client and server share a MAC secret key and a session encryption key.



- **MAC:** HMAC-SHA-1, HMAC-MD5, HMAC-SHA256, ...
- **Symmetric-key encryption:** RC4, ChaCha20, Triple-DES, AES, ...
- **Authenticated encryption:** MAC-then-encrypt, AES-GCM, ChaCha20 + Poly1305².

10.4.3 TLS 1.3

TLS 1.3, approved by **IETF** in August 2018, was a major overhaul of TLS 1.2. Some of the changes made in TLS 1.3 were:

- Removed RC4, Triple-DES, CBC-mode.
- Removed MAC-then-encrypt.
- Removed MD5 and SHA1.
- Removed RSA key transport.
- Mandates AES-GCM (and optionally ChaCha20 + Poly1305).
- All public-key exchanges are based on ephemeral elliptic curve Diffie-Hellman.
- Elliptic curves include P-256, Curve25519 and P-384.

10.5 Public Key Management in TLS

Root CA keys are pre-installed in browsers. Root CAs certify public keys of Intermediate CAs.

Web servers get their public keys certified by Intermediate CAs (perhaps for a fee).

²a MAC scheme similar to GCM

- DigiCert's web server certification business: <https://www.digicert.com/>
- Let's Encrypt. <https://letsencrypt.org/>: Non-profit CA, that provides free web certificates. Automatic certificate issuance; relies on domain validation.

Clients (users) can obtain their own certificates. However, most users do not have their own certificates. If clients do not have certificates, then authentication is only one-way (the server authenticates itself to the client).

10.5.1 Example of an X.509 Certificate

Go to <https://www.cibc.com> and click on the padlock.

- Subject name: Canadian Imperial Bank of Commerce | www.cibc.com
- Issuer name: DigiCert SHA2 Secure Server CA
- Serial number: 08 E9 7A 69 . . . BB
- Signature algorithm: SHA-256 with RSA Encryption.
- Valid from: Apr 22, 2020
- Valid to: May 13, 2022
- Public Key Info/Algorithm: RSA Encryption
- Subject RSA public key info: 2048-bit n, e = 65537
- DigiCert's 2048-bit RSA signature

When a client (a person using a web browser) Alice visits CIBC's website/server (Bob), three public keys are used:

1. CIBC's RSA public key: (n_C, e_C) , private key: d_C . CIBC's public key is certified by the intermediate CA DigiCert-S (full name: "DigiCert SHA2 Secure Server CA").
2. DigiCert-S's RSA public key: (n_S, e_S) , private key: d_S . DigiCert-S's public key is certified by the root CA DigiCert-G (full name: "DigiCert Global Root CA").
3. DigiCert-G's RSA public key: (n_G, e_G) , private key: d_G . DigiCert-G's RSA public key is certified by itself (i.e., it is self-signed), and is embedded in all browsers.

Handshake-1

When Alice visits <https://www.cibc.com>, CIBC sends Alice the following:

1. *CIBC's certificate* which contains CIBC's name, RSA public key (n_C, e_C) , etc., and DigiCert-S's RSA signature s_1 on this data.
2. *DigiCert-S's certificate* which contains DigiCert-S's name, RSA public key (n_S, e_S) , etc., and DigiCert-G's RSA signature s_2 on this data.
3. *DigiCert-G's certificate* which contains DigiCert-G's name, RSA public key (n_G, e_G) , etc., and DigiCert-G's RSA signature s_3 on this data.
4. A randomly selected point $X (= xP)$ on the elliptic curve P-256, and CIBC's RSA signature s_4 on this point.

Handshake-2

Upon receiving the three certificates and (X, s_4) , Alice does:

1. *Verify* DigiCert-G's signature s_3 using DigiCert-G's public key (which is embedded in Alice's browser).
2. *Verify* DigiCert-G's signature s_2 using DigiCert-G's public key (n_G, e_G) , thereby authenticating DigiCert-S's public key.
3. *Verify* DigiCert-S's signature s_1 using DigiCert-S's public key (n_S, e_S) , thereby authenticating CIBC's public key.
4. *Verify* CIBC's signature s_4 using CIBC's public key (n_C, e_C) , thereby obtaining an authentic copy of X .
5. Select a random y and send $Y = yP$ to CIBC.
6. Compute the session key $k = \text{KDF}(yX)$.

Handshake-3

CIBC receives Y and computes the session key $k = \text{KDF}(xY)$.

Alice and CIBC now share a session key k which is used to encrypt and authenticate (with AES-GCM) all communications for the remainder of the session.

DigiCert

Alice and CIBC are relying on DigiCert to do its job well, which includes:

- Carefully check the identify of the certificate holder (CIBC's name and url), and not issue certificates to imposters (e.g., <https://www.cibc.com>).
- Carefully guard its private keys d_G and d_S from disclosure.
- Carefully guard misuse of its private keys d_G and d_S by company employees.

To engender confidence and trust in its certification practice, DigiCert publishes its **Certification Practices Statement** and has its security practices audited by external parties (such as Deloitte or KPMG).

Even though CIBC might have full confidence in DigiCert, there remains the possibility of another Root CA or Intermediate CA issuing fraudulent certificates.

There are several hundred root CA public keys in a browser.

- Mistakenly-issued certificate.
- Maliciously-issued certificate.
- Maliciously-requested certificate.

CAB (CA/Browser Forum) Voluntary consortium of CAs and browser vendors that prepares and maintains industry guidelines concerning the issuance and management of certificates. <https://cabforum.org/>

Common CA Database (managed by Mozilla) is a repository of information about CAs whose root and intermediate certificates are included in browsers. <https://ccadb.org>

CA Breaches

DigiNotar was a Netherlands-based CA, whose root CA public key was embedded in all browsers. DigiNotar issued TLS certificates as well as certificates for government agencies in the Netherlands. In July 2011, DigiNotar issued several hundred fraudulent certificates for domains including `aol.com`, `microsoft.com`, and `*.google.com`. Shortly after, $\approx 300,000$ gmail users from Iran were redirected to websites that looked like the gmail site. The fraudulent web sites would appear to be valid and secured to the gmail users (and the users would establish a secret key with the fraudulent web site using TLS). Presumably, the fraudulent web site then captured the gmail userids and passwords of the users. See <http://tinyurl.com/SlateDigiNotar>. It is suspected that the DigiNotar attack was mounted by an individual or organization in Iran, but this has not been proven.

In 2015 Google discovered that the root CA CNNIC (China Internet Network Information Center) had issued an intermediate CA certificate to an Egyptian company MCS Holdings, which in turn issued fraudulent certificates for several Google domains. See: <http://tinyurl.com/GoogleCNNIC>.

In 2016, it was discovered that the Chinese root CA WoSign had issued fraudulent certificates for several domains including GitHub and Alibaba. See: <http://tinyurl.com/GoogleWoSign>.

In response to the CA breaches, DigiNotar, CNNIC and WoSign's root CA keys were removed from the list of trusted root CA keys in browsers.

Root CA's had to meet certain requirements to qualify for Extended Validation (EV) status.

Another effort is Google's Certificate Transparency. This is a certificate logging mechanism to allow anyone to check which certificates a CA has issued. Auditors monitor CAs to watch for malicious behaviour. Domain name owners monitor the logs to check for certificates issues for their domains. <https://certificate.transparency.dev/>

Random Bit Generation

11.1 Introduction

random bit generator

A **random bit generator** (RBG) is a device that outputs a sequence of independent and unbiased bits.

Some applications of RBGs in cryptography:

1. To generate secret cryptographic keys. Examples: secret key k for AES, HMAC and AES-GCM; primes p and q in RSA; secret key in ECDSA.
2. To generate per-message secrets k in ECDSA.
3. Padding bits for schemes such as RSA-OAEP.

Main security requirement: The random bits should be unpredictable to an active adversary.

How to generate random bits in practice? Purely random bits must come from some physical process. In other words, they cannot be generated by some mathematical algorithm. The mathematician John von Neumann famously said

Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin.

Physical processes that can generate random bits include

- Repeated tossing of a fair coin.
- Elapsed time between emission of particles during radioactive decay.
- Timing between user key strokes or mouse button clicks.
- Samples from I/O buffers.
- Sample bits from a microphone.
- RDRAND: instruction for obtaining random numbers from Intel/AMD on-chip hardware random number generators.

11.4 Pseudorandom Bit Generation

Random bit generation can be very slow. So, in practice, a random bit string is used to seed a pseudorandom bit generator.

HMAC can be used to derive many pseudorandom bit strings s_1, s_2, \dots from a “random” bit string s :

$$s_1 = \text{HMAC}_s(1), \quad s_2 = \text{HMAC}_s(2), \quad s_3 = \text{HMAC}_s(3), \quad \dots$$

FIDO U2F

FIDO = Fast IDentity Online

U2F = Universal 2nd Factor authentication

FIDO U2F is developed by the FIDO alliance. Members include Amazon, Bank of America, Google, Microsoft, Facebook, PayPal, Samsung, VISA, ...

FIDO mission Reduce the world's reliance on passwords to better secure the web.

Some of the many problems with passwords:

- Selection of weak passwords.
- Reuse of passwords.
- Too many passwords to remember.
- Forgotten passwords.
- Phishing attacks.
- Servers have to securely store many passwords.

Many of these problems can be alleviated to some degree by using a password manager. However, they still have their own security risks, and still vulnerable to phishing attacks. Another countermeasure to phishing attacks is second authentication factor: something we know (e.g., password, PIN), something we have (e.g., phone, token), or something we are (e.g., fingerprint, eye). However, phishing attacks can sometimes be launched by sophisticated adversaries even if a second authentication factor is used. See more on https://citizenlab.ca/2015/08/iran_two_factor_phishing/.

12.1 U2F Protocol

1. Authenticator Registration

- Alice generates a public-private key pair (for ECDSA with the P-256 elliptic curve).
- Alice registers her userid, password, and public key with a web service.

2. User Authentication

- Alice visits the web site (via her web browser).

- She enters her userid and password (which the web server verifies).
- The web server sends Alice a random challenge r .
- Alice uses her private key to sign a message m comprising of r , the web server's URL, and the TLS channel ID of the connection (optional); the signature s is transmitted to the sever.
- The server verifies (m, s) using Alice's public key.

Alice has one public-private key pair for each account. All her private keys can be securely stored on a single *security token*. The ECDSA signature operations are performed by the security token; the private keys never leave the token. Alice can also use different userids for each of her accounts; there is then no linkability between accounts. This makes it easier for Alice to maintain anonymity.

Therefore U2F is phishing-resistant. U2F also resists MITM attacks.

12.2 Google's Titan Security Key

It implements the FIDO U2F standard for two-factor authentication. It's compatible with popular browsers including Chrome.



The tokens are built with a tamper-resistant hardware chip that includes firmware engineered by Google. This helps to ensure that the keys haven't been physically tampered with. It's also built into Pixel 3, 3a, 4, 4a, 5 phones featuring the tamper-resistant Titan M security chip.

To activate the Titan security key, press a button on a Bluetooth/USB device or tap over NFC. This physical "test of user presence" ensures that a signature happens only with the user's consent. It also ensures that malware cannot sign messages when the user is not present.

12.3 FIDO2

It's "Passwordless" authentication. It consists of **CTAP2** and **WebAuthn**.

CTAP2 is Client-to-Authenticator-Protocol specified by FIDO alliance. It uses ECDH with P-256.

WebAuthn is W3C's Web Authentication protocol.¹ This uses RSA-PKCS-1-v1.5 and RSA-PSS.

Passwords (or PINs) are only used in an initialization phase, when a user registers their security device to a browser (rather than to a website). Thus no password use when a user registers with a website, or authenticates to the website.

Support in major browsers (Chrome, Firefox, Safari,)

Native platform support (Windows, Android, iOS,)

FIDO2 is exciting because it has the potential to eliminate passwords from the web and replace them with a more secure and usable authentication method.

¹W3C = World Wide Web Consortium

The Signal Protocol

13.1 Introduction

The Signal Protocol was designed by Moxie Marlinspike and Trevor Perrin.

It is free, open source, and is used in:

- Signal (free messaging app).
- Facebook Messenger (“secret conversations” optional feature in the Messenger app on iOS and Android).
- Skype (“private conversations” optional feature).
- WhatsApp.

WhatsApp is owned by Facebook. Has over 2 billion users (India, Brazil, Mexico, France, UK, ...), and handles 10's of billions of messages everyday. Is banned, or has been temporarily blocked, in several countries. Has very low revenues (it's free, works over WiFi, no advertisements, and no user data to mine except metadata). In Jan 2019, Facebook announced plans to tightly integrate WhatsApp, Facebook Messenger, and Instagram into a single “privacy-focused platform”: <http://tinyurl.com/NYTWhatsApp>. A new privacy policy will be enforced on May 15, 2021.

13.2 Signal Objectives

Participants: Alice, Bob, WhatsApp, ThirdParty (E)

1. **Long-lived sessions.** Alice and Bob establish a long-lived secure communications session. The session lasts until events such as app reinstall or device change.
2. **Asynchronous setting.** Alice can send Bob a secure message even if Bob is offline. Messages can be delayed, delivered out of order, or can be lost entirely without problem.
3. **Fresh session keys.** Each message is encrypted/authenticated with a fresh session key. Encrypt-then-MAC: $c = \text{AES-CBC}_{k_1}(m)$, $t = \text{HMAC}_{h_1}(c)$, where k_1 and h_1 are each 256-bits.
4. **Immediate decryption.** Bob can decrypt a ciphertext as soon as he receives it.
5. **End-to-end encryption.** WhatsApp and E do not possess any of Alice's or Bob's secret keys, nor do they get access to any plaintext.

However, WhatsApp (but not E) does get all the metadata, e.g., who sent a message to whom and when, your contacts, your profile name, etc.

6. **Forward secrecy.** If a party's state is leaked, then none of the previous messages should be compromised (assuming they have been deleted from the state).
7. **Post-compromise security.** Parties recover from a state compromise (if the attacker remains passive).

13.3 Signal Protocol

Here are the cryptographic ingredients used in Signal.

1. AES-CBC: 128-bit IV, 256-bit key.
2. HMAC: with SHA-256, and a 256-bit key.
3. KDF: A key derivation function (either HMAC or HKDF, but we will not get into the details).
4. Curve25519: See page 97.
5. Elliptic curve key pairs: (X, x) .
 $x \in_R [1, n - 1]$ is a secret key and $X = xP$ is the corresponding public key.
6. ECDH.
7. EdDSA: (an ECDSA-like signature scheme).

Three stages: registration, root key establishment and message transmission.

Note:

All of Alice's and Bob's message are sent via WhatsApp's servers.

All communication between Alice/Bob and WhatsApp is encrypted/authenticated using a TLS-like protocol.

Alice (and Bob) always deletes a secret key as soon as she no longer needs it.

13.3.1 Registration

1. After Alice has downloaded the WhatsApp app, she sends WhatsApp:
 - ID_A : her identifier (cell phone number)
 - A : her long-term public key
 - U : her medium-term public key
 - $\text{Sign}_A(U)$: her signature on U
 - S_1, S_2, \dots, S_ℓ : one-time public keys (and Alice securely stores her secret keys $a, u, s_1, s_2, \dots, s_\ell$).
2. Similarly, Bob sends WhatsApp: $ID_B, B, V, \text{Sign}_B(V), T_1, T_2, \dots, T_\ell$.

13.3.2 Root Key Establishment

Alice (initiator) wishes to connect with Bob (responder).

1. Alice \rightarrow WhatsApp: request to create session with Bob.
2. WhatsApp \rightarrow Alice: $B, V, \text{Sign}_B(V), T_1$ (and deletes T_1).

3. Alice does the following:

- (a) Verify $(V, \text{Sign}_B(V))$ using B .
- (b) Select an ephemeral key pair (X_1, x_1) .
- (c) Compute root key $\text{root}_0 = \text{KDF}(aV, x_1B, x_1V, x_1T_1)$. (root_0 has bitlength 256 bits).

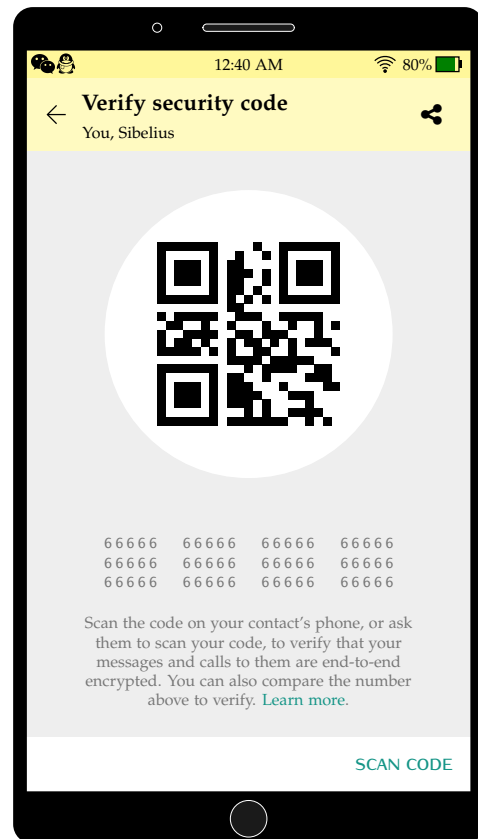
Note that given A and X_1 , Bob can compute $\text{root}_0 = \text{KDF}(vA, bX_1, vX_1, t_1X_1)$. Also note that WhatsApp or any other third party cannot compute any of the four ECDH shared secrets vA, bX_1, vX_1 and t_1X_1 , and thus cannot compute root_0 .

Why use four ECDH shared secrets to derive root_0 ? This is done to give Alice and Bob protection even if some of their secret key material is compromised.

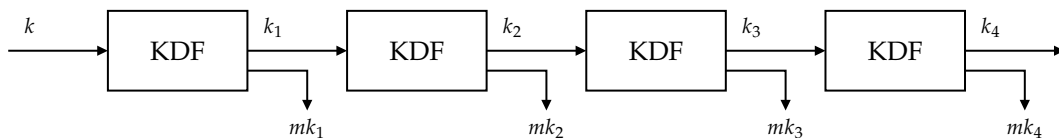
What remains in the root key establishment stage is for Alice and Bob to verify that they have both computed the same root key, and that no one else can compute root_0 . In other words, that Alice and Bob were not the victim of a malicious intruder-in-the-middle attack. Indeed, WhatsApp could launch a malicious intruder-in-the-middle attack by sending to Alice and Bob its own set of public keys and claiming that these are Alice's and Bob's public keys.

How can Alice and Bob verify that they indeed have authentic copies of each other's long-term public keys A and B ? In WhatsApp, the authenticity of long-term public keys is done manually.

QR codes and 60-digit numbers encode identifiers and long-term public keys; (Alice, A) and (Bob, B). Alice and Bob should verify these prior to sending each other messages. This verification is done in order to ensure that WhatsApp is not performing a malicious intruder-in-the-middle attack. However, this verification is rarely done in practice.



Suppose that Alice and Bob share a secret key k . They can *ratchet* k and derive message encryption keys mk_1, mk_2, mk_3, \dots as follows:

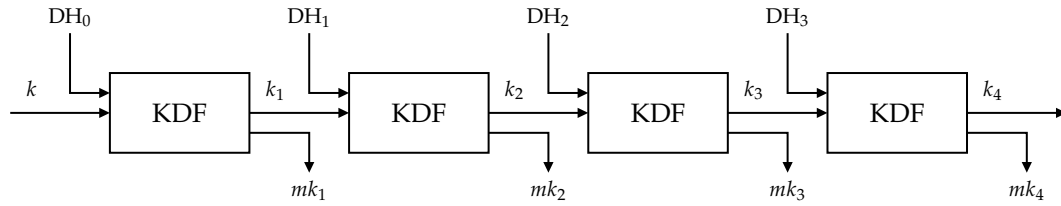


Keys are deleted as soon as they are no longer needed.

For example, k is deleted as soon as k_1 and mk_1 are computed. Also, mk_1 is deleted as soon as it is used to encrypt (or decrypt) a message.

Suppose that E learns k_2 and mk_2 (by gaining access to Alice's device). Then E can compute $k_3, mk_3, k_4, mk_4, \dots$. But E cannot compute mk_1 . Thus, ciphertext that was generated using mk_1 cannot be decrypted by E .

In order to achieve post-compromise security, a fresh ECDH shared secret established by Alice and Bob is used each time the KDF is applied.



Here, $DH_i = \text{ECDH}(X_i, Y_i)$, where X_i is contributed by Alice, and Y_i is contributed by Bob.

Suppose that E learns k_2 and mk_2 . Then E cannot compute k_3, mk_3 unless she also learns x_2 (or y_2).

13.3.3 Message Transmission

Alice maintains three key chains:

1. A **root key chain** (used to seed the other two chains).
2. A **sending key chain** (to generate message sending keys).
3. A **receiving key chain** (to generate message receiving keys).

Bob also maintains three key chains:

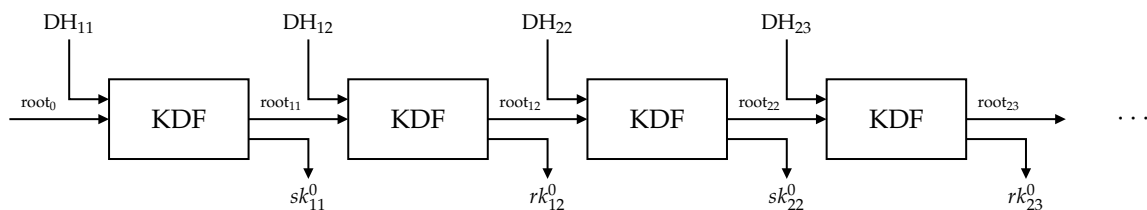
1. A *root key chain* (the same one as Alice's).
2. A *receiving key chain* (the same as Alice's sending chain).
3. A *sending key chain* (the same as Alice's receiving chain).

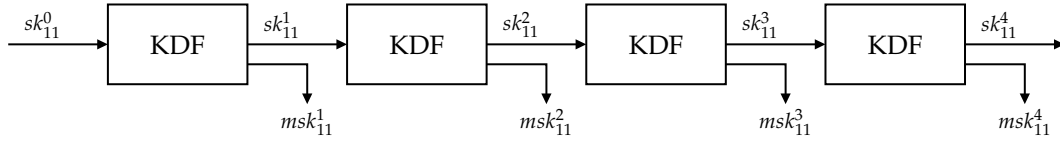
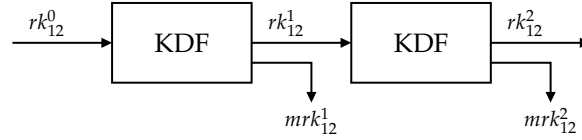
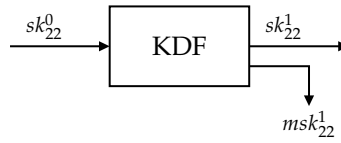
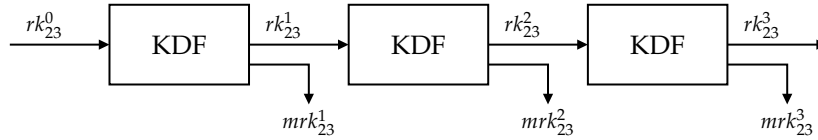
Example of Message Transmission

Consider the following example:

1. Alice \rightarrow Bob: $M_{11}^1, M_{11}^2, M_{11}^3, M_{11}^4$.
(Alice's first sending chain of 4 messages)
2. Alice \rightarrow Bob: M_{12}^1, M_{12}^2 .
(Alice's first sending chain of 2 messages)
3. Alice \rightarrow Bob: M_{22}^1 .
(Alice's first sending chain of 1 messages)
4. Alice \rightarrow Bob: $M_{23}^1, M_{23}^2, M_{23}^3$.
(Alice's first sending chain of 3 messages)

Alice's root key chain



Alice's first sending key chain**Alice's first receiving key chain****Alice's second sending key chain****Alice's second receiving key chain**Alice's root key chain

1. $DH_{11} = \text{ECDH}(X_1, Y_1), Y_1 = V$
2. $\text{KDF}(\text{root}_0, DH_{11}) \rightarrow \text{root}_{11}, sk_{11}^0$
7. Receive Y_2
8. Compute $DH_{12} = \text{ECDH}(X_1, Y_2)$
9. $\text{KDF}(\text{root}_{11}, DH_{12}) \rightarrow \text{root}_{12}, rk_{12}^0$

Alice's sending key chains

3. $\text{KDF}(sk_{11}^0) \rightarrow sk_{11}^1, msk_{11}^1$
4. $\text{KDF}(sk_{11}^1) \rightarrow sk_{11}^2, msk_{11}^2$
5. $\text{KDF}(sk_{11}^2) \rightarrow sk_{11}^3, msk_{11}^3$
6. $\text{KDF}(sk_{11}^3) \rightarrow sk_{11}^4, msk_{11}^4$
15. $\text{KDF}(sk_{22}^0) \rightarrow sk_{22}^1, msk_{22}^1$

Alice's receiving key chains

10. $\text{KDF}(rk_{12}^0) \rightarrow rk_{12}^1, mrk_{12}^1$
11. $\text{KDF}(rk_{12}^1) \rightarrow rk_{12}^2, mrk_{12}^2$
19. $\text{KDF}(rk_{23}^0) \rightarrow rk_{23}^1, mrk_{23}^1$
20. $\text{KDF}(rk_{23}^1) \rightarrow rk_{23}^2, mrk_{23}^2$
21. $\text{KDF}(rk_{23}^2) \rightarrow rk_{23}^3, mrk_{23}^3$

12. Select X_2
13. Compute $DH_{22} = \text{ECDH}(X_2, Y_2)$
14. $\text{KDF}(\text{root}_{12}, DH_{22}) \rightarrow \text{root}_{22}, rk_{22}^0$
16. Receive Y_3
17. Compute $DH_{23} = \text{ECDH}(X_2, Y_3)$
18. $\text{KDF}(\text{root}_{22}, DH_{23}) \rightarrow \text{root}_{23}, rk_{23}^0$

sk	chaining key for sending key chain
rk	chaining key for receiving key chain
msk	message sending key
mrk	message receiving key

Alice sending M_{ii}^j : $C_{ii}^j = \text{AE}_{msk_{ii}^j}((A, B, X_i, j, L_{i-1}), M_{ii}^j)$, where L_{i-1} is the length of Alice's $(i-1)$ th sending chain.

Here

$$\text{AE}_k(T, M) = \text{AES-CBC}_{IV, k_1}(M), \quad \text{HMAC}_{k_2}(\text{AES-CBC}_{IV, k_1}(M), T), \quad T$$

where $k = (IV, k_1, k_2)$ with $IV \in \{0, 1\}^{128}$, $k_1, k_2 \in \{0, 1\}^{256}$.

13.4 References

1. The Double Ratchet Algorithm <http://tinyurl.com/DoubleRatchet>
2. Signal source code <http://tinyurl.com/SignalProtocol>
3. WhatsApp Encryption Overview <http://tinyurl.com/WhatsAppEnc>
4. A formal security analysis of the Signal Messaging Protocol <https://eprint.iacr.org/2016/1013>

Post-Quantum Cryptography

14.1 Quantum Computers

Conceived by Yuri Manin (1980) and Richard Feynman (1981), quantum computers are devices that use quantum-mechanical phenomena such as superposition, interference, and entanglement to perform operations on data.

A qubit is the quantum analogue of a classical bit, and can be in two states at the same time, each with a certain probability.

An n -qubit register can be in 2^n states at the same time, each with a certain probability.

When a function f is applied to an n -qubit register, it is simultaneously evaluated at all 2^n states.

However, when the n -qubit register is measured, it reverts to being in one of the 2^n states according to its underlying probability distribution. So, quantum computers are not “massively parallel machines.”

14.2 The Threat of Quantum Computers

14.2.1 Shor

The public-key systems used in practice are:

- **RSA:** Security is based on the hardness of integer factorization.
- **DL:** Security is based on the hardness of the discrete logarithm problem.
- **ECC:** Security is based on the hardness of the elliptic curve discrete logarithm problem.

Shor's Algorithm: In 1994, Peter Shor discovered a very efficient (polytime) quantum algorithm for solving these three problems.

So, all RSA, DL and ECC implementations can be totally broken by quantum computers.

14.2.2 Grover

Let $F : \{0,1\}^n \rightarrow \{0,1\}$ be a function such that

- F is efficiently computable; and
- $F(x) = 1$ for exactly p inputs $x \in \{0,1\}^n$.

Grover's Algorithm

Grover's Algorithm (1996) is a quantum algorithm for finding an $x \in \{0,1\}^n$ with $F(x) = 1$ in $2^{n/2}/p^{1/2}$ evaluations of F .

Exhaustive key search: Consider AES with an ℓ -bit key. Suppose we have t known plaintext-ciphertext pairs (m_i, c_i) , where t is such that the expected number of false keys is very close to 0.

Define $F : \{0,1\}^\ell \rightarrow \{0,1\}$ by $F(k) = 1$ if $\text{AES}_k(m_i) = c_i$ for all $1 \leq i \leq t$; and $F(k) = 0$ otherwise.

Then Grover's algorithm (with $p = 1$) can find the secret key in $2^{\ell/2}$ operations. Thus, 256-bit AES keys should be used in order to achieve a 128-bit security against quantum attacks.

When will quantum computers be built?

- 1998: (Jones & Mosca) 2-qubit quantum computer
- 2000: 7 qubits
- 2011: 14 qubits
- 2017: 50 qubits (IBM) <http://tinyurl.com/IBMqc50>
- 2019: 53 qubits (Google)

The largest number factored using Shor's algorithm on a quantum computer is $21 = 3 \times 7$.

Note: Larger numbers have been factored on adiabatic quantum computers (e.g., DWAVE computers), but these implementations do not scale.

A quantum computer that can factor a 2048-bit RSA modulus using Shor's algorithm will need (at least) 2048-qubit registers.

These qubits will have to be fault tolerant, i.e., be error resistant.

The physical qubits that have been built so far are not fault tolerant.

So, the plan is to use quantum error correction to combine many (imperfect) physical qubits into one (almost perfect) logical qubit.

Optimistic estimates are that thousands of physical qubits will be needed to build one logical qubit.

So, factoring 2048-bit RSA moduli might need millions of physical qubits.

In Oct 2019, Google announced that they had achieved **quantum supremacy**:

Perform some task (not necessarily useful, but not too contrived) on a quantum computer much faster than is possible on classical computers.

Google's task was to apply a random quantum circuit to the all-0 53-qubit initial state; it did this in 3 minutes.

Google estimated that a classical supercomputer would take 10,000 years to perform the same task.

IBM researchers lowered that estimate to 2.5 days using a different classical technique.

It is important to note that Google's 53-qubit quantum computer is not fault tolerant.

So, while it is major scientific achievement, it does not in any way threaten the security of presently-deployed cryptosystems.

The next major milestone is to build a single logical qubit.

It's still too early to be able to predict when scalable quantum computers will be built. "It is highly unexpected that a quantum computer that can factor 2048-bit RSA numbers will be built within the next decade."

On the other hand, there is no fundamental reason why a large, fault-tolerant quantum computer cannot be built.

What does this mean for internet security? Automatic software updates, TLS. Should we care? The NSA and other organizations are capturing and storing large amounts of internet traffic right now.

What, if anything, should we do to mitigate the threat? When should we take action? Now? In 5 years? In 10 years? In 15 years?

According to the National Security Agency:

Algorithms often require 20 years to be fully deployed on National Security Systems (NSS). NSS equipment is often used for 30 years or more. National security information intelligence value is often 30 years (sometimes more), although it may vary depending on classification, sensitivity, and subject.

Recommendation: Information security systems should be crypto-agile. In other words, they should be designed so that new cryptographic primitives can be employed without significant change to system infrastructure.

NSA's August 2015 Announcement

IAD will initiate a transition to quantum resistant algorithms in the not too distant future. Based on experience in deploying Suite B, we have determined to start planning and communicating early about the upcoming transition to quantum resistant algorithms. Our ultimate goal is to provide cost effective security against a potential quantum computer. We are working with partners across the USG, vendors, and standards bodies to ensure there is a clear plan for getting a new suite of algorithms that are developed in an open and transparent manner that will form the foundation of our next Suite of cryptographic algorithms.

Until this new suite is developed and products are available implementing the quantum resistant suite, we will rely on current algorithms. For those partners and vendors that have not yet made the transition to Suite B elliptic curve algorithms, we recommend not making a significant expenditure to do so at this point but instead to prepare for the upcoming quantum resistant algorithm transition.

See *A riddle wrapped in an enigma*

PQC Standardization

U.S. National Institute of Standards and Technology (NIST)

Solicited proposals for quantum-resistant signature and encryption (key encapsulation) algorithms.

Nov 30, 2017: 69 submissions in Round 1.

Jan 30, 2019: 26 submissions selected for Round 2.

Jul 22, 2020: 7+8 submissions selected for Round 3.

Evaluation and standardization are expected to take 2+ years

14.3 Quantum-Safe Candidates

- (Classical) symmetric-key cryptography
- Hash-based signatures
- Code-based public-key encryption
 - Error-correcting codes
- Lattice-based public-key encryption and signatures
- Multivariate polynomials signatures
- Isogeny-based key agreement
 - SIKE: elliptic curve Supersingular Isogeny-based Key Encapsulation
- Quantum key distribution

Large-scale experiments

1. **Google's ongoing experiment** with quantum-safe cryptography in its Chrome browser: ECDH + Lattice-based key agreement
2. **Cloudflare's ongoing experiment** with quantum-safe cryptography in TLS 1.3: ECDH + Lattice-based key agreement, ECDH + SIKE
3. **Amazon's AWS Key Management Service** now supports hybrid post-quantum TLS.

Commercialization

- ISARA (Waterloo): <https://www.isara.com>
- evolutionQ (Waterloo): <https://www.evolutionq.com>
- Crypto4A (Ottawa): <https://www.crypto4a.com>
- Infosec Global (Toronto): <https://www.infosecglobal.com>
- ID Quantique (Switzerland): <https://www.idquantique.com>
- Post-Quantum (UK): <https://post-quantum.com>
- PQShield (UK): <https://pqshield.com>
- CryptoNext (France): <https://cryptonext-security.com>

Bitcoin

15.1 Paper Cash

Coins (including paper bills) are issued by the Bank of Canada in accordance with an economic policy.

Suppose that Alice wishes to give a coin to Bob (in return for some goods or services).

Bob can examine the coin to ensure that it is valid (i.e., not a forgery).

Double spending is not a concern because Alice cannot give the same (valid) coin to two different parties.

Payer anonymity (during payment) and payment untraceability (after payment) are facilitated.

15.1.1 Features of Paper Cash

- **Recognizable** (as legal tender)
- **Portable** (easily carried)
- **Transferable** (without involvement of the financial network)
- **Divisible** (has the ability to make change)
- **Unforgeable** (difficult to duplicate)
- **Anonymous** (no record of who spent the money)
- **Untraceable** (difficult to keep a record of where money is spent)

Note: Many of these features are not available with credit card payments.

15.2 Bitcoin

An electronic cash scheme invented by Satoshi Nakamoto (a pseudonym) in 2008.

Bitcoin is decentralized, i.e., no central authority such as a “Bank” or “Government” is required.

- Q: Who creates coins?
- Q: How can the creation of coins be regulated?

- Q: How does the recipient of a coin ensure it has not been previously spent?

Anyone can use Bitcoin:

- Download a wallet from <http://bitcoin.org>.
- Obtain bitcoins by “mining” or from an exchange such as Coinbase or kraken.

Note: Payer anonymity and payment untraceability are not primary goals of Bitcoin.

The first bitcoins were generated by Satoshi Nakamoto on Jan 3 2009.

The basic unit of bitcoin currency is 1 BTC.

Each BTC can be divided into 100 million pieces, the smallest of which, i.e., 0.00000001 BTC, is a satoshi.

Bitcoins can be generated (i.e., mined) by anyone.

They are generated at the rate of R BTC every 10 minutes (on average).

Initially, $R = 50$. On Nov 28 2012, R was lowered to 25. On July 9 2016, R was lowered to 12.5. On May 11 2020, R was lowered to 6.25.

R will be halved over time, until the year 2140, when a total of 21 million BTC will have been generated.

As of March 25 2021, around 18.5 million BTC have been generated.

Value of a bitcoin: The US dollar value of 1 BTC has fluctuated wildly:

Date	Value
†May 22 2010	\$0.0025
Jul 17 2010	\$0.08
Jan 1 2011	\$0.30
Feb 9 2011	\$1.00
Jun 8 2011	\$31.91
Jan 1 2013	\$13.30
Apr 9 2013	\$223.10
Jul 6 2013	\$69.31
Nov 30 2013	\$1128.82
Jan 1 2014	\$747.56
Jan 3 2015	\$289.86
Jan 2 2016	\$433.23
Mar 29 2017	\$1183.65
Dec 17 2017	\$19205.11
Apr 2 2018	\$7083.80
Mar 20 2019	\$4087.11
Apr 7 2020	\$7366.26
Mar 25 2021	\$51955

†10,000 BTC for a \$25 [pizza order](#)

Why use bitcoin?

1. It's decentralized.
 - Not under the control of any government.
 - Not under the control of any bank, credit card company, or other financial institution.

- Anyone can use it (even if you don't have a credit history).
 - It's (relatively) easy to use.
2. Bitcoin's monetary policy is fixed and public.
 3. Transactions are irreversible.
 4. Transaction fees are low (even across borders).
 5. Transactions can be anonymous (with a bit of care).

Of course, there are many reasons not to use Bitcoin. We will not dwell on those for now...

The most important contribution of Bitcoin was the implementation of a public decentralized storage system (also known as a distributed ledger) in the form of a blockchain:

A sequence of data organized in blocks with the following properties:

- **public**: readable by everyone.
- **consensus**: all participants agree on the data.
- **liveliness**: writeable by everyone.
- **persistence**: unchangeable by anyone

15.2.1 Elements of Bitcoin

1. **Transaction**: The transferring of a coin from one user to another. All transactions are public and are broadcast to all users.
2. **Peer-to-peer network**: The users of Bitcoin are organized in a peer-to-peer network.
3. **Blocks**: Every 10 minutes or so, the latest transactions are verified and collected in a block. This block is hashed and (cryptographically) linked with previous blocks. The block is broadcast to the peer-to-peer network.
4. **Blockchain**: The list of blocks is called the blockchain. It contains a record of *all* past transactions.
5. **Mining**: The process of verifying transactions and compiling a block is called mining. A successful miner receives a *reward* (new bitcoins).
6. **Proof-of-work**: To successfully compile a block and receive a reward, the miner has to solve a cryptographic challenge.

Main Cryptographic Ingredients

SHA-256 hash function and ECDSA with secp256k1 elliptic curve:

$$E : Y^2 = X^3 + 7 \text{ over } \mathbb{Z}_p,$$

where

$$p = 2^{256} - 2^{32} - 2^9 - 2^8 - 2^7 - 2^6 - 2^4 - 1.$$

$n = \#E(\mathbb{Z}_p)$ is a 256-bit prime. $P \neq \infty$ is a fixed point in $E(\mathbb{Z}_p)$.

Key Pairs (for ECDSA)

Each user selects $a \in_R [1, n - 1]$ and computes the elliptic curve point $A = aP$.

The user's ECDSA private key is a ; the user's ECDSA public key is A .

We will denote Alice's key pair by (a, A) , Bob's key pair by (b, B) , Chris's key pair by (c, C) , etc.

In Bitcoin, a user's public key A serves as a pseudonym for the user Alice.

More generally, a user can select a different key pair for each transaction.

See <http://bitaddress.org> for a JavaScript client-side key pair generator.

The first bitcoins were generated by Satoshi Nakamoto on Jan 3 2009.

Here, S is Satoshi Nakamoto's public key

The genesis block is created by Satoshi Nakamoto (S) on Jan 3 2009. Block 0 is embedded in the Bitcoin software.

15.2.2 Transactions

A transaction is the transfer of a coin from one user to another user.

Suppose that Alice has a coin, say of value 1 BTC.

The transaction in which Alice obtained this bitcoin is represented by the string TXA.

Suppose Alice wishes to give this coin to Bob. The transaction is represented as follows:

$$T_{AB} = \{\widetilde{T_{XA}}, A, B, 1 \text{ BTC}\}_A,$$

where \widetilde{m} denotes the hash of m , and $\{M\}_A$ denotes a message M and its ECDSA signature with respect to the public key A .

This transaction is broadcast to the entire network. Bob (and anyone else) can verify the authenticity of the ECDSA signature using the public key A .

Transaction Chain for the First Bitcoins

Note: The transaction chain contains Alice's and Bob's public keys, but not their names.

Questions: How can the recipient verify that a coin has not been double-spent? (without using a trusted central authority.) How are the other bitcoins generated?

Only one of the transactions T_{AB} , T_{AC} should be accepted as valid; the other transaction should be rejected.

Recall that all transactions are broadcast to all users.

Any user (called a miner) with public key W can volunteer to collect all transactions T_1, T_2, \dots, T_n that it received in an interval of time, say the previous 10 minutes.

The user W verifies that these transactions are valid and that the corresponding coins have not been previously spent.

The user forms a block consisting of the hash of the previous block, the user's public key W , a nonce, and T_1, \dots, T_n .

The nonce is incremented until a hash value that begins with t zeros is obtained.

The block is broadcast to the network.

15.3 The Blockchain

Users will accept a block if all the transactions in it are valid, if the coins have not been previously spent, and if the hash value begins with t zeros.

Users show their acceptance of the block by using its hash as the “previous hash” for the next block, thereby growing the blockchain.

The blockchain serves as a public decentralized ledger that records all transactions.

There is a possibility that two blocks are created around the same time by two different users.

This causes a fork in the blockchain.

To remedy the fork, users will trust the longest chain and continue to grow that chain. More precisely, users will trust the chain that that was most difficult to generate.

The blocks that are not part of the longest chain are dropped and the valid transactions in them are returned to the miners’ memory pool of unverified transactions.

Incentive: The block creator (W) is awarded R BTC (currently, $R = 6.25$) [mining].

Work factor: The target t is updated every 2016 blocks (2 weeks) to ensure that the average time it takes to generate a block is 10 minutes.

Currently, the bitcoin network is generating hashes at the rate of approximately $2^{67.1}$ per second. The hash difficulty is approximately $t = 77$.

A PC can do approximately 2^{22} hashes per second. So, one PC will take about 700,000,000 years to generate one block.

Mining pools: Users form mining pools and share an award.

If A gives B a coin, then B should complete the transaction with A only after the transaction T_{AB} appears in the blockchain, perhaps followed by several more blocks (e.g., six blocks). Transactions are *not* instantaneous. If the transaction is accepted instantaneously, B has to accept the risk that A might double spend the coin.

Since all transactions are public, payer anonymity and payment untraceability are not guaranteed.

Bitcoin is “secure” as long as honest users collectively control more CPU power than any cooperating group of users.

Bitcoins can be combined and split.

Suppose that Alice (A): received 25 BTC from Bob (B) in Transaction TBA received 20 BTC from Chris (C) in Transaction TCA.

Suppose that Alice wishes to: give 30 BTC to David (D) leave 14 BTC to herself as change give 1 BTC as a transaction fee.

Here is the corresponding transaction:

The transaction fee is claimed by the miner who validates this transaction.

Suppose that Bob owns two public keys, B and H . Suppose that Bob received 9 BTC in transactions T_1 and T_2 :

T_1 : 1 BTC from A to B
 3 BTC from D to E

2 BTC from C to H

T_2 : 5 BTC from C to F

6 BTC from G to B

Suppose Bob wishes to give 2.5 BTC to F , 3 BTC to I , 1.5 BTC as change to B , 1.75 BTC as change to H , and offer a transaction fee of 0.25 BTC.

He forms the transaction T_3 :

Inputs: $(\tilde{T}_1, 1), (\tilde{T}_1, 3), (\tilde{T}_2, 2)$

Outputs: $(F, 2.5), (I, 3.0), (B, 1.5), (H, 1.75)$

T_3 has 3 signatures, with public keys B, H, B .

If an attacker obtains a copy of Alice's wallet (and thus her private key a), then the attacker can spend the coins associated to A . If Alice deletes (or loses) her private key a , then all her coins corresponding to A are lost forever.

Alice could store her bitcoins at an exchange.

Mt. Gox, a Bitcoin exchange based in Tokyo, "lost" 850,000 BTC and declared bankruptcy in February 2014. Later, it "found" 200,000 BTC.

In early December 2018, the founder of Canadian cryptocurrency exchange QuadrigaCX died. Apparently he was the only person who knew the password to access \$190 million worth of QuadrigaCX's customer cryptocurrencies (including bitcoin). See <http://tinyurl.com/WPQuadrigaCX>

Alice could use a hardware wallet (e.g. Ledger) or a cosigner service (e.g. BitGo).

Scalability: The maximum size of a block limits the number of transaction to about 7 per second. In contrast, VISA processes about 2,000 transactions per second.

Mining costs: Mining requires hardware and electricity.

Energy costs per transaction: \approx Cdn \$25.

Proof-of-stake:

Miners put up collateral (stake) and are chosen based on the size of their stake.

Misbehaving miners are penalized (e.g. by slashing their stake).

Examples: Algorand, Cardano, Tendermint, EOS, Casper.

Several technical details have been omitted including:

A public key is identified by its 160-bit hash value.

Merkle trees are used to minimize the size of a block.

Simplified payment verification (SPV).

Today, Bitcoin is rarely used for (legal) transactional purposes. It is mostly used:

- illegal transactions (moving funds out of a country, illegal purchases on the internet, ransomware, etc.);
- as a mechanism for storing value in countries that are experiencing hyperinflation (easier to store and use than gold or diamonds);

- as a (highly) speculative investment.

The future of Bitcoin is uncertain, and many challenges remain: technological, economic, government regulations.

Nonetheless, Bitcoin has been a remarkably successful proof-of-concept, and there is now a very high rate of innovation in the cryptocurrency/blockchain space.

- Bitcoin magazine: <http://bitcoinmagazine.com>
- Download a wallet: <http://bitcoin.org>
- Live blockchain: <http://blockchain.info>
- Bitcoin Blocks: <http://explorer.btc.com>
- Genesis Block: <http://tinyurl.com/BTCBlock0>
- Block 1: <http://tinyurl.com/BTCBlock1>
- Block 100,000: <http://tinyurl.com/BTCBlock100000>

15.4 Ethereum

Invented by Vitalik Buterin in 2013. <http://tinyurl.com/NYTButerin>

Blockchain-based decentralized computing platform.

Underlying cryptocurrency is called ether.

Supports a Turing-complete programming language.

Permits full smart contract functionality.

Potential Decentralized Applications (DAPPs) include:

- Timestamping and notarization of documents.
- Record asset ownership (domain names, stocks, student transcripts, home ownership, etc.).
- Contract signing (without a “trusted” lawyer).
- Crowdfunding (a.k.a. Kickstarter).
- Democratic autonomous organizations (DAOs).

Enterprise Ethereum Alliance (EEA) members include: Microsoft, Intel, Cisco, J.P.Morgan, Credit Suisse, PriceWaterhouseCoopers, Ernst & Young

Feb 28, 2017

Together, we will learn from and build upon the only smart contract supporting blockchain currently running in real-world production – Ethereum – to define enterprise-grade software capable of handling the most complex, highly demanding applications at the speed of business.

<https://entethalliance.org/>

- **Zcash**: Anonymous cryptocurrency.
- **Algorand**: Proof-of-stake blockchain platform.
- **Diem**: Facebook’s proposed cryptocurrency.

- **Filecoin:** Cryptocurrency-incentivized decentralized and verifiable file storage network.

Wrap-up

16.1 Boring Crypto

This course has been about BORING CRYPTO:

Cryptography that is well studied, widely standardized, and widely deployed. MAIN PROBLEM: Alice and Bob wish to communicate securely.

They encrypt/authenticate data using: AES-GCM or ChaCha20/Poly1305.

They establish a shared secret key using: ECDH (or RSA key transport).

The public keys need to be authenticated:

- Certificates issued by a CA. (e.g. TLS).
- Preinstallation of a public key (e.g. in an operating system).
- Visual inspecting each other's public keys (e.g. Bluetooth, WhatsApp).

So is Cryptography a Solved Problem? OF COURSE NOT!

A lot of challenges remain:

- The possibility of cryptanalytic advances
- The threat of quantum computers
- Efficient implementation
- Secure implementation
- Key management
- The insecurity of the Internet of Things (IoT)
- Balancing the privacy rights of individuals and the needs of law enforcement

16.2 Cool Crypto

- Lightweight cryptography
- Quantum-safe cryptography

- Lattice-based cryptography
 - Isogeny-based cryptography
- Cryptocurrencies
- Differential privacy (e.g. Google's RAPPOR)
- Multi-party computation
- Computing with encrypted data
 - Fully homomorphic encryption
 - Encrypted databases
 - Privacy-preserving machine learning

Index

C

collision-resistant hash function 43
confidentiality 8

D

data integrity 8
data origin authentication 8

E

elliptic curve E over F 91
exponential-time algorithm 85

F

F -rational points on E 91

G

generic attack 46

H

hash function 41

I

input size 77

K

Key management 109

M

message authentication code 55

N

non-repudiation 8

P

point multiplication 94
polynomial-time 77
polynomial-time algorithm 85
preimage resistance hash function 42
public-key cryptography 73

R

random bit generator 116
running time 77

S

second-preimage resistance hash function .. 43
secure MAC scheme 56
secure public-key encryption scheme 84
security level of a cryptographic scheme ... 14
security of a signature scheme 88
security of SKES 13
subexponential-time algorithm 85
symmetric-key encryption scheme 11