GLOSSARY

Glossary of Information & Intelligence

Note: this is about information, intelligence, security it, and measurements on these. Communicating it is not here, especially the portions of manipulation of media to convey it. Nor are generic reliability concepts.

access control The element of *identity management* that regulates access to a resource. Determines what a user (subject) can do and can't do with a resource. Determines if system is in an uncorrupted state. May be contingent on time of day, location, channel of access, so on

see also *authorization, biometrics, capabilities*

access level

1. Private
2. Controlled
3. Restricted
4. Confidential
5. Public

assign level of access and trust Too many parties assigning trust may allow too high of access. Social flaws: if you don't give permission 'you've wronged that person.' Too few assigning parties has weaknesses too. Restricting 'certificates' can create an artificial scarcity leading to inflated prices and poor organizational administration.

attribute based e.g. 18 and over, level of trust in the identity

changes Changing restrictions is operating with authorization for a known secure path only.

discretionary access control Defines how something (subject) may have access privileges to an entity (object).

grants Rights granted based on (see *security evidence*)

- Information gathered
- Listed and path caps on access
- Limits to knowledge and proof
- Security policy

mandatory access control The computer maintains access restrictions in all cases; the intention is to prevent accidents and compromises by the user. This ensures that anything (that contains elements of the restricted information) is kept at least at the same level of restricted access as the original information.

matrix Access control matrix

***Table 1:** Access control matrix example*

Object	*Operation*	*Users*
1	operation 1 operation 2 operation 3	set of users
2	operation 1 operation 2	

FILE G:\My Documents\Glossary\Glossary InfoSec;7.doc

	operation 3
3	operation 1 operation 2 operation 3

privileges

Privileges (or permissions) typically include at least read, write, execute, list permission, and change permission.

time and region limits

Information can be limited with a marked Geographical Zone distribution:

- Current zone
- Host
- Local network
- External network

Anything that employs it inherits restrictions on zone access as well. This reduces accidental release of information outside of the zone.

accountability

ADVISE
analysis, dissemination, visualization, insight and semantic enhancement

US government program.

See *TANGRAM*

analysis

Logical accounting for all individual parts of a main body of data.

Know the functions and limits of the individual parts

missing parts

nonassociated parts

1. Study data in its original form, look for obvious or significant parts

2. Look for signs of unusual conditions

aircraft tracking

Aircraft info: Altitudes, numbers and type

Intercept track

Call sign, and frequency to station location for azimuth-range position reports

recognize ELFAIR-relevant traffic,

recognize irrelevant chatter

automotive tracking

Virginia’s Department of Motor Vehicles: 1997 Operating costs were in part, paid by selling lists of data.

Ameritech’s GovServices

Civilink: Computer system to maintain public records, usually court records

Ameritech is also often hired to perform those government operations

approaches

- Evidence-based security: information gathered about the code
- Code-access security: examines calling chain to cap access; Can annotate code with security limits and requirements.
- Defined process to verify assembly – type safety, operation use, modifications
- Role-based security: security ids and attributes based on a role, title, task

- Cryptography
- Separate application Domains

assesses

performance

capabilities

vulnerabilities

what it is doing at any given time

attack surface *cliché* don't use.

attribute May be static characteristics or dynamic properties (functions or behaviours). Documentation often includes:

- Definition of attributes
- Explanations for measurements of attributes
- Guidance as to data sources, quality and data
- Explanations for missing values or inappropriateness of a measure
- Any other information to provide clear explanations of the data

Predictor attribute; dependent attribute

See also *property*

auditing

audit trail Values of the relevant shared variables are recorded prior to an update

authentication A process of *identity management* that maps an unknown user (etc) to an identity (approximately a known person). May use credentials.

This is, or is similar to, certification. It may have an expiration time.

see *biometrics, credentials, password*

level of Increasing security level of unknown person. More destructive or dangerous operations require a higher level of authentication.

May require more than just the identification of the person; it may require certification that the person has character.

of message Properly identifies sender. Message digest of the transmission using A's private key.

A encodes a known message using its private key.

authorization The element of *access control* that regulates access to a resource based on the accessor (subject); ensures that contents/commands are done only by authorized parties

authorization check Can I do $A_1.A_n$ accesses on $D_1..D_n$?

1. Can anyone do these?
2. Do I have a set of keys that allow those access?
3. If I am authenticated, retrieve my level of access information. Do I have sufficient level of access for an item of this type? (Each item has a level of classification code)
4. Am I (or my authentication Id) on a list of people allowed to A1..An for this? (ACL groups many objects can share such a list)
5. If classified at a higher level than you, an intermediary can send you a (possibly redacted) portion, or answer some of your're queries about it.

biometrics "The literature defines biometrics as distinguishable (rather than unique) physiological and behavioral traits that may be used for identification and authentication." "Most biometric applications rely on an assessment of similarity between stored templates created at enrollment and biometric samples taken during user authentication. The matching process suffers from imprecise standards (such as the rigidity of thresholds to define performance precisions) for the measurement of similarity. Stricter (or more relaxed) matching requirements result in higher rates of false rejections (or false acceptance)."

Chandra, Akhilesh; Thomas Calderon. "Challenges and constraints to the diffusion of biometrics in information systems" Communications of the ACM, *December 2005*

see also *authentication, credentials*

block cipher system Groups of plain text are treated as large binary integers (200+ digits). This number is transformed by an encryption function to create another large integer (the cipher).

botnet A system created when several home & office computers (up to tens of thousands) are illicitly infected with a virus or worm known as a `bot'. This system may allow the bots to be commanded to do some limited tasks, such

http://www.lurhq.com/phatbot.html

as:

- send out spam,
- send out other bots,
- attack a site or computer network (usually in a massive denial of service attack),
- install or manage key-logging (and other `spy' software)
- execute, copy, modify, upload or download files.
- open the system for further attacks.

The bots may be controlled thru custom network software, or they may connect to popular IRC (internet relay chat) or other P2P systems (include. Gnutella, WASTE, etc.) and await commands. It is reported that botnets controlled by eastern European hackers are available rented for as little as $20. Agobot is the most common bot family; its source code is widely available.

break

A method of generating a (fake) message with a specified signature, with less complexity than a pure brute force attack. Most breaks are not security risks – they are modified brute force attacks, and have a complexity that makes them impossible to employ.

See also *dictionary attack, rainbow table*

broadcast flag

To prevent signal theft and redistribution of materials. There is a great deal of heated debate, mainly from parties with poor legal background or a liberals arts background. The devices for 5C compliance internally are not required per se, merely to prohibit recording, just redistribution.

CALEA
communications assistance for law enforcement act

1994. Major telecommunications carriers quietly cooperate to prevent the losses publicity would cause if they required (or fought) warrants. CALEA made it physically easier (rather than remove q's on warrants)

call sign

a basic call sign may be assigned to a function

supplemented with geographic reference, or number suffix

capabilities
StarOS style

Modify (t/f) Can modify object property

Destroy (t/f) Object can be removed from system

Copy (t/f) If the object & capabilities can be copied

Restrict (t/f) If the other capabilities can be removed

Pointer to item.

Type:

- NULL: Slot is empty
- Representation: Pointer is to a representation object
- Abstract: Pointer points to an abstract object
- Token Type: Pointer is a special extended privilege for the list holder

Capabilities list: configuration data on what operations A can perform on other objects. F(object, caps) → set of allowed selectors.

method

Complex expression is rewritten into a 'normal form' (B). Usually this is wrt the operations. Reduces the variation to a more manageable amount. The normal form is translated into the target set.

functions

Funds transfer, licensing format, security, spam control / denial of services, privacy control, certificates, door badges, access smart card, IO communication, messaging application

Movie & music redistribution & control, buying music, licensing, document change control, smart cards (building access), keyless entry (start, door, trunk)

Broadcast systems, smart card with satellite tv, broadcast flags, DVDs.

Distributed resources: leases, GC exit state (trumps all others), phone system , access control.

Computer: execution privileges, device channel (to get boot access, reduce egg-shell security issues with physical access). Databases, IO model (switch/fabric has this) NT security tokens in API's.

capability search model

Tokens in feedback: tokens relations with cash (money), resource scarcity, options & tokens

Spam control: token is used to indicate email was paid for / should be sorted / or processed in some way. Each machine / sender has a credit account; when it runs of credit, no one will forward any more messages. This effects a feedback, slowing down trojan's / zombie machines. Alternatively, billing occurs only when tokens are redeemed; people only redeem those when recipient rejects them. Thos doesn't seem as fast to sop spam & virii.

Palladium / Nexus: tokens are used by special processor to hand-off control of main processor, one thus happens, usually the Os in charge.

certificates

A type of credential, where an Agency says file/service conforms to level XYZ.

Some Certifications – like that this person is trustworthy – should be time limited (the person may change). Others – like that a person was exposed to such and such training – should not be,

Restricting certificates can create an artificial scarcity, leading to inflated prices and poor organizational administration.

CIA

Directorate: if subject doesn't answer, this is considered evidence that they are guilty (e.g. had special training for evasiveness) and should be promoted to enhanced interrogation.

classification

Extract features and classify. Categories, grouping, classification hierarchy, taxonomy. A document may cover multiple topics. Constructing a taxonomy. Identifying relationships between documents.

codes of standard terminology

Consist of a standard name for a procedure, a number or other identifier assigned to represent it, and a brief description of the procedure.

Organized into logical groups to aid finding the appreciate code quickly, with easy. To provide a uniform language that accurately describes procedures & services – a means of reliable communication over many different kinds of people (e.g. a nation). Compilations are copyrightable if creativity in their selection or their arrangement. [Argued that if the effort produces a useful work, it should not be copyrightable]

codewords

Shared between US, Canada, UK, Aus, NZ; rules:

1. 5-letters, 2. pronounceable, 3. infrequently used words, 4 not derogatory or profane, 5 do not reveal the meaning

see also *cover term*

bookbreaking

Codewords / phrases related to secrecy?

compartmentalization

A document is in at least one compartment and labeled with a codeword relative to the department. A compartment is a special control structure. Compartments have unique codewords, used to protect specific information. And a channel.

confidentiality

Hiding part or all from eavesdroppers

containment

paged or segmented memory ; files; vdisk; user group

Corona

Spy satellites in use from 1959-1972. The satellites ejected film canisters, which were retrieved by C130 jets. The highest orbit was 12500 feet. Cloudy skies made for a nifty horizon.

cosine measure

Measures the angle between normalized vectors (is this term as valid covariance?). This is most often used to estimate text similarity, when categorizing texts.

see also *naïve Bayesian*

costs
incremental

The cost of keeping up with CERTS is very low on OpenVMS, since OpenVMS is the subject of very few CERTS vulnerability alerts.

http://h71000.www7.hp.com/openvms/whitepapers/TCS_2004.pdf

cover term[1]

These are characterized by

- The type of cover term
- The definition of its intended use, and what that signifies
- The authority (if any) in providing or assigning the term
- The length of the term
- The alphanumeric rules
- The security classification rules; the term alone and it's meaning
- Exclusions: words that can't be employed, such as those on other lists

credentials

Used in *authentication* to establish that an identity is warranted. Based on:

- Who you are: uses authenticated identifier
- What you've got: token or key
- What you know: a password or secret

The credentials may have an associated trust level.

see also *identification*

crime
motives

It is commonly asserted that profit is the motive. Others disagree: "few computer criminals are indeed motivated purely by profit. Employees become criminals during employment to solve personal problems that may involve money, sabotage, or espionage. They are often motivated by debt, relationships gone bad with other employees or spouses, personal dissatisfaction, or an attempt to hide poor or unethical business decisions."

Parker Donn B; Letter to Communications of the ACM*, June 2008, cites* Crime by Computer *(1976, Scribners), and* Fighting Computer Crime *(1983, Scribners)*

Curveball

defector to Germany, made it all up. Based on what interrogators asked, would research and extrapolate new "info" for them.

database

Regular structure for information; as a database's scope expands, relational tables approach matrices.

data
classes

Based on availability, performance, cost

- Business critical
- Business operations
- Business internal data
- Departmental & Remote office data
- Reference data
- Archive data

[1] V2N4

dictionary

Defines the attributes in the model, including their type and value range.

See also *attribute, model*

data mining

"Goals common to all… the detection, interpretation and prediction of qualitative or quantitative patterns in [existing databases]"

see also *data quality, modeling methods*

Ramakrishnan, Naren; Ananth Grma, "Data Mining: From Serendipity to Science" Computer, *August 1999, p34-37*

core techniques

Clustering, Classification, Association, Time-Series Analysis, Trend & Forecasting

issues

- Hard to get the proper source data
- Unrealistic expectations about the quality of data
- Plan and et al inflexible & inappropriate
- Usually the data has enough data errors, inaccuracies and inconsistencies to be of concern

scope

Enterprise-wide: Data-warehouse
Business-unit specific: Data-mart
Subject-specific: Data-mart

industries

Industries that employ data mining: insurance, direct mail
IBM Advanced Targeted Marking for Single Events. Business Intelligence @ Fingerhut, Inc. Used IBM Probabilistic Estimations data-mining

IBM Underwriting Profitability Analysis, for Farmers Insurance Group.

MCI 'Friends & Family' (1991) relative small sub-graphs in long distance call graph → calling circles. Sought to make these calling circles regular customers

Minimum Length of time a telecomm is required to maintain call records (US Federal Law) – 2 years.

data normalization
rules of

Attempts to reduce redundancy in a database, and provide useful constraints by striving to define the data as a set of relations in which all of the attributes are functionally dependent on only the primary key. Each step of normalizing produces more tables than in higher-precision forms, with fewer columns per table. The rules are:

1. Eliminate repeating groups
2. Eliminate redundant data
3. Eliminate columns not dependent on key
4. Isolate independent multiple relationships
5. Isolate semantically related multiple relationships

Note: structuring a model purely based on the normalization rules will yield a poor model. The complexity of the work rapidly rises with the increase of tables, usually by an order of magnitude. I.e., normalizing table A (lets say O(n) access time) splitting it into two tables, would produce an access time of $O(n^2)$.

data notation

Does the data belong to one information category or many?

What is the coding system?

- numeric
- alphanumeric
- alphabetic
- mixed
- mnemonic

extending an existing standard to meet the needs

differentiation from related data

data quality

Data often has enough inaccuracies, errors, and inconsistency to be of concern:

- Assessing data quality
- Adjusting processes
- Cleansing to resolve problems
- Techniques to make inferences and analysis robustly in the face of this
- Control matrices
- Sampling

data taxonomy

What is the range of data produced and consumed?

What is the best way to *store* the data, *retrieve* it, *manipulate* it, and *serve* it? Is it transactional? Atomic? What are the consistency requirements? Integrity? How do you do this reasonably efficiently and scalably? How redundant are you expecting it to be? Service levels?

delegation

The element of *identity management* that allows one to perform actions on the behalf of another. It forms an algebra of operation. The basic operations look like

> A speaks for B
> A says Y

Example rules:

> A says (B speaks for A) then it is believed; other's may not be believed. (B speaks for A) is true
> A says (A says C) then same as (A says C)
> (A speaks for B) and A says C then B says C

There is an ordering of security level

> T(x) x is a type
> $T_{\text{security level}}(x)$ there is a type, secret-x (not quite the same as type x)

DES
data encryption standard

Encryption technique developed in the early 1970's by IBM, with NSA feedback. (Most assertions that the NSA weakened DES are inaccurate; their feedback was crucial to strengthening it). Uses either 40-bit or 56-bit keys (the later, and any other encryption using 56-bit or longer keys, required an export license until 2000). The standard was adopted in 1976 as a federal standard for transmitting both industry data and classified federal data. The legal regulation dates from the 1940s; until the early 1990s strong encryption required special equipment. (The use of dedicated computers was an impractical possibility). Since the early 90's encryption time has fallen, due to computers capabilities; this is one reason n that longer keys have been required to ensure strength.

DES was replaced with AES in 2000 as the recommended standard.

dictionary attack

Stores a bit table correlating a dictionary of words (and other inputs) to a digest/hash output. This allows mapping an encrypted password to the password (or message).

See also *rainbow table*

digital signatures

Were first tried in the 70s but failed as cumbersome and expensive; study of the algorithms has continued.

Domain and type enforcement

Partitions system resources. Processes have an associated domain; objects have an associated type. Two global tables defined allowed access: Domain-definition table, domain interaction table.

doppelganger

Jon Orwant

Events. Succession Events: Fields include: Organization, Post, In&Out (links to below), VacancyReason, Comment

In&Out. fields include: Person, NewStatus, OnTheJob, OtherOrg, Comment

Organization. Fields include Name, Alias, Descriptor, Type, Name

Person. Fields include:

enforcement Concentrated in a small part of the system. Concentrates assurance effort on that part.

Difficulties:

1. Multilevel security
2. Techniques are very similar
3. Attitudes toward what's already there.

Three approaches:

1. Live with it, but patch it
2. Replace it
3. Extend it

Notes. All systems are replace eventually, the rate varies and the strategy of replacements. Identify common mistakes and better approaches. Records are important.

Practical, Messy: Complicated, or add-ons: virus checkers, firewalls

Practical, Sound: Replicate low security to high security system

Impractical, Sound: Orange Book (Trusted Computer Systems Evaluation Criteria), Red book, requires too much change in products. Breaks legacy systems.

electronic intelligence
ELINT

operational

technical

collection manager

analyst

entropy
conditional

Entropy a data source has and that an adversary might have knowledge about.

unconditional Entropy ignoring such knowledge

espionage "Espionage is defined as the practice of spying on or spying by governmental and military entities to gain information."

Executive Order 13292 Applies FIPS 199 military plans, weapons systems, operations, foreign government information, intelligence activities (including special activities), intelligence sources or methods, and foreign relations.

fact estimation noisy-or model to estimate facts from how often it is repeated.

Simple form. The probability is the extraction rule's accuracy. Not very good, ignores sample size:

P(x is correct | statement X appears k times) = $1-(1-p)^k$

Better:

P(x is correct | statement X appears k times in n documents) =

$$\frac{\sum_{r\in num(c)}\left(\frac{r}{s}\right)^{k}\left(1-\frac{r}{s}\right)^{n-k}}{\sum_{r\in num(c-e)}\left(\frac{r}{s}\right)^{k}\left(1-\frac{r}{s}\right)^{n-k}}$$

c = correct statements

e = erroroneous apps

Redundancy helps cope with noise (Information by redundancy), correlation of correct relation much higher than correlation of errors.

P(conclusion), P(conclusion|given) can be a probability distribution. Use Poisson approximation

finger print

A hash of an encryption key

FIPS 199

Protocols for determining confidentiality levels, data integrity levels, and availability level.

Created in Feb 2004; mandated by E-Government Act of 2002 Title |||, "The Federal Information Security Management Act."

firewall

"barriers between us and them for arbitrary values of *them*" "A point of logical flow of a program where the validity of logical constraints is checked. If the constraint is satisfied, execution proceeds. If the constraint is violated, the firewall triggers and takes appropriate action."

FISA
foreign intelligence surveillance act

1978. Established special, secret courts (without oversight) to allow spying on US citizens, especially by the NSA. The NSA was required to show probable cause, and typically looked at 10-20 people at a time. The NSA always thought of it as a hindrance to their moral superiority and desired to collect info on larger numbers of people.

In 2001, changed with little to no oversight. Standard was lowered to "reasonable belief." NSA typically monitors 5000 people at a time.

grouping items

Find sets of items in a transactional database purchased together so as to "improve the placement of items in a store or layout of mail-order catalog pages and web pages."

Item set – a "set of item's that appear together in many transactions"

Item set's support – "the percentage of transactions that contain an item set"

Market basket – "a collection of items purchased by a customer in an individual customer transaction."

Process goes thru transactions building a pool of best item sets and discarding weaker ones. With databases too large to go thru, randomly selects regions.

Is-a hierarchies to generate abstract item sets.

Ganti et al, "Mining Very Large Databases" Computer, April 1997, p38-45

heap spraying

Lots of exploit code placed throughout memory in an attempt to overlap with a hole (buffer overflow).

HIPAA
health insurance portability and accountability act of 1996

aka, Kennedy-Kassenbaum Act, Public Law 104-191.

If the healthcare providers choose to use electronic transactions, specifies the mandatory standards for electronic transactions (in health insurance). Typical claims, prior, cost $6-$8 to process. 2001 estimate was that an electronic claim cost $0.17 to process. A number of packages and organization process have come into being (and are still maturing) to handle the issues identified in HIPAA, as well as the standards.

http://www.cms.hhs.gov/hipaa/

ANSI X12.837
standardized data exchange

Standardized data exchange, security and privacy standards (specified in ANSI X12.837). Six areas of standardization are:

1. transactions – claims, enrollment, eligibility, payments, referrals
2. code sets – identifying diseases, procedures, equipment, drugs, transportation, ethnicity
3. identifiers – to identify providers, payers, patients, etc.
4. benefit coordination when more than one health plan is involved
5. security of information and privacy
6. electronic medical records

Rolled out in stages:
Oct 2002. Healthcare organizations compliance with transaction and data set standards
Apr 2003. Compliance with privacy standards

Format: http://www.com1software.com/c1088.htm
Dictionary of elements in the data exchange: http://www.ihs.gov/AdminMngrResources/HIPAA/docs/ded4010.pdf

Unkown. Security, identifiers, etc

Security Guidelines: Requires that an organization implement controls for accessing records, auditing, and authorization tasks; data and entity authentication. It is up to the organization to define these, “maintains or transmits health information shall maintain reasonable and appropriate administrative, technical and physical safeguards.”

identification Do not confuse with *authentication*

1. A persons signature (e.g. drawer or endorse on a check)
2. The sending banks password (the issuer key) in an ETF
3. Personal Identification Number
4. Biometric identification
5. Personal information known to both account holder and the other party (usually written on a card)
6. SSN, Taxpayer Identification Number
7. Bank Identification Number
8. Machine Readable code on the bottom edge of cheques
9. CUSIP number

see also *credential*

identifiers Start with an uncommon letter, e.g. Q or K, to make it easy to find on a page or display.

Use checksum on identifier to catch typos

Print as groups of 3 or 4 characters.

Eliminate pairs of letters that sound the same on a phone.

identity based on Pointer to memory region.

“state” the memory it points to. Require everyone to create a sensible definition.

identity management authentication, authorization, roles, delegation, interchange of identify information between domains.

informatics The development of techniques and systems for more efficient organization, storage and dissemination of recorded scientific information.

information

Type of information	*Technique*
structured information	relational DB, etc.
population information	statistics
unstructured information	keyword tables

Table 2: *How structured is the information we have and how do we work with that?*

information cost When it is too inconvenient to get better information, people will use the information they already have.

Mooers, 1969

If the people who (have to) use a tool/technique/system do not benefit from it, it may fail, be worked around, undermined, or be subverted by the workers.

Jonathan Grudin

information diffusion Information diffuses out: it moves to more systems, it becomes inaccurate, less precise and/or outdate. It gets public.

information extraction Filling fields and records of a database from text or loosely structured inputs.

Segmentation – find the start & end boundary of the text snippet

CLASSIFICATION – which database field to use?

ASSOCIATION – which fields belong in the same record

NORMALIZATION – putting the information into a standard format

Identifying names, places, organizations, phone #'s, dates, times, internet address, currencies.

Identify key points using part of speech estimation and sense disambiguation to identify interesting phrases.

information flow
lattice model of

Security applications (secure information flow)

Denning

Lattice based access control restriction; lattice defines level of security an object may have and access to

Labels attached to objects

Labels attached to data (contained by objects)

See also *access control*

policy

A policy of a set of allowed accessors:

owner$_1$: accessor$_1$, .. , accessor$_n$
owner$_2$: accessor$_m$, … accessor$_p$

information lifecycle management

Information Lifecycle Management requires more finely detailed information than most file systems provide.

- Tiered storage (keeping some files fast, and archiving others) requires fine attribute details and policies
- Requires atleast application support to be successful
- Applications to create (atleast) the data that couldn't otherwise be deduced or inferred
- Summary tables created by applications or extraction tools
- Backup SW and HSM SW that uses this information to best stored the data cheaply or for fast access, etc.

criteria

- Relevance
- Comprehensiveness
- Freshness
- Presentation

information market

Markets that exist to aggregate information. They might "bet" on the likelihood of an event (to form the group estimate the likelihood), or sell information such as measured facts.

- Coordination Systems
- Information Sharing
- Sale of Reynolds numbers and lab work

information practices
code of fair information practices

Five principles

There must be no personal data record-keeping systems whose very existence is secret

There must be a way for a person to find out what information about the person is in a record and how it is used

There must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the persons consent.

There must be a way for a person to correct or amend a record of identifiable information about the person

Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data.

Department of Health, Education, and Welfare, 1973; simson 2000

Canadian Principles

1. Accountability
2. Identifying purposes
3. Consent
4. Limiting collection
5. Limiting use, disclosure and retention
6. Accuracy, adequate for purpose
7. Safeguards
8. Openness
9. Individual access
10. Challenging compliance

information processing

Needs to protect managers ”from irrelevant distractions of their attention.”

Simon p144

“allocate the time [people] have available for receiving information [in such a way] that they will get only the information that is most important and relevant to the decisions that they will make”

Context input processing product model

Evaluation is to be conducted:

Swanson, p79

1. Needs analysis
2. “In an appraisal of strengths and weaknesses of alternatives
3. “During implementation
4. “In a comparison of the implementation with expectations.”

Information retrieval without world/situation

Look person by name: direct, relational

look person up by property: relational query

person property: dereference, relational query

approximate methods

fuzzy methods

bayesian

vector space

information security

At least partly concerned with users, actions, objects, and media/channel – and combinations thereof. Is the action possible? Can the user carry out the action? Can the user access on the object? Can this user perform this action on the object? And so forth.

1. Confidentiality – must also be able to handle non-confidential data
2. Information Integrity – validity, completeness, and precision of information
3. Possession
4. Information Authenticity
5. Availability – data must be in a useful form
6. Utility
7. Validity
8. Completeness
9. Precision
10. Possession

information security policy

An information security policy should address: Systems security, product security, community security, and corporate information security. It should include managerial oversight: periodic checks, establish clear policies, self-audit and control functions in each department. For items to be protect it should identify:

- Categories of users; the object would have a subset of these that are permitted to access it

- An extent to which access if provided
- Various levels of access (that is, what may be accessed) and some guidelines to help decide which level of access each item should have
- Relevant time restrictions, periods and duration of use

Other issues

- Propagation of privileges
- Restriction
- Revokation
- Anonymous – broadcast keys, public keys, capability sets, public/world access privileges

multi-level security

Objects, Users, Communication Channels have a security privilege (or permission) level assigned to them; classically this is: classified, secret, top secret, Pink Toto. Access and operations compare the security levels. This is a fast operation on a per access basis.

- A > B is true if A is more secure than B.
- A User can read Obj if his security level is ≥ Obj's security level
- A User can create Obj if it's security level is ≥ his security level
- A User can post to a CommChannel if his security level is ≤ the CommChannel's Security Level.

Other methods include Access Control Lists, Capability sets, Group/User/World permissions. This system is imperfect and his tends to create a some characteristic problems

- Most objects are classified at a higher security level than intrinsically necessary
- People in turn are promoted to higher levels of security just to gain access to the objects, etc.

multiple routes of access

There is often more than one means of accessing a given resource -- and may not be adequately protected

information theory

A signal, chosen form a specified class, is applied to a channel, and the output probabilistically linked to that application.

Shannon: a message is encoded into a sequence of those signals, and the recipient decodes them by mapping them to the message.

Wiener, applies a signal to a channel, and the recipient estimates some property of the input.

"explains organized complexity in terms of the reduction of entropy… that is achieved when systems.. absorb.. from external sources and convert it to a pattern or structure." (simon 127)

information value

= Expected value(x) – Expected Value(no information)

insurance actuarial calculations

Involves the loss experienced by many customers. Risk of rare loss incidents are not measurable, not controllable, and ergo not manageable by at-risk parties

Why insurance makes for a bad mitigation driver in security. Future Risk Involves:

- Determination by unknown assailants not by victims
- Unknown time of attack
- Unknown methods and resources
- Unknown purposes
- Unknown and known vulnerability

- Unknown consequences

This assessment has little shown methods to change victim behaviour and reduce risk

intelligence$_2$

An information gathering process, especially for information that is difficult to obtain. This information, usually, is intended to support decisions by policy makers.

- The information gathering methods and discipline
- How the information is analyzed
- How the information will be used
- How the conclusions or information is revised
- The sensitivity to misinformation

see also *traffic analysis*

derived intelligence

technical intelligence

- How gathered, which limits it to kind of information gathered, and couples with analysis techniques

- Images, source test interception (ala government documents) voice/video interception, reports from people (confessions, boats, etc.)

Discovering problems, opportunities.

***Table 3:** Different types of intelligence*

Terms	*Sources*
COMINT	Communications intelligence
diplomatic intelligence	Assess intentions
ELINT	Electronic means
HUMINT	Talking to people, interrogation
military intelligence	assess capabilities
IMINT	photo from plane, balloon, satellite
SIGINT	Signals intelligence
Standards & Measurement	

Distinction of the terms

***Table 4:** Distinctions*

Term	*Meaning*
communication	How to form and interpret information, communicated between parties
diagnostic	Trace to causes
intelligence	Forming the who-what-why-how information of signals, communication, and circumstances
safety	How to prevent or reduce risks or harms
security	Preventing, reducing, etc others from finding the information above

analysis

Derived info about organization structure

facts about their current state

plans for future organization structural actions

business

Concerned with facts and measures about a business's internal operations, its industry, customers, competitors, and other market conditions.

There are some flaws. First, it is very expensive (often several hundred million) and takes years to deploy. Second, the tools don't reasonably support basic information; Every Burger King has an SQL Server, but none can immediately answer `How many Whoppers have I sold today?' Third, the tools are too separate for end users. The user thinks in terms of a series of related steps serving one purpose. They not given one cohesive application; rather, they have several, complex applications that operate differently, and they have to forcibly switch between them.

From an IT perspective, these tools required deploying a network on top of the network -- server support, application support, protocols, update "router configurations," security and etc.

goals

Seeks to answer:

- "*Do we have* an enemy/adversary? If we do, *who* is he and *why* is an opponent?
- "*Where* is he and in what *strength*?
- "What are his *intentions?*
- "What are his perceptions of himself, *his own purpose and goals?*
- "Are his intentions *consistent* with his strength (capability? If not , might he be practicing *deception?"* [v2n12]

open source

Concerned with handling information gathered from sources open to the public.

signals

For electronic, see above measurement & signature

HF direction finding

callsign (operation). Different systems of callsign

Traffic analysis / network flow

identify the cipher system, codebook.

What plain text language?

intelligence agency

gather intelligence

analysis

covert action

who is doing what to whom, and for what reasons

intelligence analysis management

cycle:

plan

collection: instrument / acquire / measure

analysis

processing & exploitation

interpretation

distribution

intercept linguist

A linguist determines

- the content of the message or the voice transmission
- its intelligence value, if any
- the appropriate vehicle for reporting the information

Kerkoffs, Auguste

Wrote “La Cryptographie Militaire” (1883), most famous for its separation of strength of cryptography system and keeping its mechanisms secret. This has since become an ideological assertion that one should not keep its mechanisms a secret, a misreading. He also insisted that the tools “must be easy to use and must neither require stress of mind nor the knowledge of a long series of rules”

These are graceful degradation of secrets.

labeled

Such systems have a means of identifying all the things in the system – called tags.

Each item gets a *label* – a list of tags.

These are tracked. X can only speak to Y if it (or its secrecy label) hasn’t seen any secretes. X’s permission cap can only be decreased.

Data polities: items can change their labels by adding and removing tags; items can allocate tags.

This system is very similar to capability systems.

legal aspects

legal issues: evidence, chain of custody, signal theft, legal discovery

legal impact

Collecting so much private information may overwhelm the legal system (esp. w. discovery). Everything in US is being preserved, include. Memory sticks, cellphones, bills, PDAs. The cost of lawyers is $125-$600/hour, with up to 200 lawyers pouring over the info,

Merits of case are even more likely to be lost in the tactics.

In our adversarial system, opponents decide what to keep.

In “inquisitional legal systems” (France & Germany) the discovery is proportional to the case, with judges determining relevancy.

levels of system security

DRM style enforcement of contracts ‘doesn’t work’ – alienates too many people, may never work, may be circumvented.

Auditing behaviour on demand (tracks sale / distribution)

Super distribution: transparency, accountability.

loss function

consequences of wrong decision

market
brining data to market

Bringing data to market

1. What data do we own or have access to?
2. Which of these are of sufficient quality to withstand scrutiny?
3. Which will others pay for?

Options:

1. Provide new content: now, novel, new uses, better quality, personalized
2. Repackage: filter, synthesize, reformat from others
3. Refit current offerings to create and deliver more useful data
4. Unbundle: extract data from existing product, and sell
5. Do you have better access to or understanding of data?
6. Find who to sell to level playing field
7. Provide identifiers
8. Provide sift and understanding tools (behaviour, trend, structure, etc.

measurement and signature intelligence

deviation from norms

MASINT signatures: distinctive characteristics

describing

identifying

detecting

tracking

example radar:

measurement finite metric parameters of a subject

signature distinctive features of subject as sensed by instrument

Radio/transmitter strength

manufacturer model, make

same vs different transmitter

direction, triangulation, distance

propagation path

frequency, modulation form

antenna

memory systems

- Relational complexity in storage, wallowing complexity
- Streaming systems: metadata & stream, continuing where one left off, chunk transfer, transfer encoding
- Transactioning system
- Compression: input data / output data

metrics

precision % of relevant documents retrieved out of all relevant documents in database

recall % of relevant documents retrieved out of all retrieved documents

modeling Elements:
See also *attributes, data dictionary*

money laundering legitimate businesses, transactions, but for more than the booked price. The later is also used to lower taxes.

mutual information

$$mutual \inf o(x, y) = \log \frac{P(x, y)}{P(x)p(y)}$$

X can be John, y can be calls(DC), calls(Bogota), etc.

can be used with a similarity measure, e.g. cosine similarity.

nodal analysis What a person does, and how they live

nonrepudiation Providing auditable proof of origin, delivery and other elements of transmission or agreement.

orientation Process-oriented: focus on performing operations with data

Data-oriented: focus on the structure, storage and maintenance of data.

pattern recognition Adjaciewz grammar

Tries (if x=b) do x; goto next item in sequence, goto Rule N

Adaptive resonance theory – similarity is called ‘resonance’

permissions
conditional

“let an entity perform a certain action or set of actions under the condition that it will assume certain additional responsibilities.

Lalana Kagal, Tim Finin, Anumpam Joshi, Sol Greenspan, “Security and Privacy Challenges in open and Dynamic environments” IEE computer, June 2006

photogrammetry

Photo info: authenticity, time of day, ephemeris

Fog/haze removal

Panopticon

photo analysis

quality assessment

clustering

near-duplicate detection; aggregate & select

face detection

concept detection

importance detection

coverage-importance

interpretation of photo

Surface illumination:

- What are the light positions?

- Fog?

Something is at location x: it’s height, type (plane, tower, etc), make, model,

# of people

Items of interest

photograph

who: who is in the photograph? who are these people?
what objects are I the photo?
time: when the photo was taken?
location: where the photo was taken: the event, other photos from the event.”

PICL

Presidents, intelligence, check list. PICL (Pickle) Factory

pin–tumbler lock

Apparently, the ancient Egyptians invented such devices, but they were lost with the decline 4,000 years ago. Linus Yale also invented some, in an attempt to create one that he couldn’t pick

policy

“An explicit representation of constraints and rules that govern an agent or system’s behavior.. Explicit policies can define permissions, obligations, norms, and preferences for an agent’s actions and interactions with other agents and programs. Such policies, especially those expressed in high-level declarative languages, can form the basis for electronic contracts and provide a sublanguage useful for negotiating agreements and commitments.” (p89-

Lalana Kagal, Tim Finin, Anumpam Joshi, Sol Greenspan, “Security and Privacy Challenges in open and Dynamic environments” IEE computer, June 2006

privacy
contextual integrity

Four classes of variable / datum set

1. Principle of transmission: describes the basis on which information flows. This can be contractual terms, legal rights, gossip, etc.
2. The capacity(s) in which the individuals sending and receiving the information are acting.
3. The context of a flow if information
4. The types of information involved

Linear Temporal Logic is being investigated to evaluate such instances.

Helen Nissenbaum, New York University

Adam Barth

privacy laws

health: Health Insurance Portability and Accountability Act (HIPAA)

education records: Family Educational Rights and Privacy Act (FERPA)

financial records: Gramm-Leach-Bliley (GLB) Act

Wiretap Act

Pen Register and Trap& Trace Act – dialing, routing, addressing, and signaling information

principles of

OECD Basic Privacy Principles of National Application

- Limits to collection of personal data; knowledge & consent of subject
- Quality of data, its relevance for purpose or use; requirements on keeping it accurate, complete, and up to date.
- Specifying the purpose for collecting it
- Limited use, and limited disclosure
- Safeguarding and securing the data against unauthorized access, destruction, use, modification or disclosure
- Openness about the developments, practices, and policies; provide means to establish existence and nature of purpose of personal develop, as well as other features listed hear.
- Individual participation. The individual should have the right to obtain confirmation of whether or not they have data relating to him, to have the data communicated to him, in a readily intelligible form. To be given an explanation if denied, and to be able to challenge the denial; to have (on successful challenge) the data rectified, erased, completed, or amended.
- The company (and its representatives) should be accountable for complying with these principles.

types of

Types of privacy include: territorial (of space), bodily, communication, information.

Issue of control over disclosure

private

Data transmission is encrypted using the targets public key.

public key

Create session symmetric key. Typical and max duration to use the key for the session.

quality of data/information

true/unquestionable… captured information

high degree of reliability … lesser value

suspected garbles; project / expected info (but not observed)

information in its original form

compromised information

highly suspect information

question and answer

Classify question, generate list, rank.

See also *information extraction*

Table 5: **Type of query and mode of response**

Word/phrase	*Answer entity*
How far / how long	Dimension
How many	Number
How much	Currency
What/which	Person, organization, company, place
When	Time
Who is	Organization
Who/whose	Person

rainbow books

Colored book protocol (rainbow book specs) ~1979
Blue Book: file transfer
Pink book: ethernet protocols

Yellow book: network independent transport services
Green book: TS29 Terminal protocol
Red book: Job transfer and manipulation protocol
Grey book: Mail protocol
Orange book: Cambridge ring 82 hardware & protocol specs
Fawn book: screen management protocol

rainbow table A variation on dictionary attack, for fast reversal of encrypted passwords. Specific to the hash function used.

reference monitor Every access for a given resources thru a monitor.

Jim Anderson, Computer Security Technology Planning Study*, 1972*

ring based protection Limits the privileges to resources, access, instructions, and capabilities available to processes within the ring. Higher privilege rings have all the privileges of the lower privilege rings. Each ring is enumerated, typically with lower numbers denoting higher privilege levels. Kernels typically exist in Ring 0; applications typically exist in ring 3.

Switching between two ring levels. Often this involves – a means of a lower privilege process to communicate with a higher privilege level. It is a type of call, that employs special entries inaa descriptor table, and an associated selector.

risk management either accept it, mitigate it, or transfer it.

role The element of *identity management* that groups operations (and/or other roles). A role is a set of actions, and/or a role. A role may have a password. Description of allowable access characteristics of host use. Determines scope of access a user has to files and privileged operations. Identifies the use of privileges on a case by case basis. Set once by the access path (and never allowed to change).

RSA Security derives from the difficulty of factoring N. The values M^F and C^P are large and tedious to compute.

decryption $C=M^E$ mod N

encryption key $C=M^E$ mod N

where C is the cipher, M is the plaintext

private key A secrete, large number that has no common factors with the product (P-1)(Q-1) of the public key

public encryption key A number E such that ED mod (P-1)(Q-1)=1

schelling point

Schengen Schengen Information Systems (via the Schengen Treaty)

SOX Sarbannes Oxley “A formal, explicit specification for the integrated use of multiple data sources in a semantic way”

“A conceptual representation of the types and properties of data and knowledge..

“A vocabulary of terms and relationships to model the domains specifying how to view the data sources.”

Sox *Sect 404 “Management Assessment of Internal Controls.” Sect 302 “Corporate Responsibility for Incident Reports.”*

secure boot The history includes:

1970’s – US Government agencies & Military explore ideas. Ford Aerospace, Comm Corp. (now in Lockhead-Martin) opened trusted OS divisions.

1980s – SecureWare (by SecureWare, Inc) was first commercial Trusted Operating System

1993 – market was gone, too expensive, low demand

William Arbaugh, primarily. (University of Pennsylvania). “Security for Private Intranets.” IEEE Computer, September 1998. and others in 1997-1998. Looks like Palladium

Sandra Kay Miller, “Trusted OS Makes Comeback” IEEE Computer, Feb 2001, p16-19

? Only load files from controlled source; files are lock as read-only to prevent changes during loading.

asset identification

security policies and threats definition

risk assessment

protection measures design

validation

security attack Types of cases: confidentiality case (getting access to information they aren’t supposed to see). Denial of Service (availability case). Integrity Attack (the kind meant to be seen, the kind meant not be seen)

controls countermeasures to address risk assessment.

evidence

- Unusual requests for resources
- Origins, signature
- Storage place (in application directory, temporary folder, network drive, etc.)

forward security if the system is compromised at some point, it does not stay compromised forever.

guidelines

1. Secure the weakest link
2. Practice defense in depth
3. Fail securely
4. Follow the principle of least privilege
5. Compartmentalize
6. Keep it simple
7. Promote privacy
8. Remember that hiding secrets is hard
9. Be reluctant to trust
10. Use your community resources

John Viega, Gary McGraw ;Building Secure Software: How to Avoid Security Problems the Right Way Addison Wesley, 2001, http://www.buildingsecuresoftware.com/

in depth The depth is the *sequential* steps that must be taken to achieve bad things. Deeper security has more required steps. Often confused with multiple security measures that operate in parallel.

levels of None: no security
Light: connection is validated or authenticated
Moderate: Connection is validated and authenticated
Stringent: Conversation is Private.

token small electronic tag or smart card. Used instead of a password

zone Security zone: range from unsafe to very stringent. Specifically in windows: NoZone*, Untrusted, Internet, Trusted, Internet, My Computer

secure software construction

principles of a security design

1. Economy of mechanism; too many different ways to do things leads to failure
2. Fail-safe defaults
3. Complete mediation (see *reference monitors* that are always used, consistent mechanism to protect resources)
4. Open design – obscurity isn’t a critical part of the design
5. Separation of duties – multifactor verification for access
6. Least privilege

Saltzer, Jerome; Michael Schroeder The Protection of Information in Computer Systems, 1974

7. Least common mechanism – resources aren’t implicitly shared
8. Psychological acceptability – system understandable by users.

signed application Detects *new bugs* – and any other code alternation – since it was signed. It does not protect against bugs already in the code. Often implemented by signing the set of files in the application, as a whole. Signed with certificates, and need to chain back to some level. Must verify the certificates authenticity, the responsible party’s identity, and credentials, etc. Certificates can be revoked. Reduces the variety of ways that a file can be corrupted and not detected, but makes for many ways that the security system can be violated.

Loader is responsible for verifying and loading executable content, and (possibly) transferring control to the executable module.

signed file Identifies origin (similar to a watermark) and other info in a hard to fraudulate manner.

social media channel, credentials, media it accepts

software connections The ways in which software modules can be interconnected: files, IO, semaphores, mutex, shared memory, data structures, API, signals, Interrupts, DMA.

speech / spoken / voice interception techniques

record techniques

crypto-linguistics identify the language

transcribe: note garbled words, note estimated words

linguistics Translate

Specialized topic? Vocabulary idioms, etc

Special ‘code words’ related to communication

speaker recognition

speech recognition

- transcribe

- sentiment

- literal / polarity (sarcasm)

split horizon Information about a network (i.e., name to IP address, ARP information, routing information, etc.) can only be sent to interfaces other than the one it was received from. That is, if the information was received on en0, it can’t be answer queries with that on en0. This prevents various problems.

spoofing Falsifying messages

standards Department of justice maintains a standards registry.
http://it.ojp.gov/jsr/public/list.jsp

summarization Key elements of text, text tiling, structural recomposition, tracking related events in multiple documents, sentiment analysis, attitude and disposition

surveillance “Surveillance involves monitoring persons or locations to identify behaviors, activities, and other changing information.” “adversarial surveillance that is to gather information in preparation for an aggressive action and likely criminal in nature.”

TANGRAM NSA. Suffers from guilt by association. Related to TIA. Normal behaviour is classed as criminal.

taxonomy focus on noun concepts, organizes, translations to automation, allows interpreting knowledge base

terms
categories of terms

1. “primarily of concern to traffic analysis”

2 “terms of secondary concern in traffic analysis”

3 no concern,

4. cross reference.

see also cover term

text fusion

Several articles (*texts*) about the same topic are merged and summarized.

- How to find the subset of articles discussing the same thing
- How to related the two articles on common areas
- How to handle disagreement on common portions
- How to identify ‘new’ information
- How to select it – or filter our non-sense, based on rumour
- How to merge into a single structure
- How to revise structure into text with prosody

See also *analogy*

text mining software companies

Autonomy, HNC, UBM

threats

Categories of threats include:

1. “Act of human error failure (accidents, employee mistakes)
2. “Compromises to Intellectual Property (piracy, copyright infringement)
3. “Deliberate Acts of Espionage or Trespass (unauthorized access and/or data collection)
4. “Deliberate Acts of Information Extortion (blackmail of information disclosure)
5. “Deliberate Acts of Sabotage or Vandalism (destruction of systems or information
6. “Deliberate Acts of Theft (illegal confiscation of equipment or information)
7. “Deliberate Software Attacks (viruses, worms, macros, denial of service)
8. “Forces of Nature (fire, flood, earthquake, lightning)
9. “Quality of Service Deviations from Service Providers (power and WAN service issues)
10. Technical Hardware Failures or Errors (equipment failures)
11. Technical Software Failures (bugs, code problems, unknown loopholes()
12. Technological Obsolescence (antiquated or outdated technologies)”

Whitman, Michael, Enemy at the Gate: Threats to Information Security Communications of the ACM, August 2003

traffic analysis

Change in behaviour, statefulness. Organizational pecking order, negotiating activities, planning activities

mask

A bitmask of signals, most commonly used for passing to sigprocmask() or pthread_setmask() to block or unblock signals. Manipulated with sigaddset(), sigdelset(), sigfillset(), sigemptyset().

pending

Generated but not yet delivered or accepted. Usually only for a moment, but for longer if the signal is blocked (see below). Can be checked for with sigpending().

synchronous

Occurs at a definite point in the program's execution. For example, a SIGSEGV is triggered right on the instruction that accesses invalid memory. A SIGPIPE is raised during a read() or write(), not after. All synchronous events are threaddirected.

trust metric One of a variety of foo bah

trust management Certifications and other's have a level of trust based on who attested to it, and how much they are trusted.

two worlds

1. A model of the user's data, used for the user's problems
2. Relational model of the system, used by the compiler / translator (this is not accessible by the users)

types

- Few people have access to information, and it is hard to gain access
- Data is duplicated everywhere and in many forms; leaving it chaotic
- Information flows in free but managed way
- Information flows beyond the immediate organization to partners, suppliers, and customers.

It is hard to mature a company thru those stages. It is costly and resource intensive to secure data. Security impedes productivity, decision making.

URI Uniform Resource Identifier: can be a URL; does not have to specify protocol. Can be a resource name but does not have to be globally unique

URL Uniform Resource Locator: the location of an object (specifies access protocol)

URN Uniform Resource Name: globally unique name

user allow action The set of all possible actions that a user can perform. Constructed by constructing a labeled transition system.

a) The definition of the users state

b) A set of inference rules:

- How the users capabilities

- preconditions required for operations on resources

- current state of users. Updated description of the system

validation of party A
A encodes a known message using A's private key.
The ciphertext is sent to B
B decodes this message using A's public key
If the message can be decoded property, then A has been validated to B.

of party B
B takes part of the original message & adds in its own information.
This is encoded using B's private key
This ciphertext is sent to A
A decodes the message using B's public key
If the message can be decoded properly, then B has been validated to A.

voting 3 ballot system

watermark Tracking the origin of a file by associating data with the file.

Safe against who & what? Safe enough vs safe as people think it is.

US British governments sold refurbished Enigma machines to 3rd world countries for several decades.

dynamic The participants change regularly (not just do to occasional failures)

open Doesn't pre-identify the set of known participants

Best to describe what could be 'a constitutional' NSA program. It is legal, and the home officer's role is to provide information about whether it is legal. Foreign operations revelations went too far.

legal directives bulk vs targeted
discriminant: definition, examples
selector
classification to cover conduct, bad use

it is easier to keep a program or approach going.