Info theory and big data

"Typical or not-typical, that is the question"

Han Vinck University Duisburg-Essen, Germany
September 2016

Content: big data issues

A definition:

- **Large** amount of **collected** and **stored** data to be used for **further analysis**

  - **too large** for traditional data processing applications.

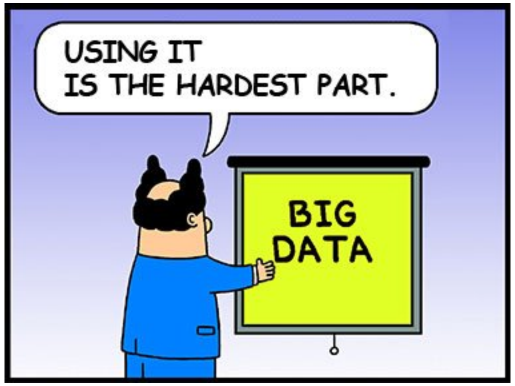
Benefits: We can do things that we could not do before!

- **Healthcare**: 20% decrease in patient mortality by analyzing streaming patient data.
- **Telco**: 92% decrease in processing time by analyzing networking and call data
- **Utilities**: 99% improved accuracy in placing power generation resources by analyzing 2.8 petabytes of untapped data

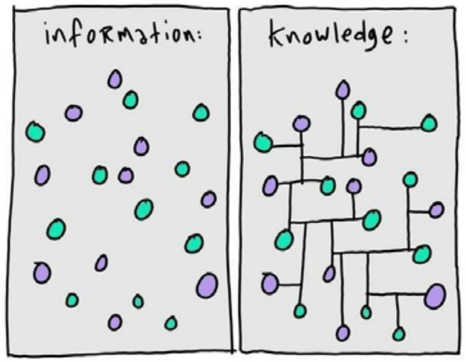
Note: Remember that you must invest in **security** to protect your information.

Big data:

- Collect - , store - and draw conclusions from the data



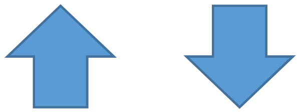
- Some problems:
  - <u>extract</u> knowledge from the data: <u>***Knowledge***</u> is based on <u>***information or relevant data***</u>
  - <u>what</u> to collect: variety, importance,
  - <u>how</u> to store: volume, structure
  - <u>Privacy,</u> <u>security</u>



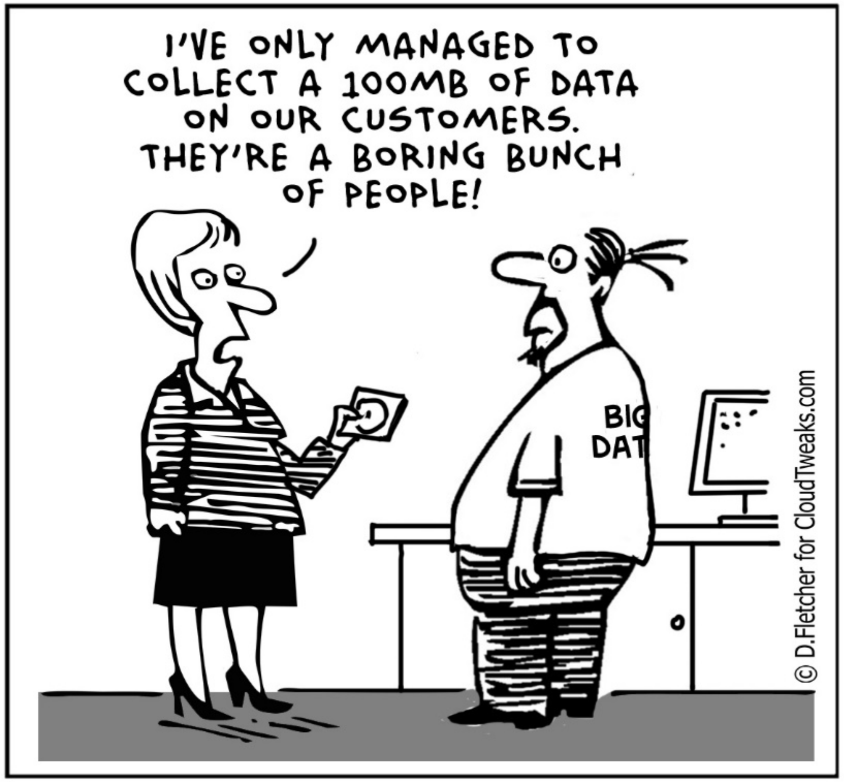
What kind of problems to solve?

<u>There are</u>:

- Technical processing problems
  how to collect and store

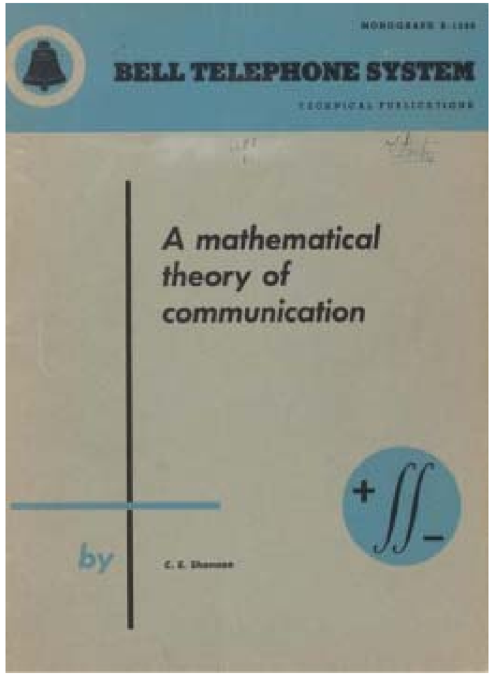


- Semantic-content problems
  what to collect and how to use



information theory can be used to quantify information and relations

claude shannon
1916-2016



Two contributions of great importance

Communication Theory of Secrecy Systems*

By C. E. SHANNON

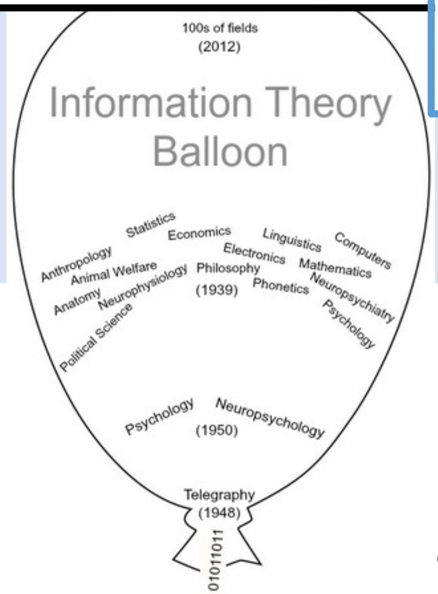
1956, Shannon and the „BANDWAGON“

- Shannon was critical about „his information theory“

tions. I personally believe that many of the concepts of information theory will prove useful in these other fields—and, indeed, some results are already quite promising—but the estab[…] is not a trivial matter of […] domain, but rather the[…] hypothesis and experim[…]

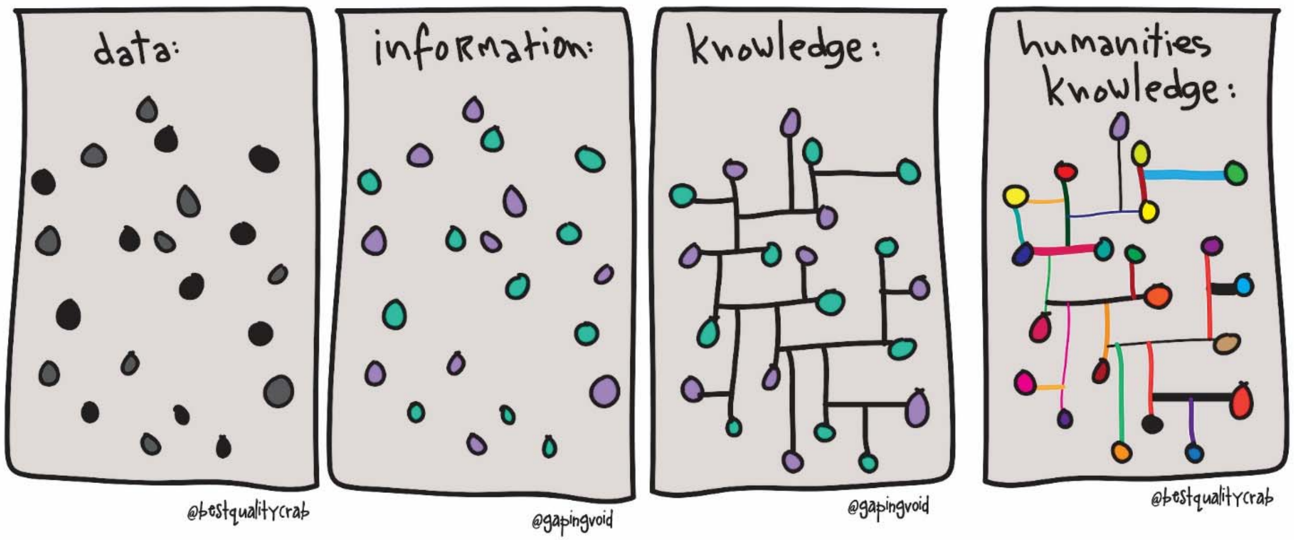
Secondly, we must keep our own house in first class order. The subject of information theory has certainly been sold, if not oversold. We should now turn our attention to the business of research and development at the highest sc[…] tain. Research rather tha[…]

tain. Research rather than exposition is the keynote, and our critical thresholds should be raised. Authors should submit only their best efforts, and these only after careful criticism by themselves and their colleagues. A few first rate research papers are preferable to a large number that are poorly conceived or half-finished. The latter are no credit to their writers and a waste of time to their readers. Only by maintaining



nice picture (often used) to illustrate the idea of content

Context =>
Understanding=>

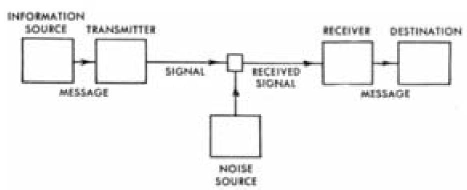


Who, what,, ... How? Why?

semantics are used to make decisions or draw conclusion

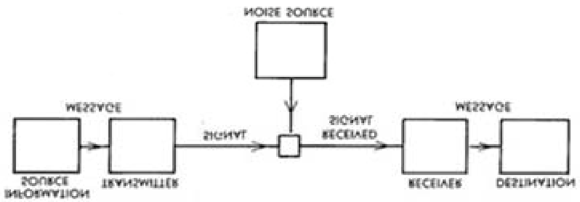
Shannon and Semantics

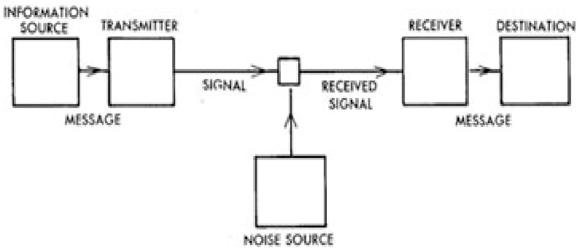
The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.



Shannon 1916-2016

Extension from the Shannon Fig.1 to the system using semantics





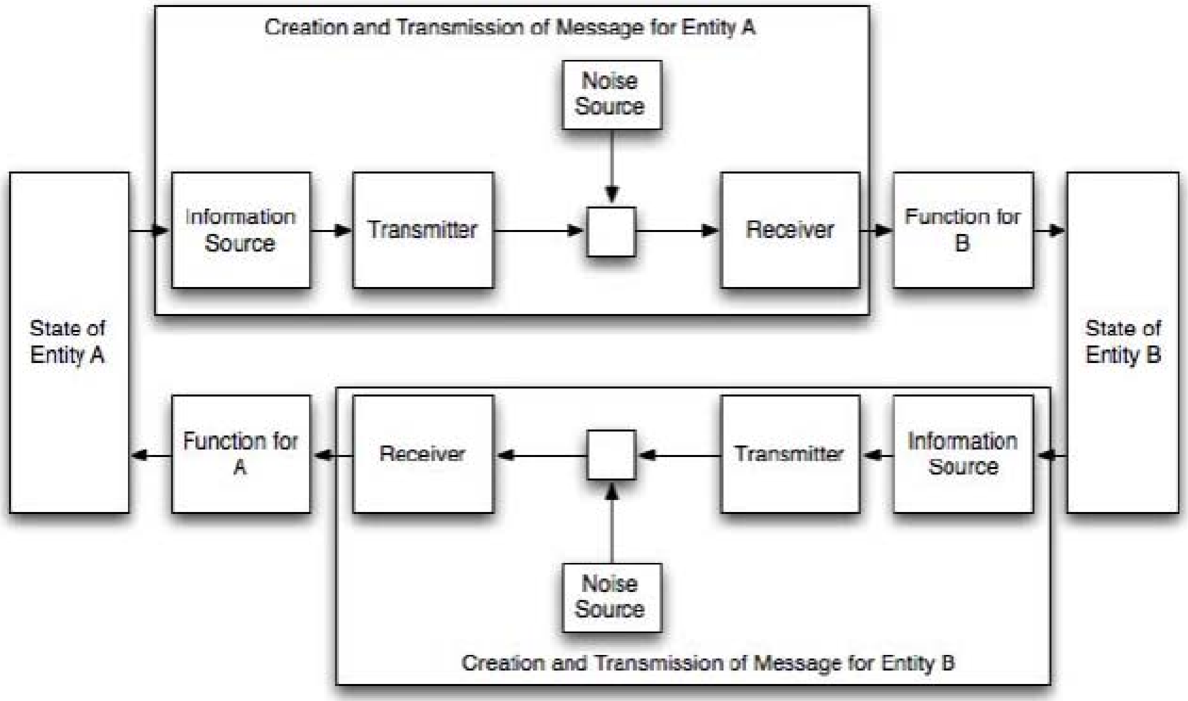
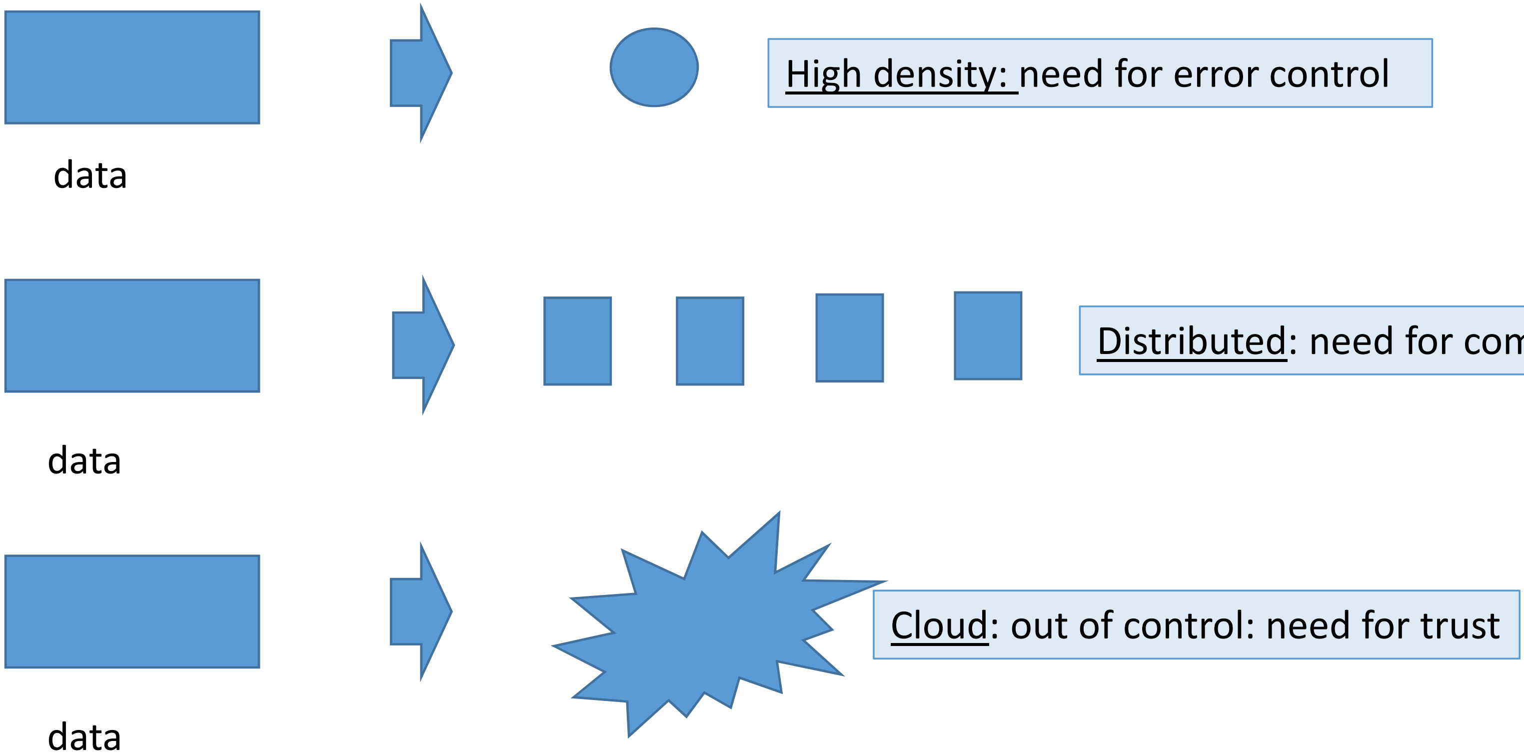


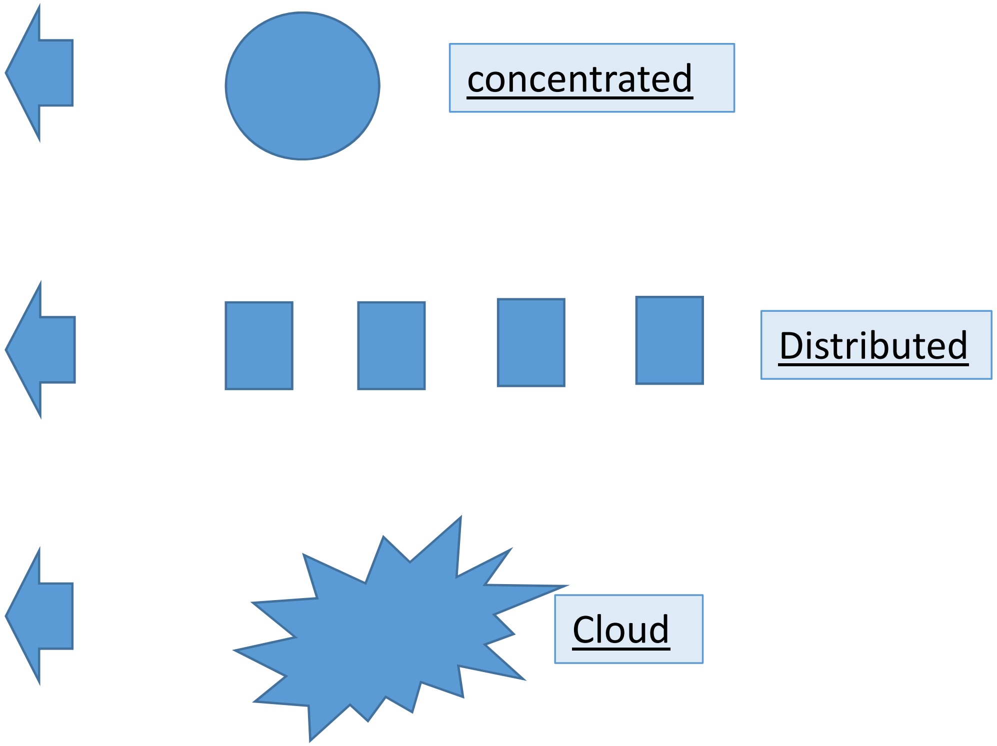
Figure 3.2. The above transactional diagram can be understood as an extension of Shannon's original diagram

How to store/ large amounts of data?



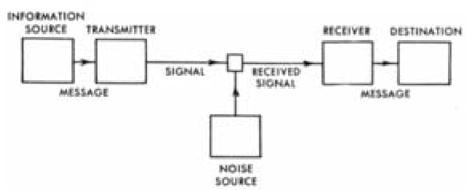
How to access large amounts of data?

Problems: - where? – who? – how?



Shannon‘s <u>reliable</u> information theory

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.



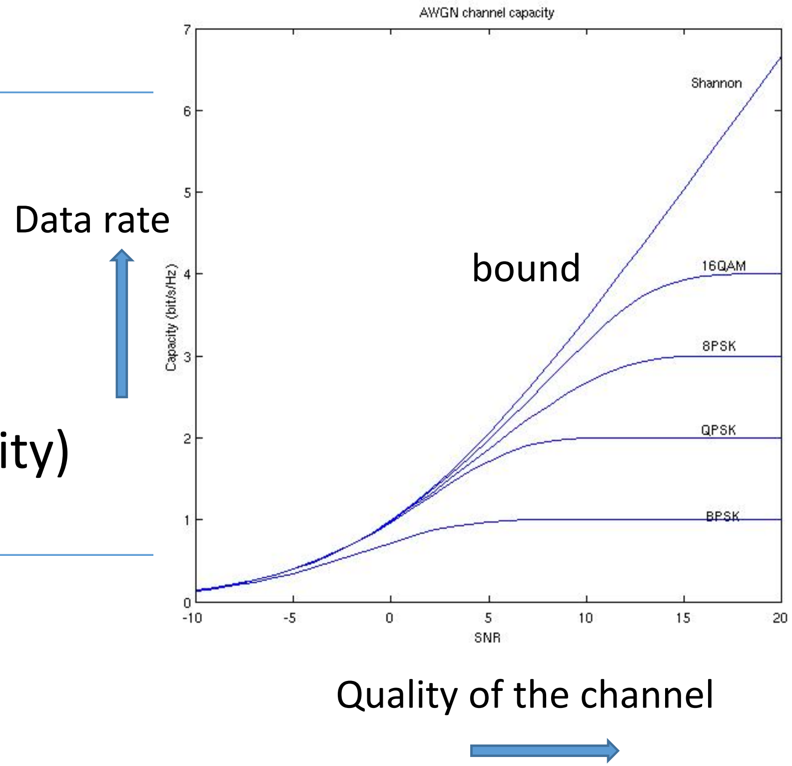
Communication: transfer of information

knowledge is based on information

Reliable transmission/storage: Shannon

NO SEMANTICS !

- For a certain transmission quality (errors):
  - codes exist (constructive)
  - that give P(error) => 0
  - at a certain maximum (calculable) efficiency (capacity)



Large memories are not error free!

- SSD drives use BCH codes that can correct 1 error or detect 2 errors.
  - Can we improve the lifetime of SSD when using stronger codes?
  - How big is the improvement?

3,8 TByte

My MsC computer (1974)
44 kB main memory!
1 Mbyte hard disk

Assuming that memory cells get defective: Memory of N words

On the Influence of Coding on the Mean Time to Failure for Degrading Memories with Defects

HAN VINCK AND KAREL POST, MEMBER, IEEE

operational state to the permanent defect state. We give bounds on the MTTF and show that, for memories with N words of k information bits, coding gives an improvement in MTTF proportional to $(k/n)N^{(d_{\min}-2)/(d_{\min}-1)}$, where $d_{\min}$ and (k/n) are the minimum distance and the efficiency of the code used, respectively. Thus the time gain for a simple minimum-distance-3 code is proportional to $\sqrt{N}$. We also

GAIN in MTTF $= \frac{k}{n} N^{\frac{d_{min}-2}{d_{min}-1}}$

For a simple d_{min} = 3 code the gain is proportional to $\sqrt{N}$

If, on the other hand, chip surface is costly or the system is unrepairable (satellite systems), then one is interested in the average amount of chip surface needed to realize a time T. The chip *surface gain* is defined as

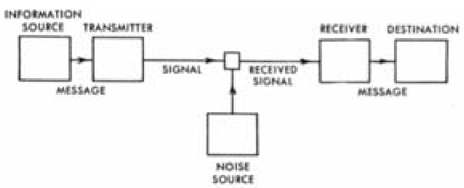
$$\gamma = \frac{\dfrac{kT}{\text{MTTF(uncoded)}}}{\dfrac{nT}{\text{MTTF(coded)}}} = \frac{k}{n}\eta.$$

Chip surface needed to realize time T

Shannon's information theory

NO SEMANTICS !

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.



- Assign - $\log_2 p(x)$ bits to a message **from a given set**
  => likely, short => unlikely, large

- Shannon **showed how** and **quantified**:

the minimum obtainable average assigned length

$H(X) = - \sum p(x) \log p(x)$ (**SHANNON ENTROPY**)

Data compression (exact reconstruction possible)

Exact: representation costly (depends on source variability!)
Need a good algorithm (non exponential in the blocklength n)

Data reduction (no exact reconstruction)

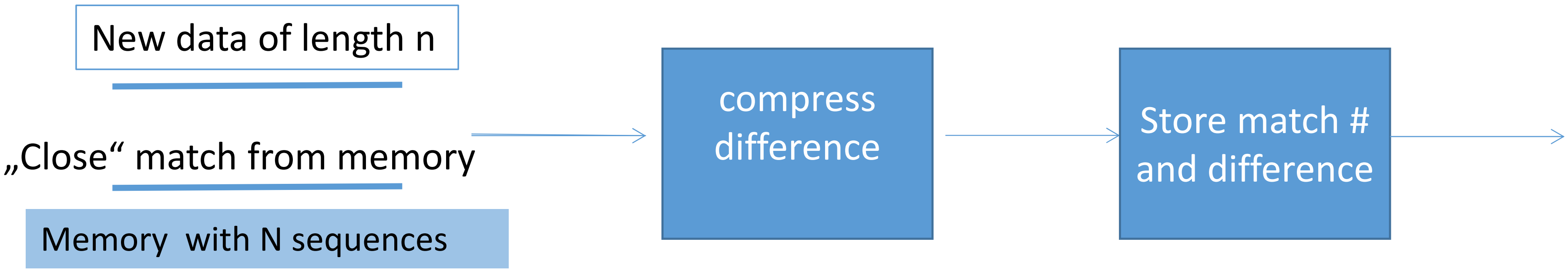
" Take this report and reduce it to an ***acronym.*** "

NOTE: In big data
we are interested in the NOISE!

No exact reconstruction: good memory reduction, but in general we lose the details

- how many bits do we need for a particular distortion?
- need to define the distortion properly!

algorithms (old techniques from the past) to avoid large data files



If we use N sequences from the memory, we need:

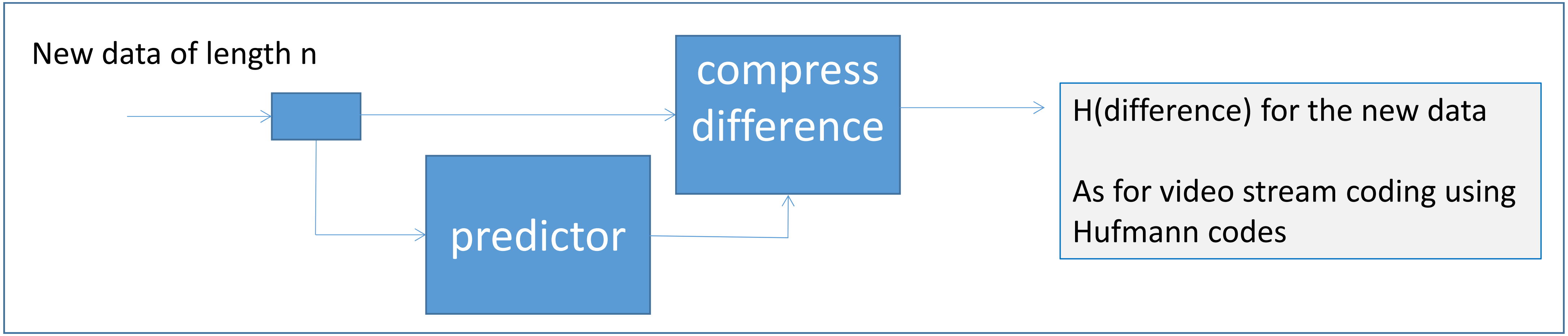
$k = \log_2 N$ bits for the memory data + H(difference) for the new data

Memory can be updated. (frequency of using a word)

Optimization: # of words in memory versus difference

Modification to save bits for sources with memory

- Use prediction



Example: video coding using DCT and Hufmann coding

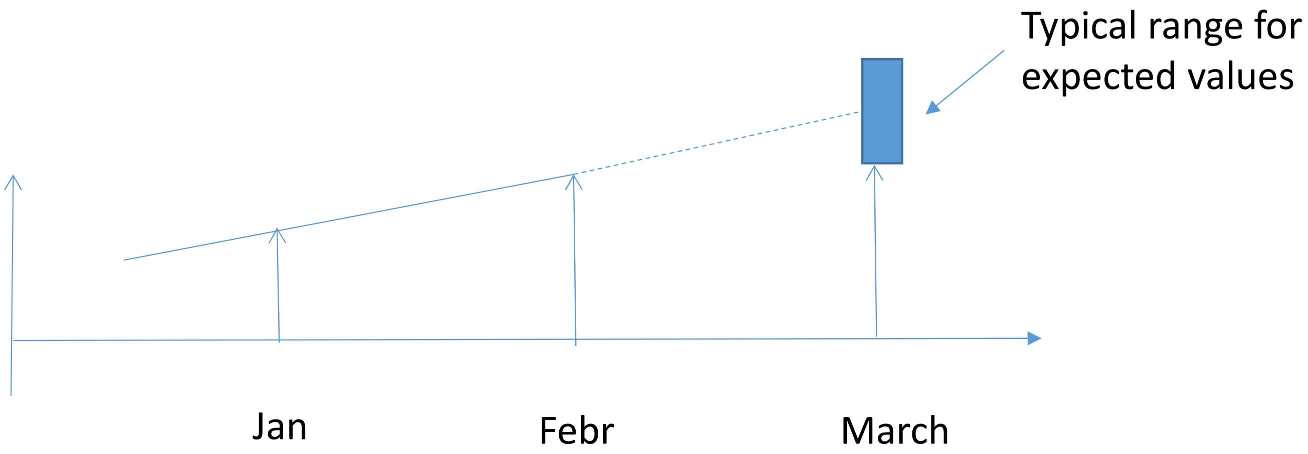
Shannon prediction of Englisch (again, no semantics)

IN A previous paper[1] the entropy and redundancy of a language have been defined. The entropy is a statistical parameter which measures, in a certain sense, how much information is produced on the average for each letter of a text in the language. If the language is translated into binary digits (0 or 1) in the most efficient way, the entropy *H* is the average number of binary digits required per letter of the original language. The redundancy,

Fig. 2—Communication system using reduced text.

Example: showing the importance of prediction

- Metering: only the difference with the last value is of interest
  - If typical consumption, within expectations, encode difference
  - If a-typical, encode the real value

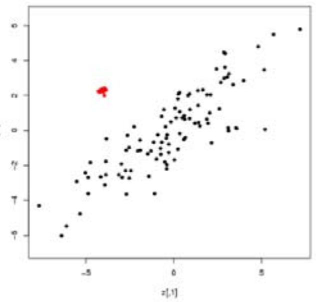


Total consumption in time

An important issue is <u>outlier</u> and <u>anomaly</u> detection

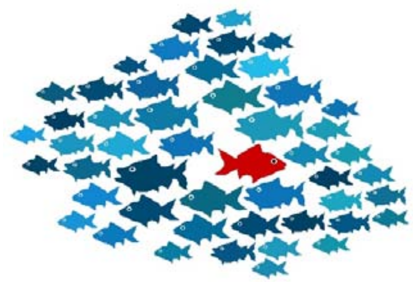
- **<u>Outlier</u>** = legitimate data point that's far away from the mean or median in a distribution

  Ex: used in information theory

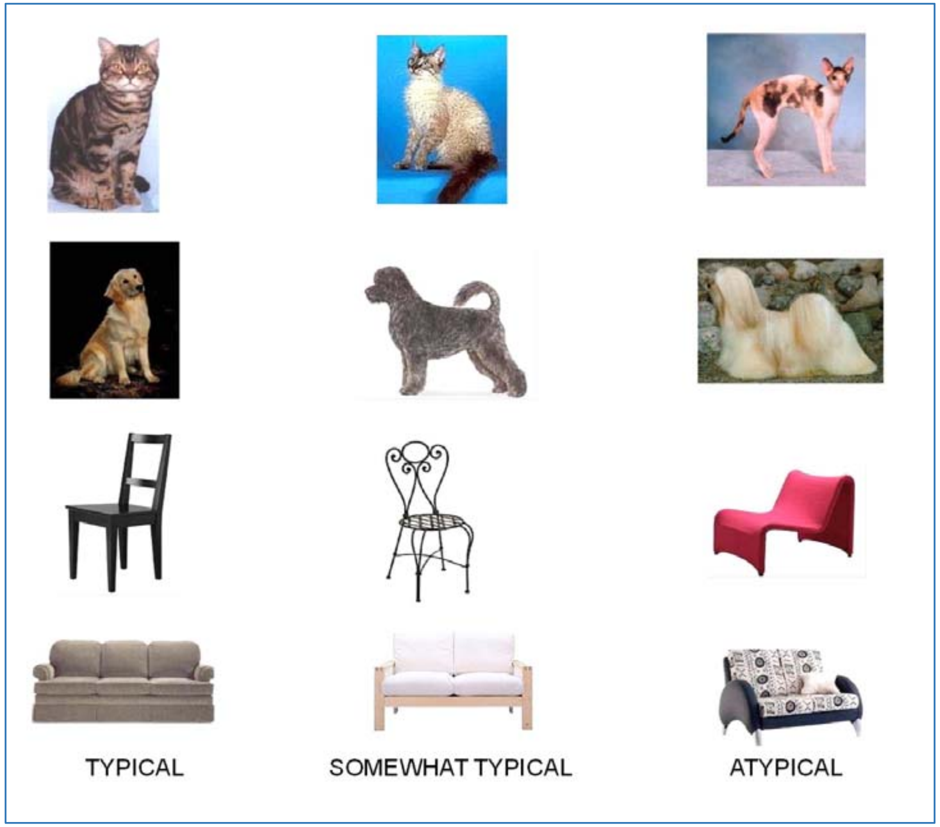


- **<u>Anomaly</u>** = illegitimate data point that's generated by a different process than whatever generated the rest of the data

  Ex: Used in authentication of data



Further problems appear for classification



What is normal?

Classical information theory approach: outliers

- Information theory focusses on typicality:
  - set of most probably outputs of a channel/source
  - uses measures like entropy, divergence, etc...

Definition 1. *Entropy-typical sequence. A sequence x^n is said to be typical with respect to an $\epsilon > 0$ and $P_X(\cdot)$ if*

$$\left| -\frac{1}{n} \log_2 P_X^n(x^n) - H(X) \right| < \epsilon.$$

Note that this is equivalent to,

$$2^{-n[H(X)+\epsilon]} < P_X^n(x^n) < 2^{-n[H(X)-\epsilon]}.$$

This notion of typicality is only concerned with the probability of the sequence and not the actual sequence itself. Next we define a stronger notion of typicality, called letter-typicality.

NO SEMANTICS

Properties of typical sequences (Shannon, 1948)

Theorem 1. *Suppose that* $0 \le \epsilon \le \mu_X$, $x^n \in T_\epsilon^n(P_X)$ *and* X^n *is emitted by a DMS,* $P_X(\cdot)$. *We have,*

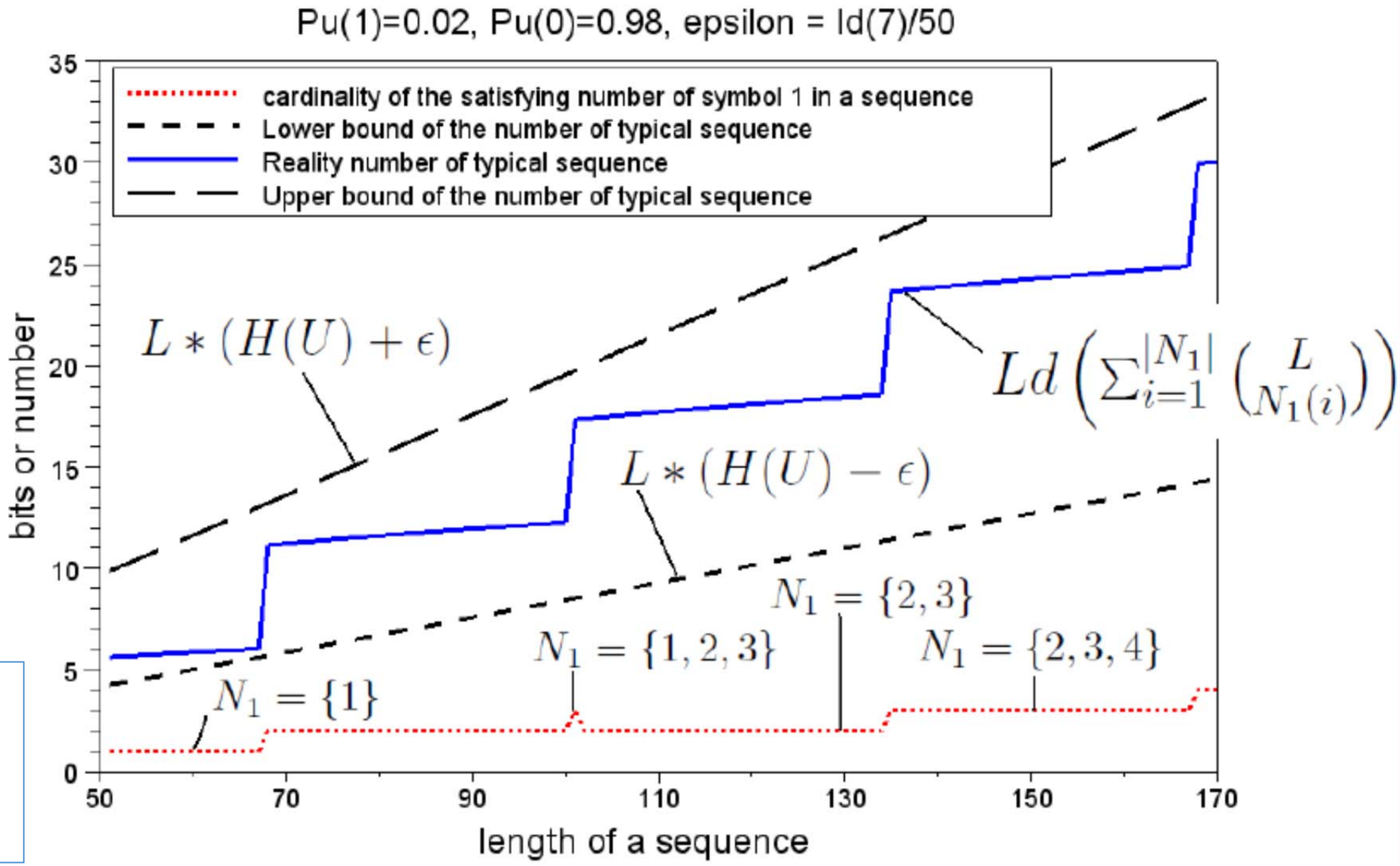
i) $2^{-n(1+\epsilon)H(X)} \le P_X^n(x^n) \le 2^{-n(1-\epsilon)H(X)}$.

ii) $(1-\delta_\epsilon(n))2^{n(1-\epsilon)H(X)} \le |T_\epsilon^n(P_X)| \le 2^{n(1+\epsilon)H(X)}$.

iii) $1-\delta_\epsilon(n) \le P[X^n \in T_\epsilon^n(P_X)] \le 1$.

For large n and small ϵ, the intuition for these results is as follows. The first results states that the probability of typical sequences is concentrated tightly around $2^{-nH(X)}$. The second result says that there are approximately $2^{nH(X)}$ sequences in the typical set $T_\epsilon^n(P_X)$ and the third result states that with high probability any sequence emitted by the DMS is typical.

example



PROBLEM:

We need the entropy!

How to estimate entropy ? or a Prob. distribution?

- Given a finite set of observations can we estimate the entropy of a source?

in the Shannon entropy [1]

$$H = -\sum_{i=1}^{M} p_i \ln p_i,$$

with the choice $\hat{p}_i = \frac{n_i}{N}$, the naive estimate

$$\hat{H} = -\sum_{i=1}^{M} \hat{p}_i \ln \hat{p}_i, \quad (2)$$

leads to a systematic underestimation of the entropy H.

Many papers study this topic, especially in Neuro science.

Ref:

Estimation of Entropy and Mutual Information

Liam Paninski
liam@cns.nyu.edu
Center for Neural Science, New York University, New York, NY 10003, U.S.A.

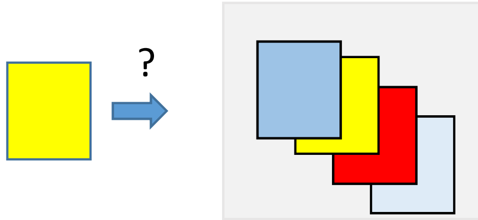
Estimation of the entropy based on its polynomial representation, Phys. Rev. E 85, 051139 (2012) [9 pages], Martin Vinck, Francesco P. Battaglia, Vladimir B. Balakirsky, A. J. Han Vinck, and Cyriel M. A. Pennartz

Information retrieval

Checking properties: questions

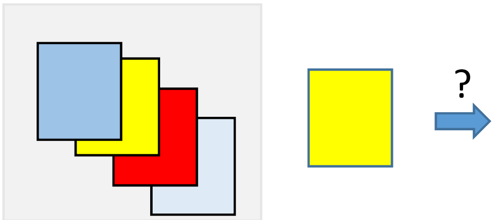
- **Do you have a particular property? (≈ identification)**

example: is yellow a property ?
=> search in the data base



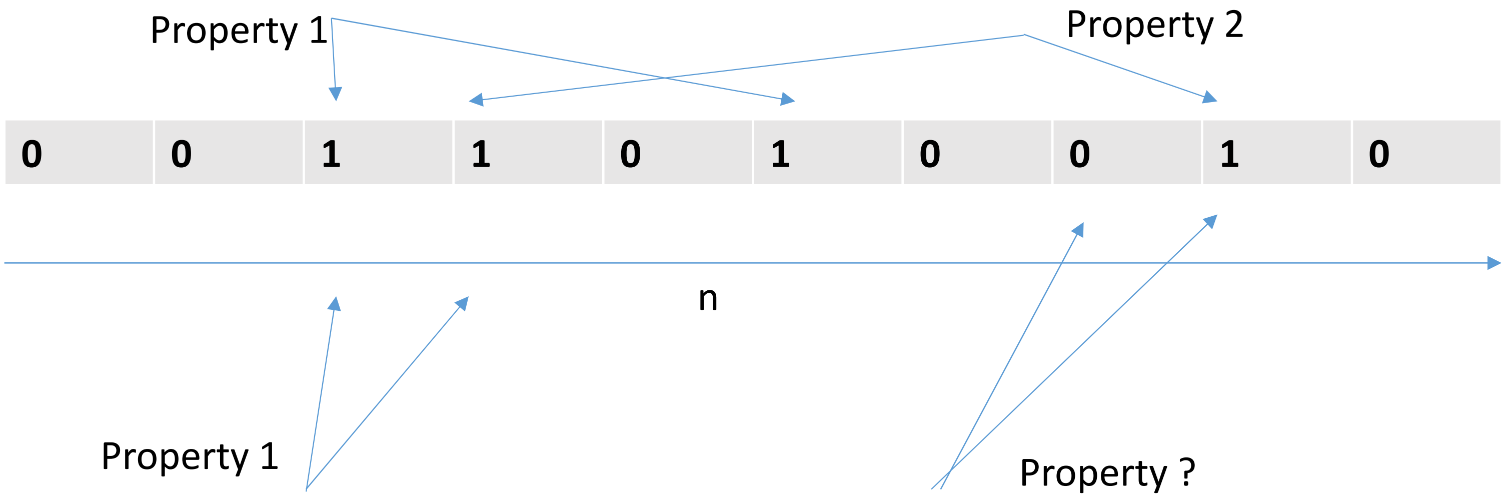
- **Is this a valid property? (≈ authentication)**

example: is yellow a valid property?
=> search in the property list



test for validity of a property can be done using the Bloom filter

- T properties, every property to k ‚1‘ s in random positions in a n array



- Check property: check the map (k positions) of a property in the n array

Performance: P(false accepted) = $\{(1 – (1-1/n)^{kT}\}^k$ => 2^{-k}, for k = n/T *ln* 2

Bloom (1970), quote.

The same idea appeared as “superimposed codes,” at Bell Labs, which I left in 1969.

Nonrandom Binary Superimposed Codes

W. H. KAUTZ, MEMBER, IEEE, AND R. C. SINGLETON, SENIOR MEMBER, IEEE

every sum of up to T different code words logically includes no code word other than those used to form the sum (Problem 2).

Superimposed codes: check presence of a property

- Start with N x n array, every property corresponds to a row. Every row pn ,1's

1	0	1	1	0
2	1	0	0	1
3	0	1	1	1
…				
N	1	0	0	0

→ n

Property: the OR of any subset of size T does not cover any other row

Signature or descriptor list: the OR of ≤ T rows

Check for a particular property: property covered by the signature?

Example:

1 0 0 1 0 1 1
1 0 1 0 0 1 0 not covered, not included in the OR
1 0 0 1 0 1 0 covered, included in the OR

Code existence: Probability(a random vector is covered by T others) => 0 for p = ln2/T (same as before) and since we have a specific code, n > TlogN

example

Superimposed Code-Based Indexing Method for Extracting MCTs

Wenxin Liang[1,4], Takeshi Miki[2], and Haruo Yokota[3,4]

Table 1. Examples of file signatures

F_1 Keyword	F_1 Signature	F_2 Keyword	F_2 Signature
Lear	1000001	Hamlet	0100001
King	0100010	King	0100010
Duke	0101000	Mother	0100100
Brother	1100000	Brother	1100000
File signature	1101011	File signature	1100111

Table 2. Examples of drops

Query keywords	Query signature	F_1	F_2
King, brother	1100010	actual drop	actual drop
King, mother	0100110	no match	actual drop
Lear, King	1100011	actual drop	false drop

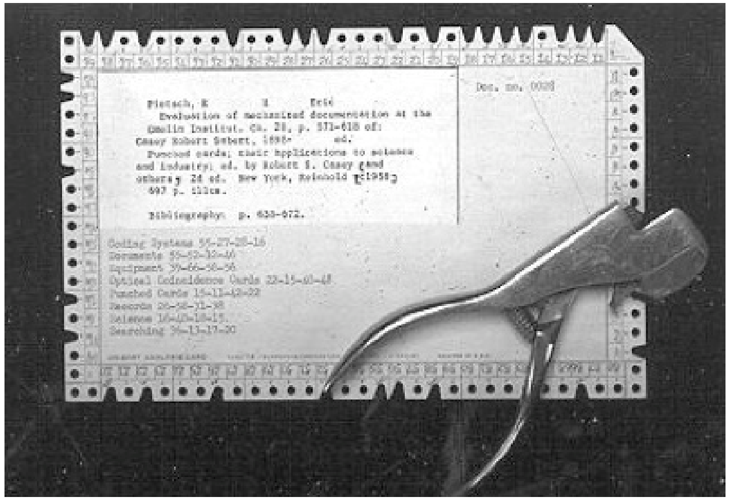
Code Example

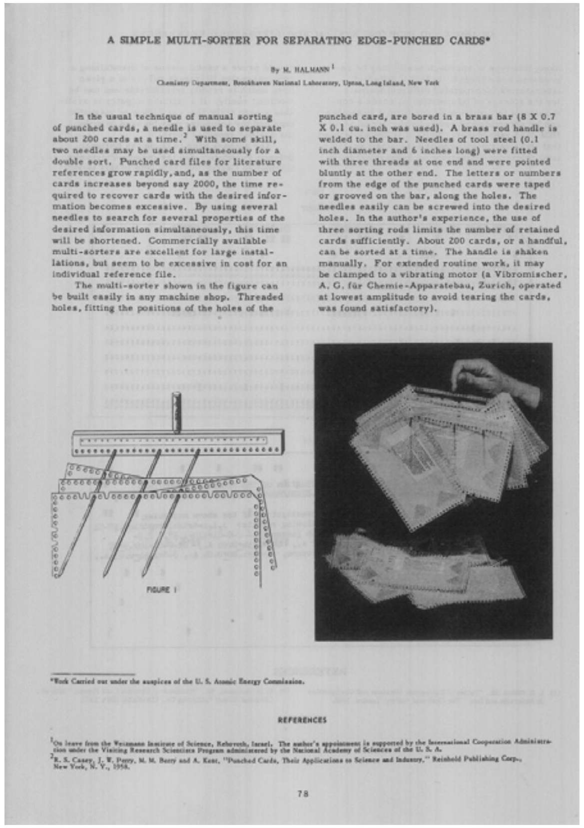
- BOUND: $T \log_2 N < n < 3 T^2 \log_2 N$

property	binary representation		
1	001	001	010
2	001	010	100
3	001	100	001
4	010	001	100
5	010	010	001
6	010	100	010
7	100	001	001
8	100	010	010
9	100	100	100
10	000	000	111
11	000	111	000
12	111	000	000

Any OR of **two** property vectors
does not overlap
with another property

How to retrieve information from a big set: Superimposed codes





A SIMPLE MULTI-SORTER FOR SEPARATING EDGE-PUNCHED CARDS*

By M. HALMANN[1]

Chemistry Department, Brookhaven National Laboratory, Upton, Long Island, New York

In the usual technique of manual sorting of punched cards, a needle is used to separate about 200 cards at a time.[2] With some skill, two needles may be used simultaneously for a double sort. Punched card files for literature references grow rapidly, and, as the number of cards increases beyond say 2000, the time required to recover cards with the desired information becomes excessive. By using several needles to search for several properties of the desired information simultaneously, this time will be shortened. Commercially available multi-sorters are excellent for large installations, but seem to be excessive in cost for an individual reference file.

The multi-sorter shown in the figure can be built easily in any machine shop. Threaded holes, fitting the positions of the holes of the punched card, are bored in a brass bar (8 X 0.7 X 0.1 cu. inch was used). A brass rod handle is welded to the bar. Needles of tool steel (0.1 inch diameter and 6 inches long) were fitted with three threads at one end and were pointed bluntly at the other end. The letters or numbers from the edge of the punched cards were taped or grooved on the bar, along the holes. The needles easily can be screwed into the desired holes. In the author's experience, the use of three sorting rods limits the number of retained cards sufficiently. About 200 cards, or a handful, can be sorted at a time. The handle is shaken manually. For extended routine work, it may be clamped to a vibrating motor (a Vibromischer, A. G. für Chemie-Apparatebau, Zurich, operated at lowest amplitude to avoid tearing the cards, was found satisfactory).

FIGURE 1

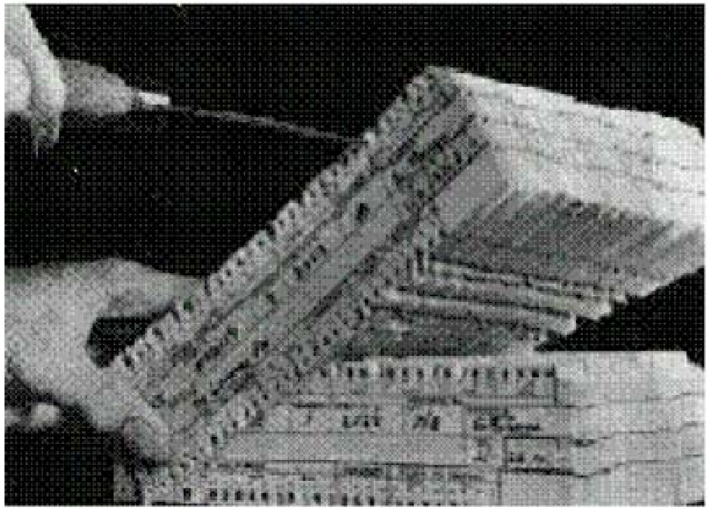
*Work Carried out under the auspices of the U. S. Atomic Energy Commission.

REFERENCES

[1] On leave from the Weizmann Institute of Science, Rehovoth, Israel. The author's appointment is supported by the International Cooperation Administration under the Visiting Research Scientists Program administered by the National Academy of Sciences of the U. S. A.

[2] R. S. Casey, J. W. Perry, M. M. Berry and A. Kent, "Punched Cards, Their Applications to Science and Industry." Reinhold Publishing Corp., New York, N. Y., 1958.

78



We need associative memory!

Nonrandom Binary Superimposed Codes

W. H. KAUTZ, MEMBER, IEEE, AND R. C. SINGLETON, SENIOR MEMBER, IEEE

a given small positive integer m, every sum of up to m different code words is distinct from every other sum of m or fewer code words (Problem 1), or logically in-

More general, take distinct for 1, 2, …, m

references

- Arkadii G. D'yachkov

- W.H. Kautz

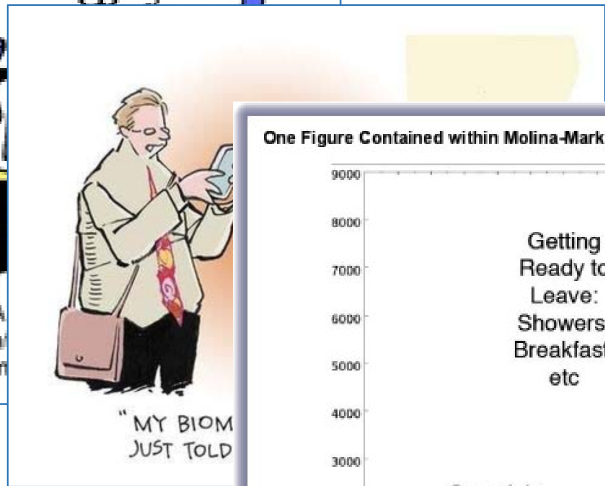
- CALVIN N. MOOERS, (1956) "ZATOCODING AND DEVELOPMENTS IN INFORMATION RETRIEVAL", Aslib Proceedings, Vol. 8 Iss: 1, pp.3 - 22

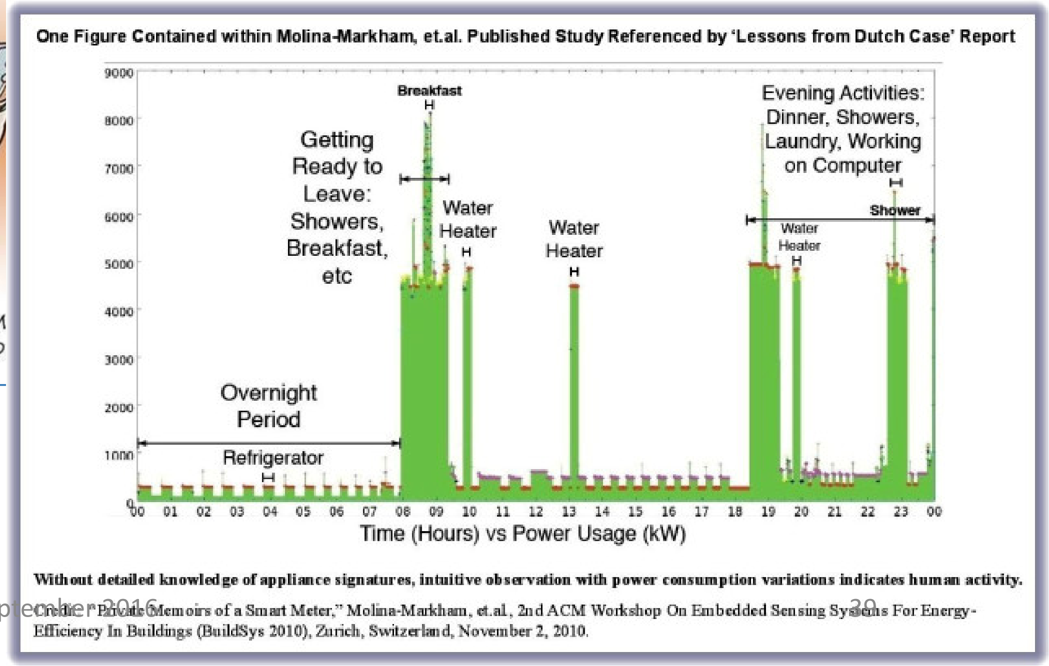
- My own:ON *SUPERIMPOSED CODES* A.J. Han *Vinck* and Samuel Martirossian

  in **Numbers, Information and Complexity**

  editors: Ingo Althöfer, Ning Cai, Gunter Dueck - 2013 - Technology & Engineering

Security and Privacy concerns for big data

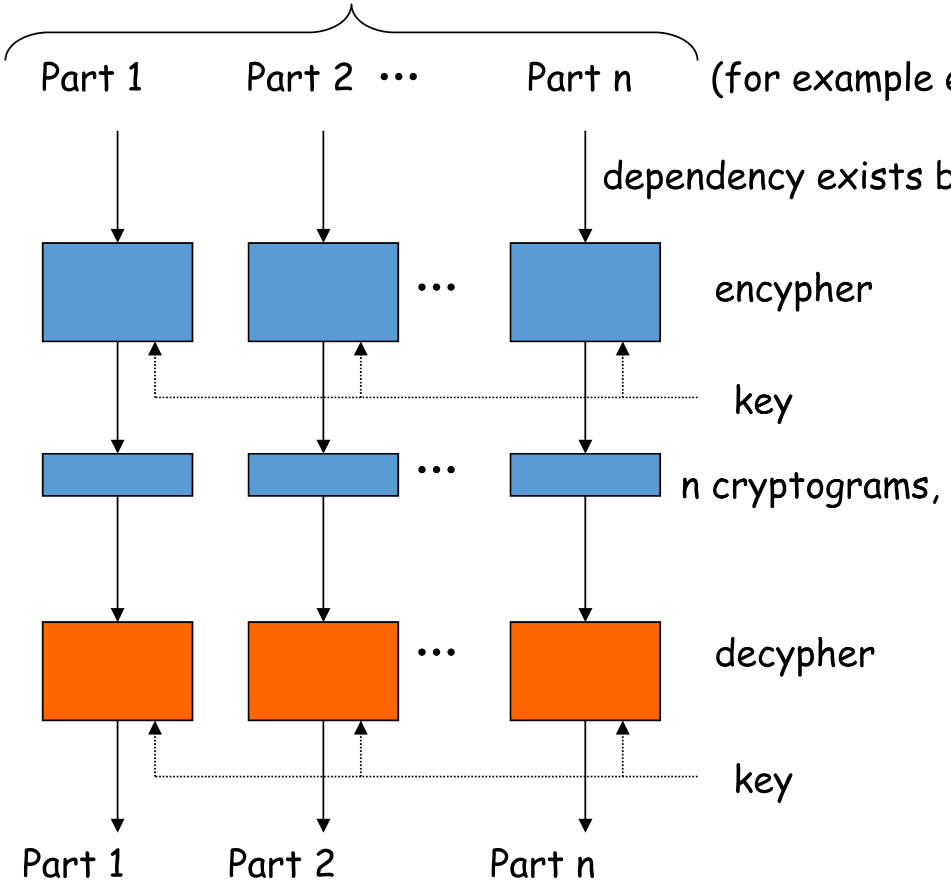




Problems:

- Data privacy
- Data protection/security

Message encryption without source coding



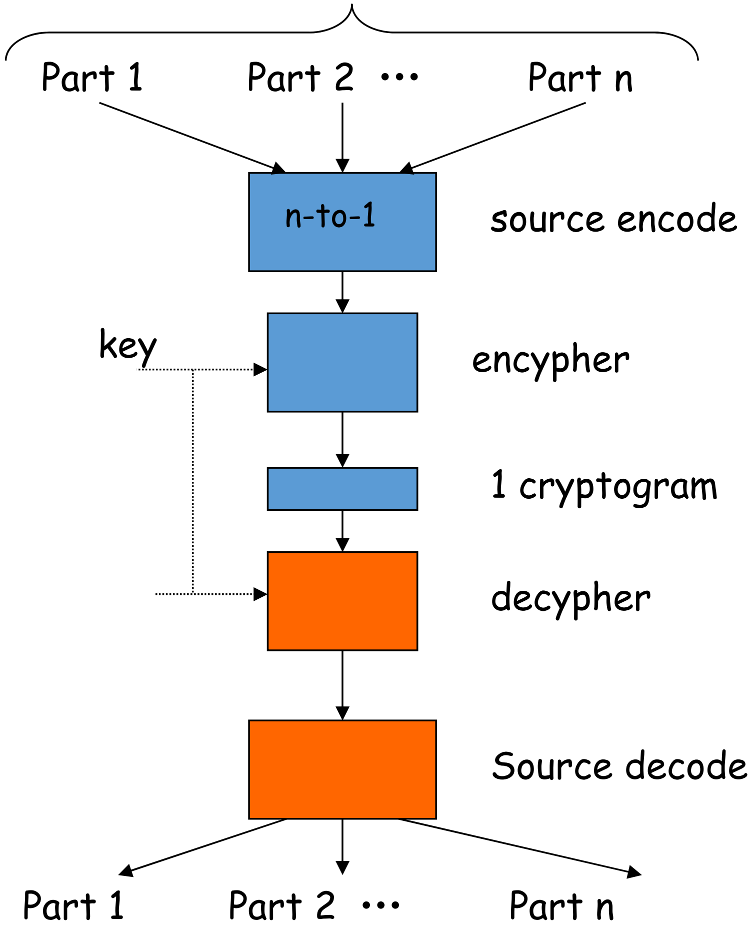
<u>dependency exists between cryptograms</u>

<u>Attacker:</u>

n cryptograms to analyze for particular message of n parts

Message encryption with source coding

(for example every part 56 bits)



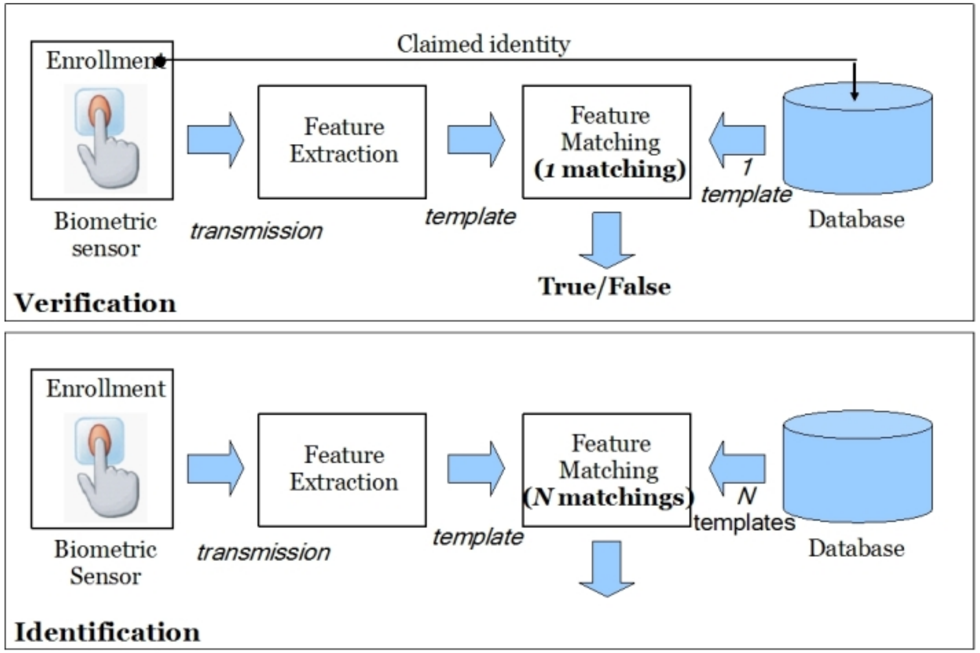
Attacker:

- 1 cryptogram to analyze for particular message of n parts
- assume data compression factor n-to-1

Hence, less material for the same message!

The biometric identification/authentication problem

1. Conversion to binary

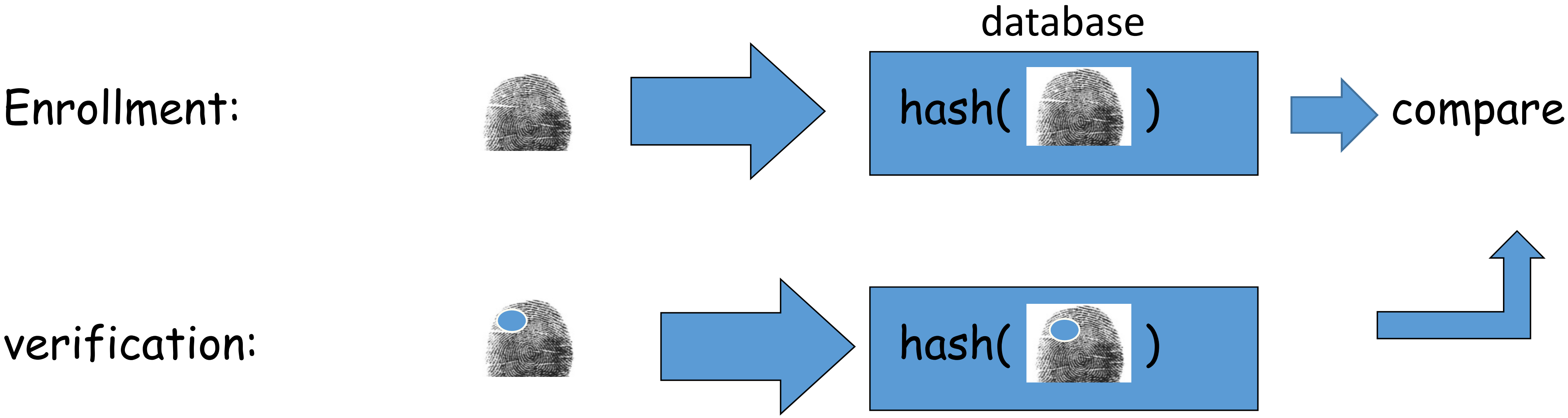


4. variations?

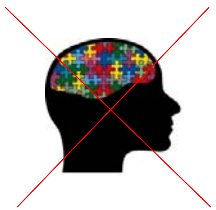
3. Privacy

2. Complex searching f(N)

Illustration of the authentication problem using biometrics

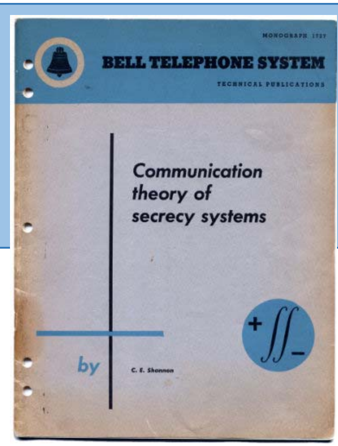


PROBLEM: BIO differs and thus also the hash!



Advantage no memorization

Information theory can help to solve the security/privacy problem



"transformed cryptography from an art to a science."

PERSPECTIVE OF SHANNON'S SECRECY SYSTEM

secret

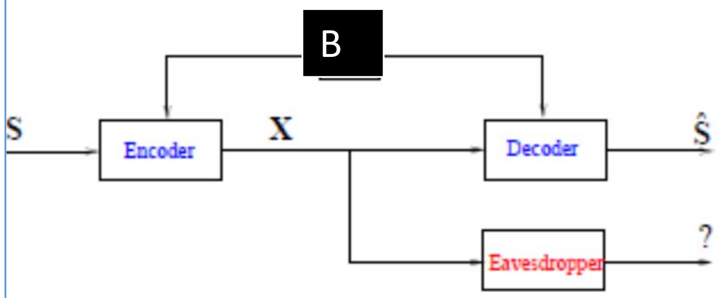


Figure 1: Shannon's secrecy system.

For Perfect secrecy we have a necessary condition:

$$H(S|X) = H(S)$$

$$\Rightarrow H(S) \le H(B)$$

i.e. # of messages ≤ # of keys

Shannons noisy key model

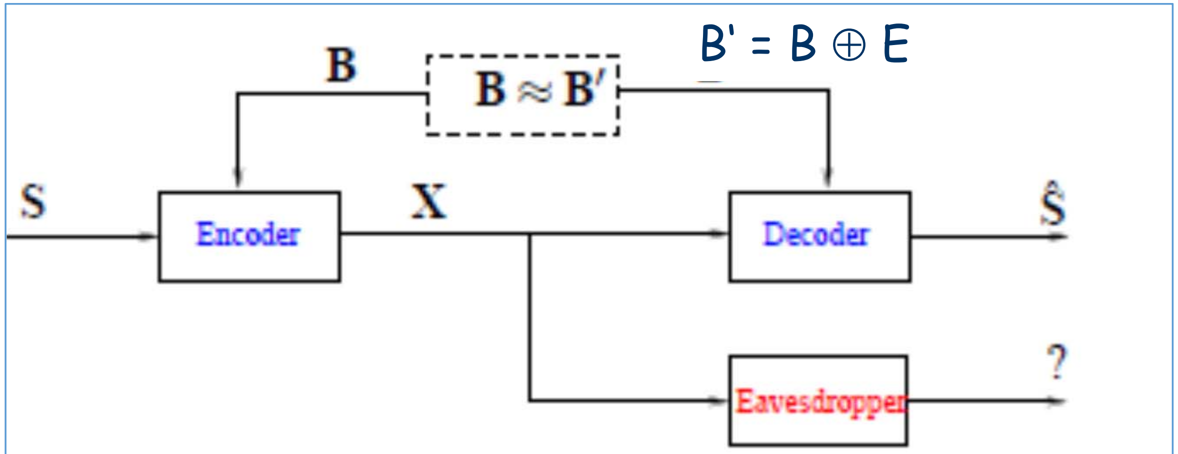


Figure 2: An extension of Shannon's secrecy system.

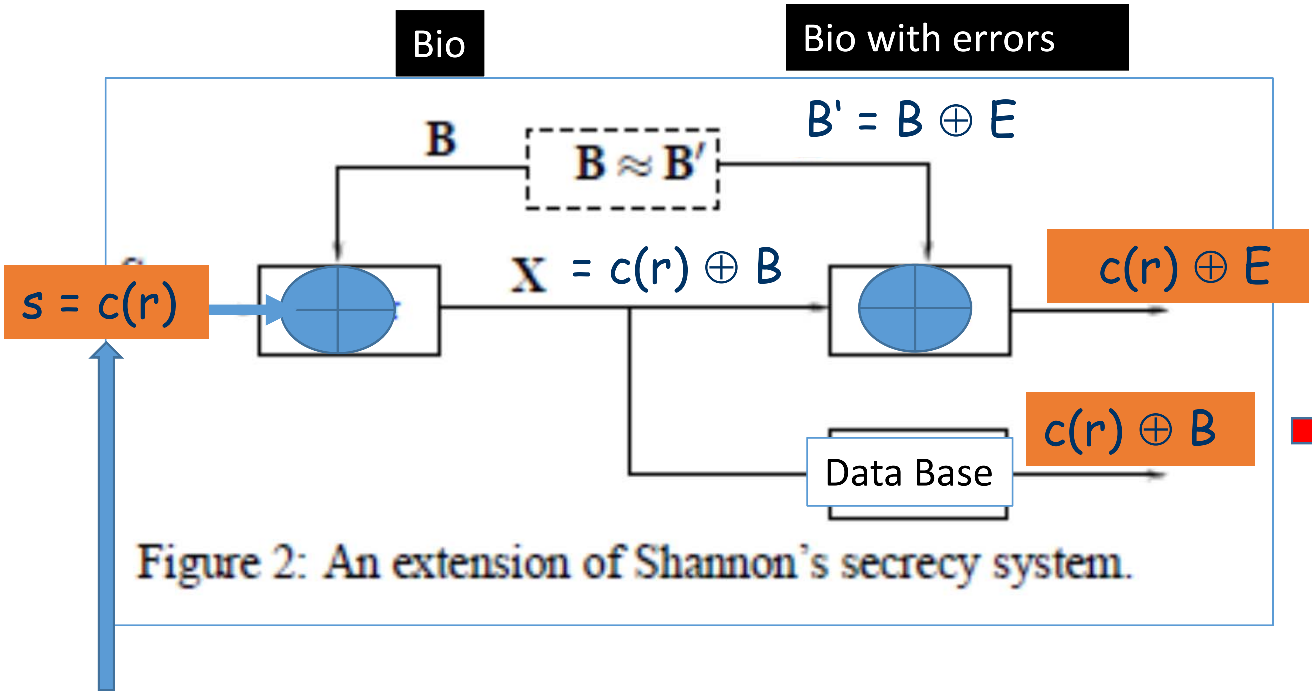
For Perfect secrecy $H(S|X) = H(S)$

=> $H(S) \leq H(B) - H(E)$

i.e. we pay a price for the noise!

Shannons noisy key model used for biometrics

Ari Juels



Decode r => E => B ⊕ E ⊕ E = B

E, r

B

K ≤ H(B) – H(E)

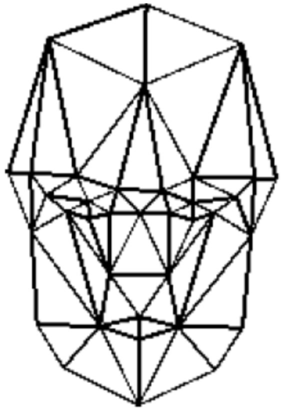
Limit on
- error correcting capability
- and privacy

Correct guess => 2^{-k}
Larger k less errors corrected, more privacy
Smaller k more errors corrected, less privacy

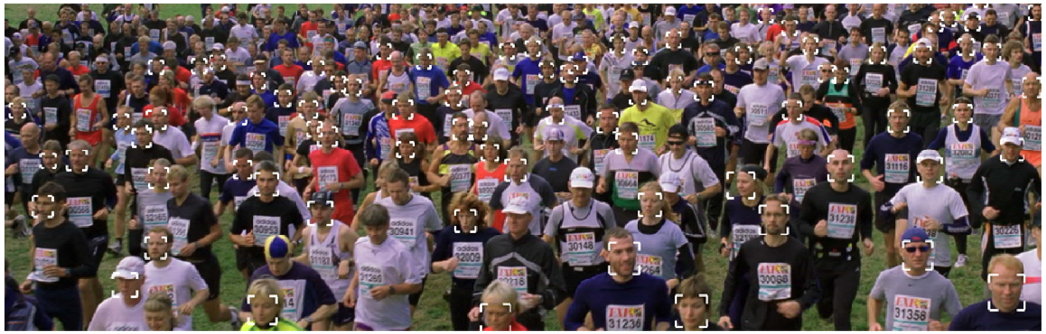
„Random“
linear codeword with k „info“ symbols

Biometrics challenge: get biometric features into binary

protection



identification



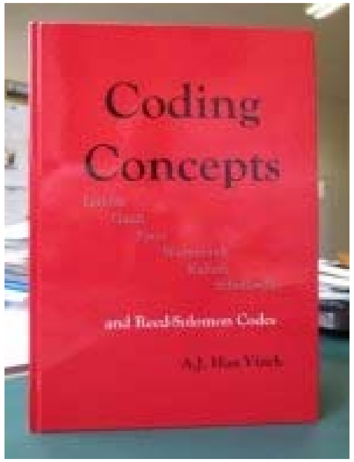
Examples where information theory helps to solve problems in big data

- data compression/reduction with/without distortion
- data quality using error correction codes
- data protection: cryptographic appproach
- outlier/anomaly/classification
- information retrieval

In theory, there is no difference between theory and practice. But in practice, there is.
Yogi Berra

The end

My website: https://www.uni-due.de/dc/



My recent (2013) book with some of my research results (free Download)

https://www.uni-due.de/imperia/md/images/dc/book_coding_concepts_and_reed_solomon_codes.pdf

PRINCETON UNIVERSITY

Motivation: Data Security

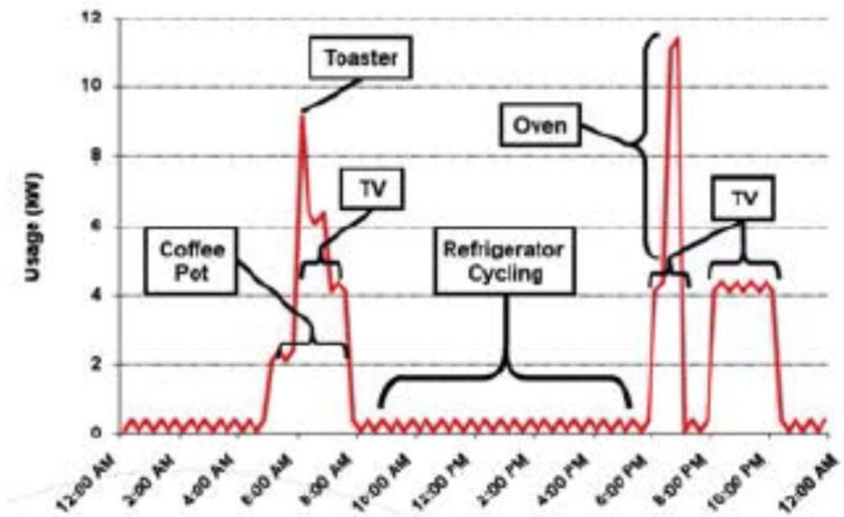
- The smart grid **cyber layer** will generate considerable **electronic data**:
  - Power flow sensors, phasor measurement units, smart meters, etc.

- The **utility** of this data depend on its accessibility.
- But, it can also **leak information that should be** kept secure, or **private**.
- How can we **characterize** this **fundamental tradeoff**?

PRINCETON UNIVERSITY

Ex. 1: Smart Meter Privacy

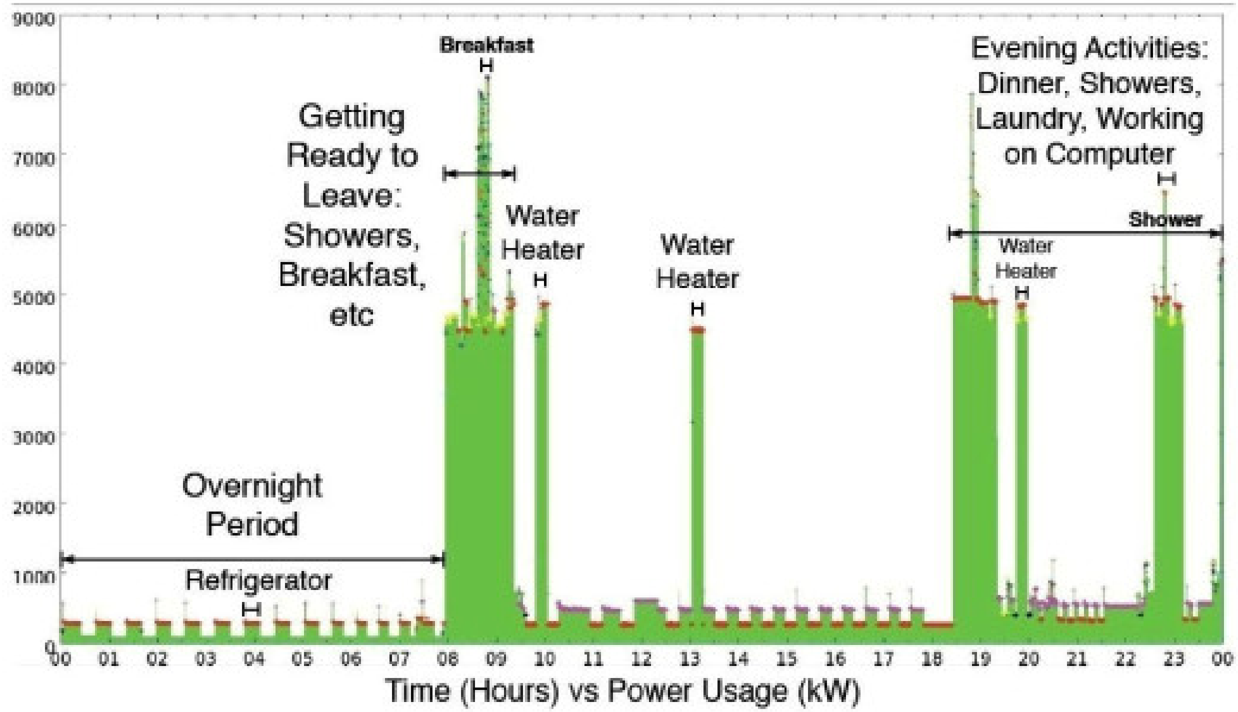
- Smart meter **data** is useful for **price-aware usage**, **load balancing**
- But, it **leaks information** about in-home activity



Privacy?

Dutch anti-smart meter logo; roughly equivalent to "Smart Metering Equals Spying"

One Figure Contained within Molina-Markham, et.al. Published Study Referenced by 'Lessons from Dutch Case' Report



Without detailed knowledge of appliance signatures, intuitive observation with power consumption variations indicates human activity.

Credit: "Private Memoirs of a Smart Meter." Molina-Markham, et.al. 2nd ACM Workshop On Embedded Sensing Systems For Energy-

Raising Public Awareness to Smart Grid, Smart Meter, and Radiofrequency (RF) Issues: Privacy, Health, Cybersecurity, Safety, Economics, Societal Impacts, Environmental Impacts, Consumer Choice and Rights

references

- http://nlp.stanford.edu/IR-book/newslides.html

Information theory: channel coding theorem (1)

- for a binary code with words of length n, and rate (efficiency) R = k/n
  the number of code words = 2^k

To achieve the Shannon Channel Capacity and Pe => 0, n => infinity
an thus also k => infinity

<u>Hence:</u>

coding problem (# of code words = 2^k how to encode!)
and also decoding problem!

Topics we can work on based on past performance

- Information theoretical principles for anomaly detection

- Biometrics and big data

- Memory systems and big data

- Privacy in smart grid

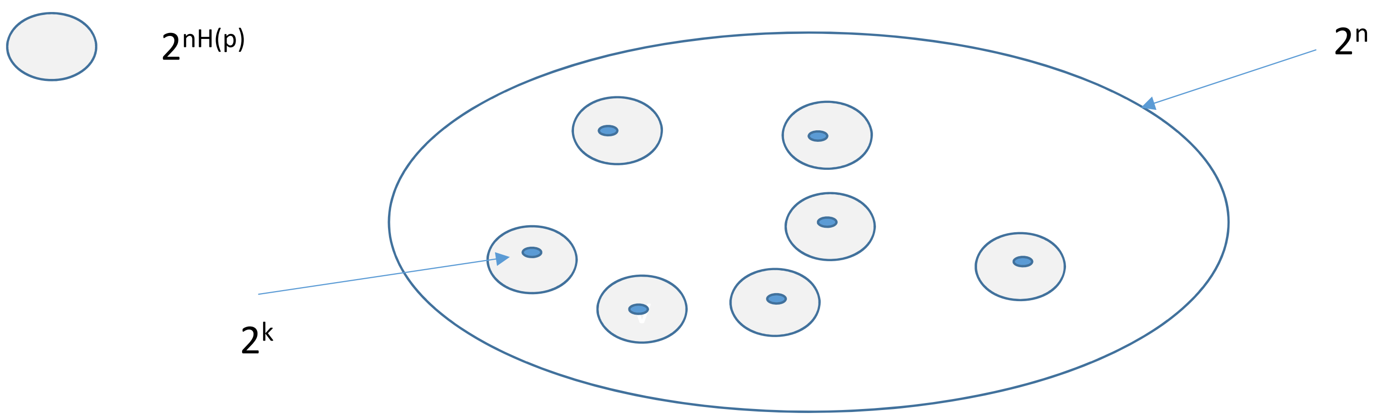
- Information retrieval and superimposed codes

Use error correcting code for noiseless source coding

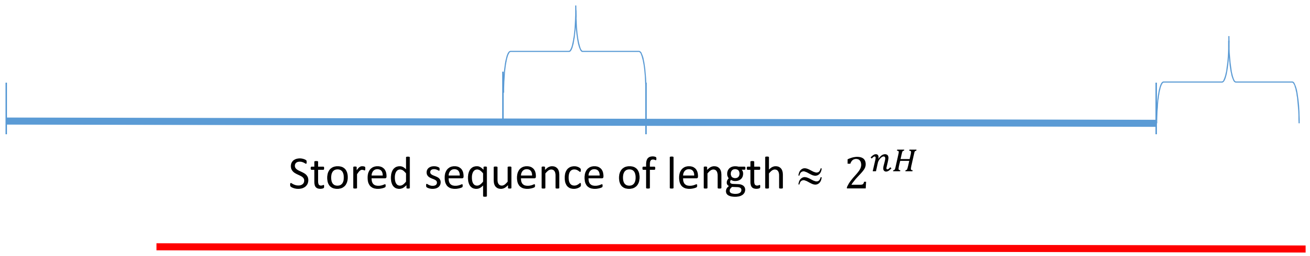
- 2^k code words of length n; Correct $2^{nH(p)}$ noise vectors

where

$$2^k \times 2^{nH(p)} = 2^n \quad \text{or} \quad k/n = 1 - H(p) \quad \text{(at capacity)}$$

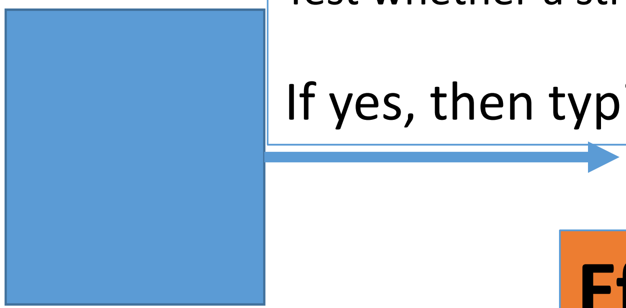


An obvious algorithm (like Lempel and Ziv)



Test whether a string of length n is in the STORED sequence somewhere

If yes, then typical   If not, then a-typical



Efficiency: => the entropy H bits/symbol

Since the probability of a typical sequence is≈ 2^{nH} we expect all typical sequences in the stored sequence

Uniquely decipherable codes

decipherable code:

the OR of $\leq$ T binary vectors of length M is unique

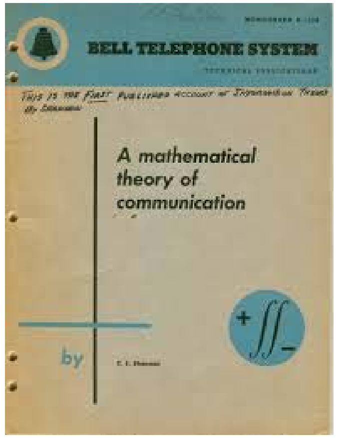
ANSWERS: WHO?

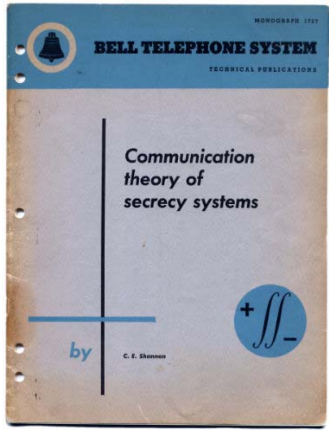
Condition on M $\quad (2^M - 1) \geq \sum_{i=1}^{T} \binom{N}{i}$

$\downarrow \qquad \downarrow$

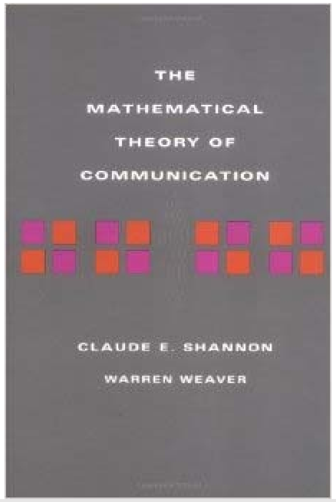
$M \geq T \log_2 N$

Some pictures



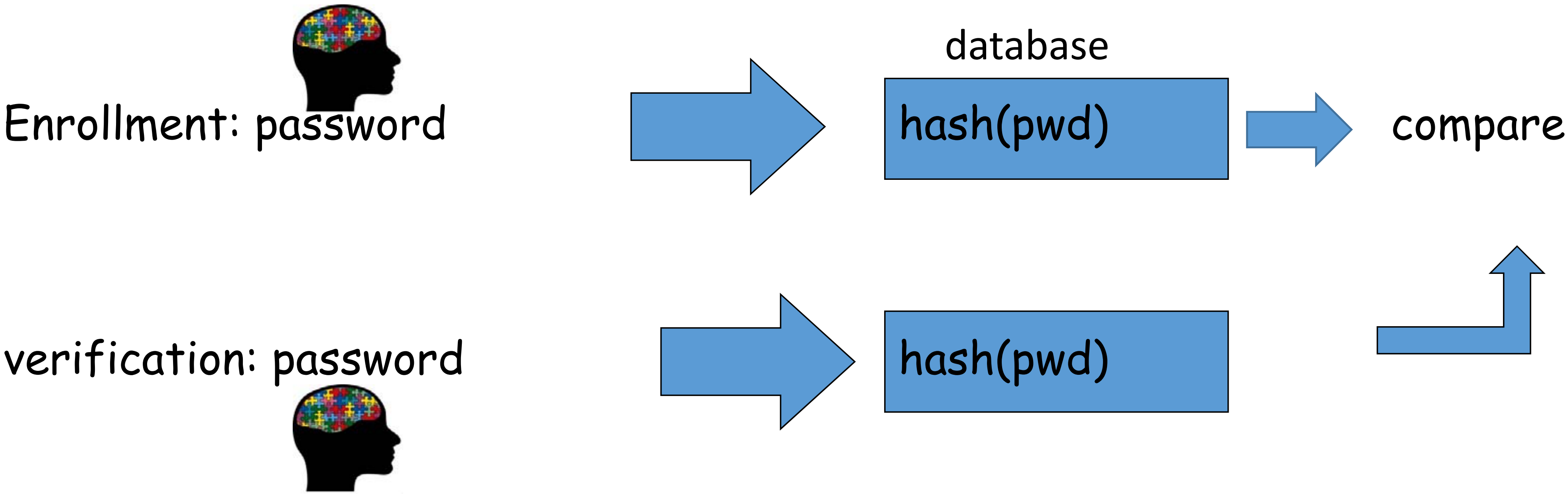


"transformed cryptography from an art to a science."

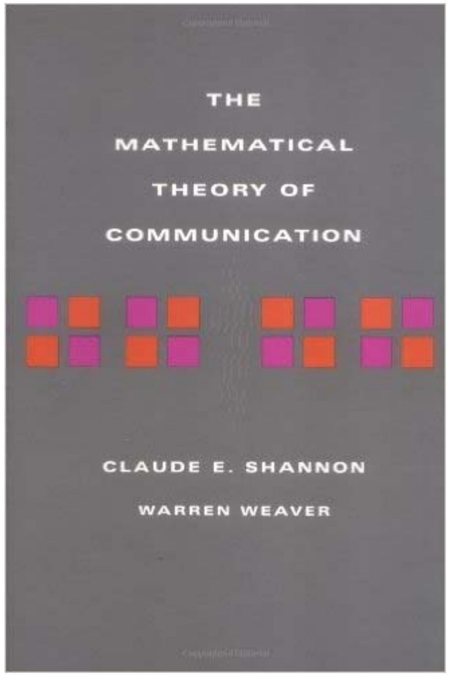


The book co-authored with Warren Weaver, *The Mathematical Theory of Communication*, reprints Shannon's 1948 article and Weaver's popularization of it, which is accessible to the non-specialist.[5] In short, Weaver reprinted Shannon's two-part paper, wrote a 28 page introduction for a 144 pages book and changed the title from "A mathematical theory..." to "The mathematical theory..."

Illustration of the authentication problem using a memorized password

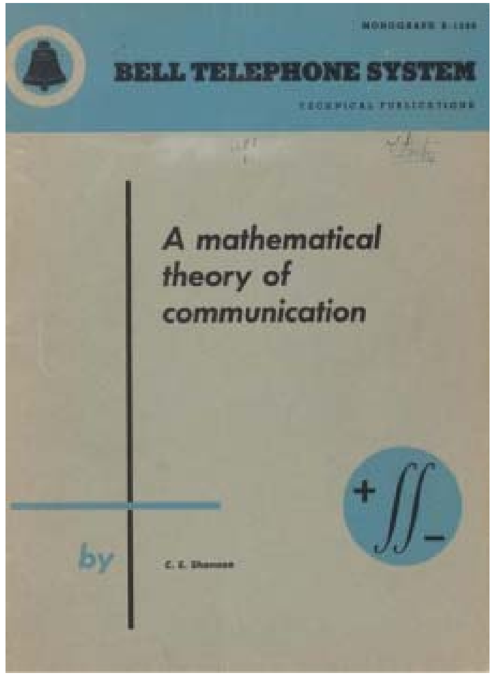


We use information and communication theory



Communication Theory of Secrecy Systems*

By C. E. SHANNON



PERSPECTIVE OF SHANNON'S SECRECY SYSTEM

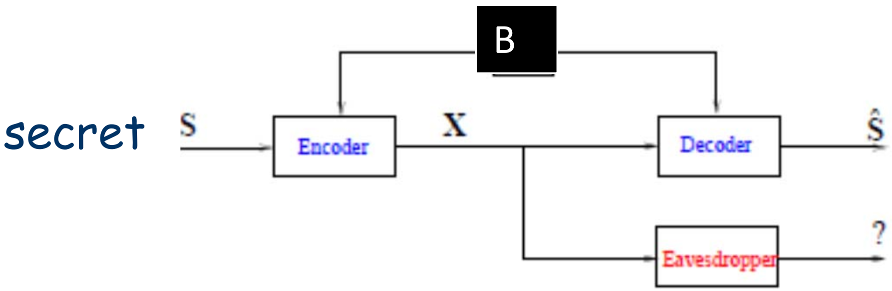


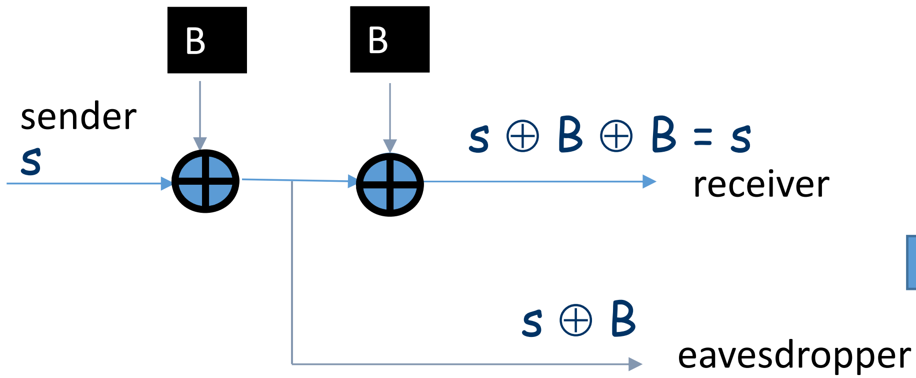
Figure 1: Shannon's secrecy system.

For Perfect secrecy we have a necessary condition:

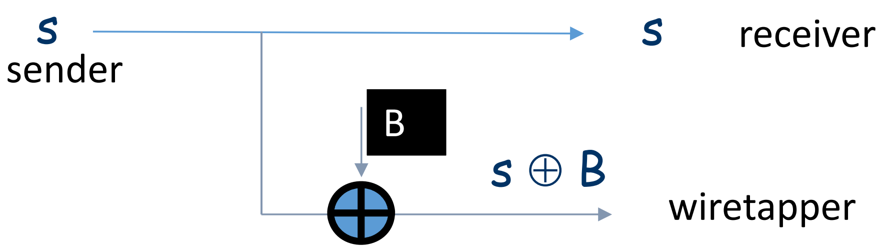
$$H(S|X) = H(S)$$

$$\Rightarrow H(S) \le H(B)$$

i.e. # of messages $\le$ # of keys



Wiretap channel model



Secrecy rate: C_s = H(B) = amount of secret bits/tr

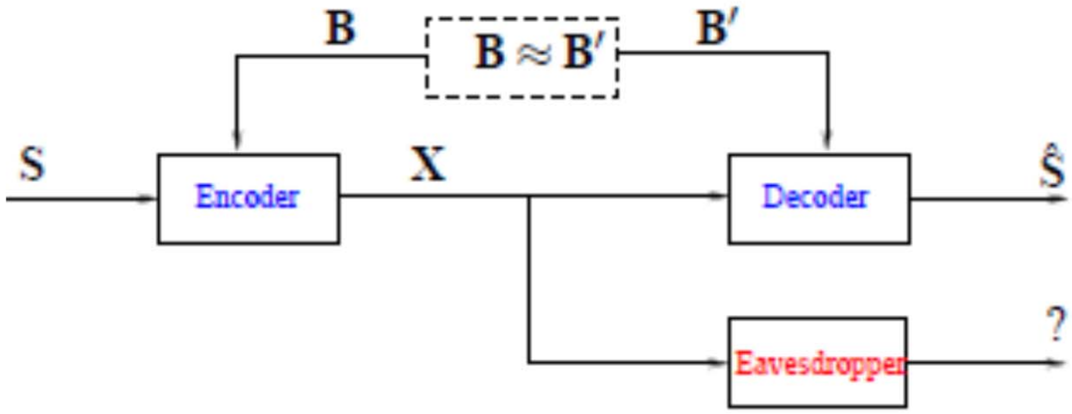


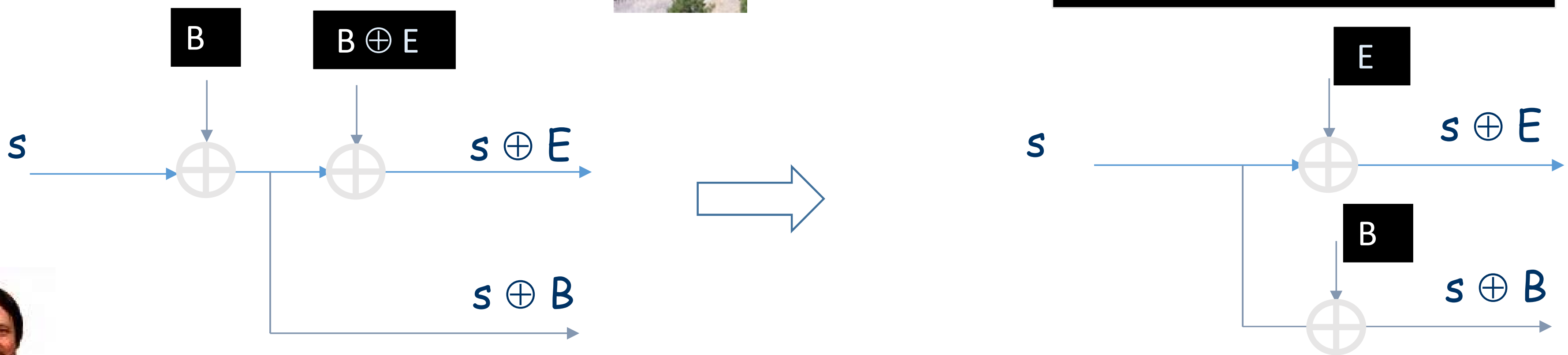
Figure 2: An extension of Shannon's secrecy system.

For Perfect secrecy $H(S|X) = H(S)$

$$H(S) \leq H(B) - H(E)$$

i.e. we pay a price for the noise!

[redacted]



Aaron Wyner

Solution given by the Juels Wattenberg scheme: USING BINARY CODES

PERSPECTIVE OF SHANNON'S SECRECY SYSTEM

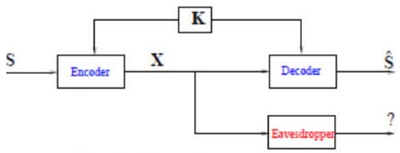
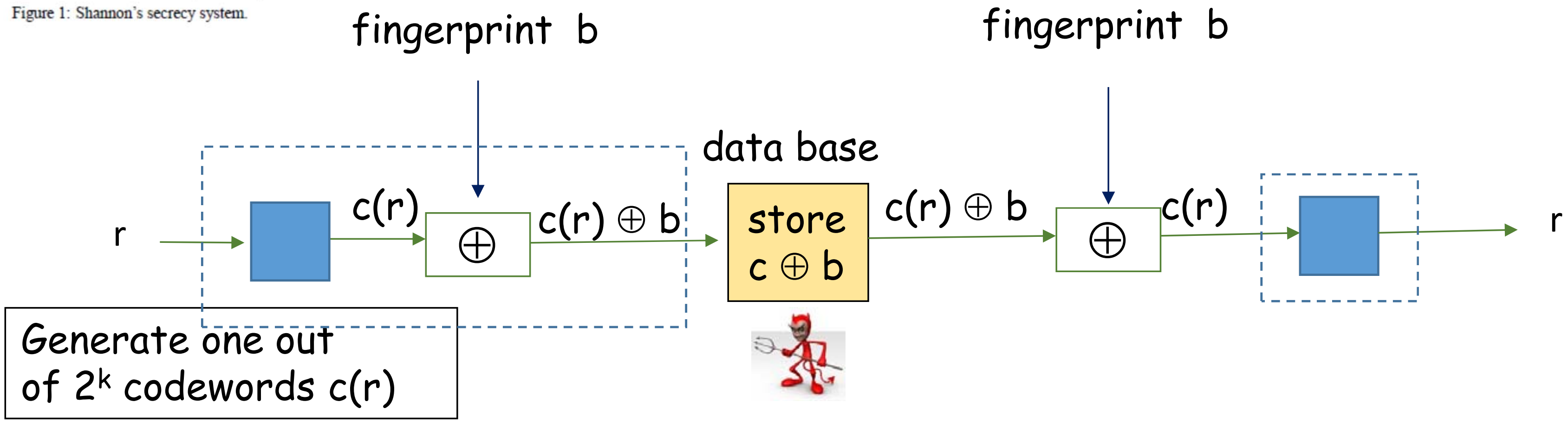


Figure 1: Shannon's secrecy system.



<u>Condition:</u> given c(r) ⊕ b it is hard to estimate b or c(r)

Guess: one out of 2^k codewords

safe storage: how to deal with noisy fingerprints ?

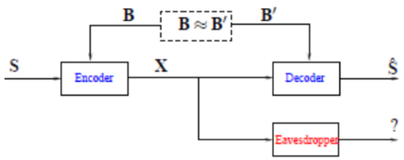
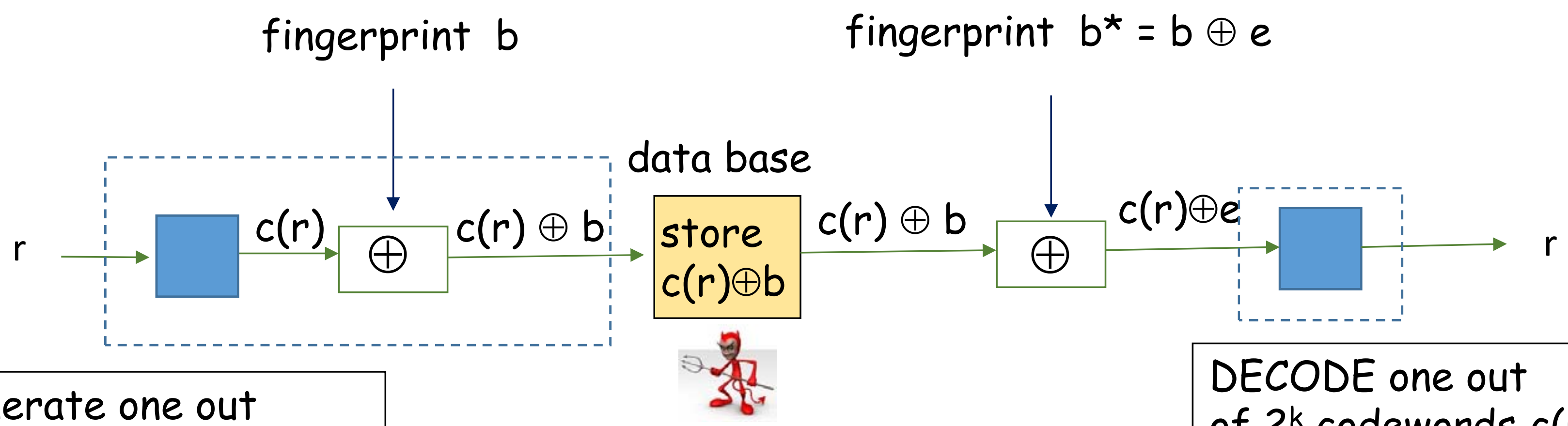


Figure 2: An extension of Shannon's secrecy system.



Generate one out of 2^k codewords c(r)

DECODE one out of 2^k codewords c(r) => r

<u>Condition:</u> given c(r) ⊕ b it is hard to estimate b or c(r)

Guess: one out of 2^k codewords

reconstruction of original fingerprint

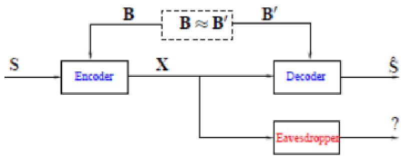
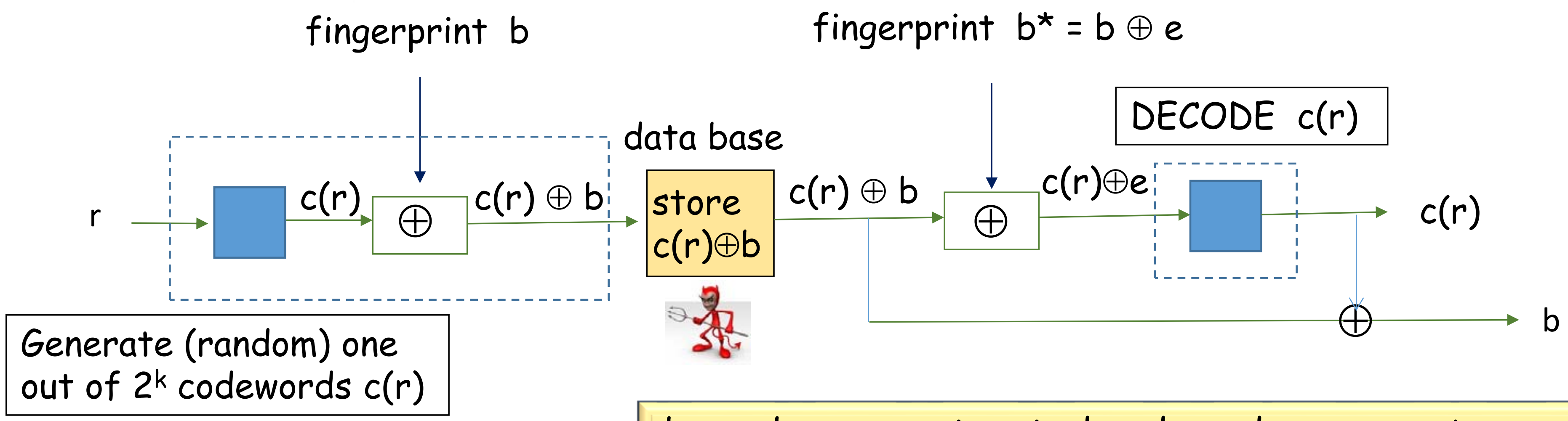
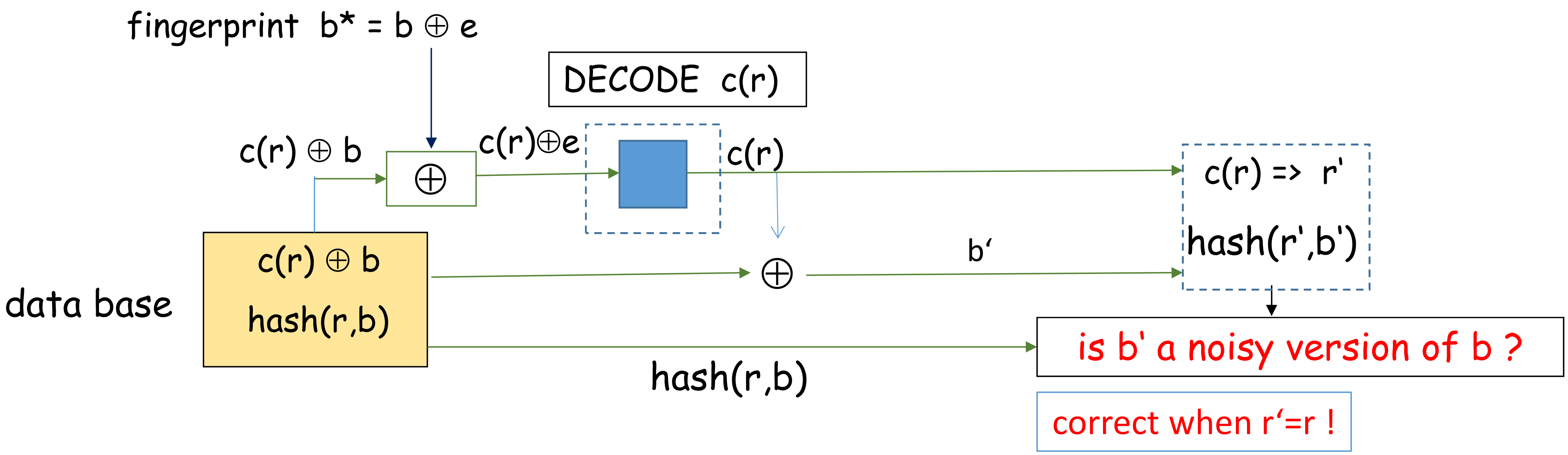


Figure 2: An extension of Shannon's secrecy system.



b can be reconstructed and used as correct password

authentication, how to check the result?



False Rejection Rate (FRR) : valid b' rejected;

False Acceptance Rate (FAR) : invalid b' accepted;

Successful Attack Rate (SAR): correct guess c, construct b from c ⊕ b

PERFORMANCE DEPENDS on the CODE! Small k gives good error protection

Entropy, mutual information

- $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$

How can we reduce the amount of data (1)

- Represent every possible source output of length n by a „binary“ vector of length m.

  - **Noiseless**: exact representation costly (depends on source variability!)
    - Need a good algorithm (non exponential in the blocklength n)

  - **Noisy**: good memory reduction, but in general we loose the details
    - how many bits do we need for a particular distortion
    - Need to define the distortion properly!

NOTE: We are interested in the NOISE!

How can we reduce the amount of data? (2)

- Assign - log p(x) bits to a message => likely, small => unlikely, large
  - Shannon showed how to do this

then, **the minimum obtainable average assigned length** is

$$H(X) = -\sum p(x) \log p(x) \qquad \text{(SHANNON ENTOPY)}$$

- Suppose that we use another assignment – log q(x)

  - The difference (DIVERGENCE) in average length is

$$D(P|Q) := -\sum p(x) \log p(x) - -\sum p(x) \log q(x) \quad \geq 0!$$

What do we need?

- Good knowledge of the structure of the data for
  - Good prediction
  - High compression rate
  - Variability for non-stationary data statistics

Anomaly: Normal or abnormal

- We need to develope decision mechanisms!

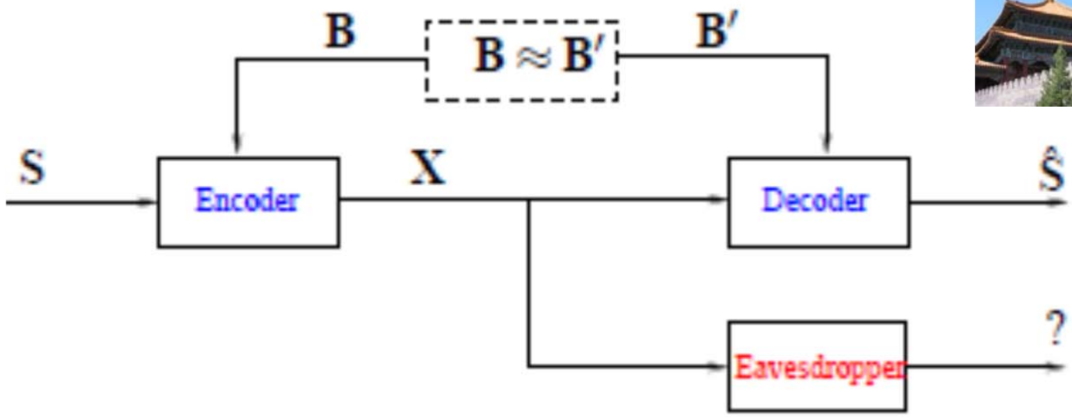
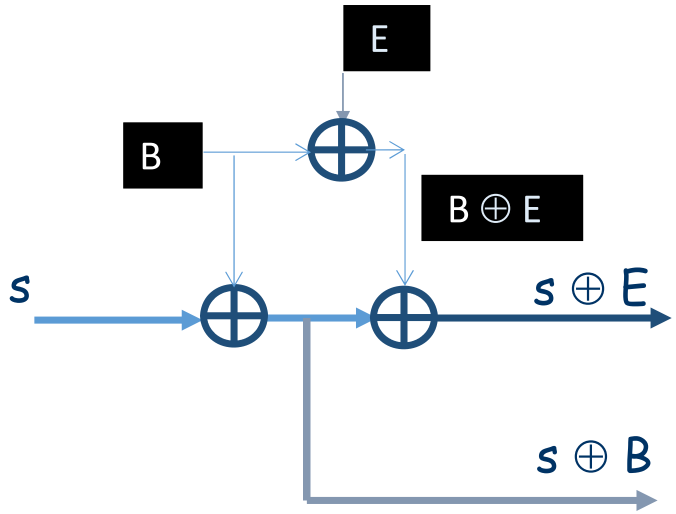


Figure 2: An extension of Shannon's secrecy system.

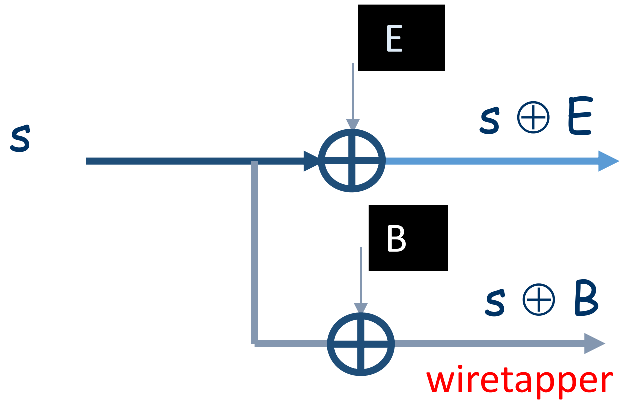
For Perfect secrecy $H(S|X) = H(S)$

$$H(S) \leq H(B) - H(E)$$

i.e. we pay a price for the noise!



Wiretap channel model



Aaron Wyner

Secrecy rate C_s = $H(B)$ - $H(E)$ = # secret bits/transmission

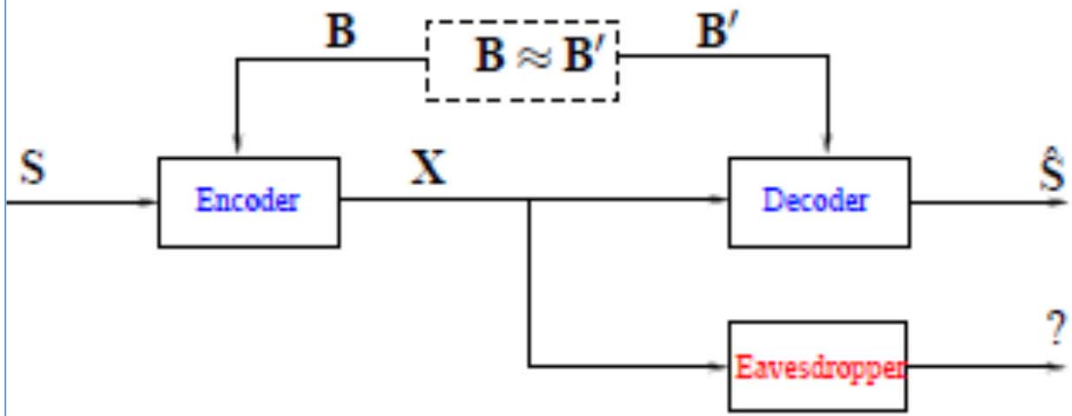
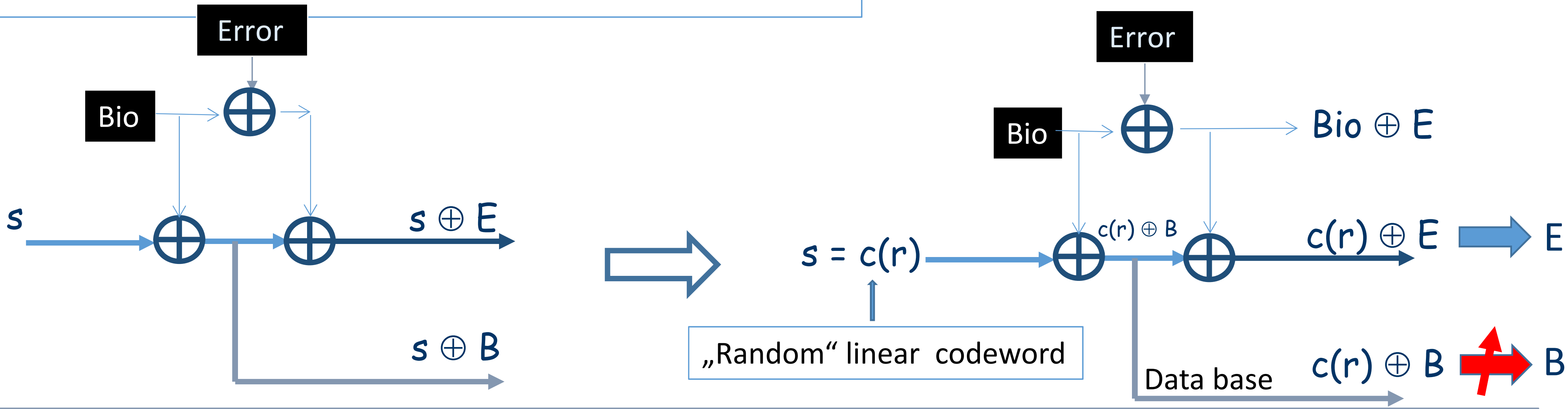


Figure 2: An extension of Shannon's secrecy system.

For Perfect secrecy $H(S|X) = H(S)$

$H(S) \leq H(B) - H(E)$

i.e. we pay a price for the noise!



Secrecy rate C_s = H(B) - H(E) = # secret bits/transmission