Cryptography

Systems and Information Security
Informatics Engineering (3rd year, 2nd sem.)

José Bacelar Almeida

Cryptography

Preamble

Basic Concepts

Applied Cryptography

Information Security

CIA triad

- **Confidentiality:** Preserving authorised restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information. A loss of confidentiality is the unauthorised disclosure of information.
- **Integrity:** Guarding against improper information modification or destruction, including ensuring information nonrepudiation and authenticity. A loss of integrity is the unauthorised modification or destruction of information.
- **Availability:** Ensuring timely and reliable access to and use of information. A loss of availability is the disruption of access to or use of information or an information system.

Cryptography & Information Security

- "*Cryptography is an important computer security tool that deals with techniques to store and transmit information in ways that prevent unauthorized access or interference*" [ISO]
- Cryptography serve as a crucial line of defence against various threats, including unauthorised access, data breaches, tampering, and eavesdropping.
- Cryptography is a **security mechanism**, used in combination with others to achieve the goals of information security.

Cryptography

Preamble

Basic Concepts

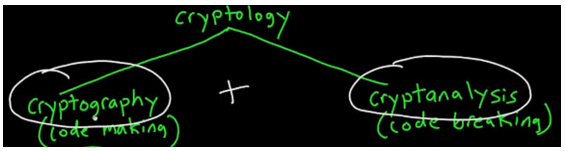
Applied Cryptography

Origins of the concept

- "Kryptós" + "Gráphein" (hidden writing)
- ...art of transmitting "secrets" (in an open channel)
- Used since ancient times, often associated with military and diplomatic activities.
- Today, the goals have expanded to encompass additional aspects of the communication process in hostile environments:
  - **Confidentiality**: secrecy of messages;
  - **Authenticity**: establish the origin of messages;
  - **Integrity**: transmitted message has not been tampered;
  - …
- Obs.: notice that the meanings of properties have been narrowed w.r.t. their use in the context of Information Security.

Cryptoanalysis

- In contrast, **cryptanalysis** aims to undermine the goals of cryptography.
- A cryptographic technique is deemed **broken** if its objectives have been successfully compromised through cryptanalysis.
- Cryptography and Cryptanalysis form an area sometimes called **Cryptology**.
- As a scientific discipline, cryptology touches on several different fields, such as probability, complexity theory, information theory, computer science, etc.



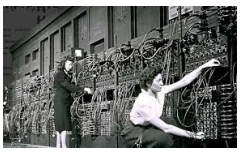
brief history

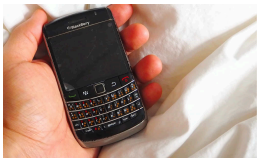
- The use of (proto) ciphers has been known since the 15th century BC.
- Armies of classical empires (e.g. Romans) regularly used cryptography

- The sophistication of the techniques has evolved over the centuries
- E.g. Enigma machine - WW2, Germany.
- Shortly after World War II, Claude Shannon (video) published two papers on Information Theory, which allowed for the formalisation of cipher security — considered the birth of Cryptography as a scientific field.

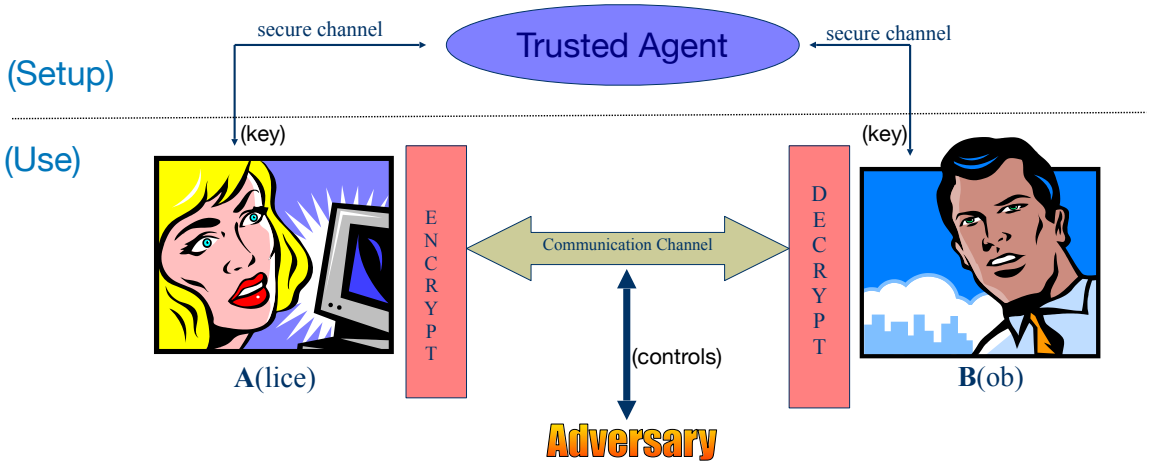
Modern Cryptography

- 1948/9 – Teoria da Informação (Claude Shannon).
- 1970/7 – Data Encryption Standard (DES).
- 1976 – Criptografia de Chave-Pública (Diffie & Hellmann)
- 2001 – Escolha do substituto do DES: Advanced Encryption Standard (AES).
- 2022/… – Criptografia Post-Quantum
- …



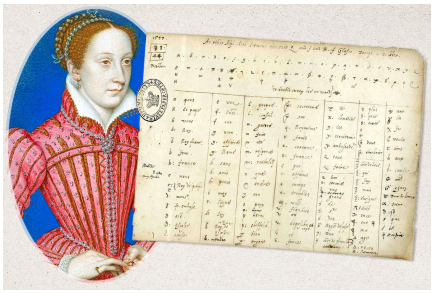

Model
(confidentiality)



Some more terminology

- **plaintext** (or **cleartext**): content of the message;
- **encryption**: operation that encodes the plaintext into an obscured message (**cyphertext**, or **cryptogram**);
- **decryption**: inverse operation of encryption
- **cryptographic system**: specification of the (possibly probabilistic) algorithms that perform some cryptographic technique (e.g. key-generation; encryption; decryption).
- **attack**: compromising the goals of a cryptographic technique (e.g. obtain the transmitted message without the corresponding key).
- **adversary** (intruder; enemy; …): hostile environment.

Security of Encryption



- Historically, security was largely attributed to the secrecy associated with the underlying technique
- ...which had disastrous consequences!
- The trend described was still present in the 20th century. However, as early as the 19th century, *Auguste Kerckhoff* established the following principle.:

  the security of a cipher must be assessed on the assumption that all the details of its construction is public knowledge

- Corollary: security can only be derived from a parameter that is explicitly secret - the **key**

Adversary

- Personalisation of the **hostile environment**
- …also known by the name of **enemy**; **spy**; **(E)ve**; **(M)allory**; …

Adversary success ⇒ ***Attack*** *on the cryptographic technique*

- **Cryptographic security** can defined as the absence of attacks:

A cryptographic technique is said to be **secure** if no adversary succeeds in attacking it.

- In practice, it's important to describe the adversary's capabilities:
  - The control he exercises over the channel (read only, read + write):
    - **Passive** - the adversary can only eavesdrop on the communication channel.;
    - **Active** - in addition, the adversary has the ability to manipulate the information circulating on the communication channel (modify / block / insert or repeat messages).
  - Computational power:
    - **computationally unbounded** - adversary is able to execute any algorithm instantaneously (with no memory limitation);
    - **computationally bounded** - adversary can only execute algorithms in some complexity class (Probabilistic Polynomial Time - PPT).
  - Additional information (e.g. previous communications);
  - etc.
- Depending on the type of adversary being considered, different notions of (cryptographic) security can be achieved
  - **Unconditional** (or **information-theoretically**) s**ecurity**— secure against a computationally unbounded adversary;
  - **Computational security** — secure against a computationally bounded adversary (PPT).

Classic Ciphers

Caesar Cipher

- Known to have been used by Julius Caesar during the Gallic campaign
- The encryption is a shift of the letters of the alphabet. .

Texto limpo:	A	B	C	D	E	F	···	T	U	V	X	W	Y	Z
Criptograma ($K = 6$):	G	H	I	J	K	L	···	Z	A	B	C	D	E	F

- To decipher, the shift is made in the opposite direction.
- The total number of possible keys is 26 (one of which is weak).
- Example: Encrypting the message `CartagoEstaNoPapo` with key K=6 results in `IGXZGMUKYZGTUVGVU`.

…attacking Caesar

- E.g. want to attack the ciphertext: FXLNTQCLOPNPDLC
- Small key-space suggests the following **strategy #1**:
  - decrypt with every possible key…
  - …and spot the one(s) that "make sense"

CRIPTOGRAMA:		FXLNTQCLOPNPDLC
+1	:	GYMOURDMPQOQEMD
+2	:	HZNPVSENQRPRFNE
....	:	.....
[MSG]+15	:	UMACIFRADECESAR (K=26-15=11)
....	:	.....
+24	:	DVJLROAJMNLNBJA
+25	:	EWKMSPBKNOMOCKB

- **Strategy #2**: frequency analysis allows for more efficient attacks…

  high frequency of 'L' suggests that it encrypts 'A' (that is, K = 'L'-'A' = 11)

Brute-force Attack

- Strategy #1 attack is known as **Brute-force Attack**

  the adversary searches the entire key space in the hope of finding the right key

- It assumes that:
  - there is enough redundancy in the original message,
  - or a plaintext/ciphertext pair is known.
- Often seen as an attack that can always be applied to a cryptosystem...
- …but its feasibility depends on the size of the key space.
- **Conclusion**: key sizes are a necessary (although not sufficient) criterion for the security of ciphers.

big numbers!

…what is as reasonable key size?

- If we consider keys to be arbitrary bitstrings, the size of the key space increases exponentially on the key size
- Example:

Key Size	**Time** ($1\mu sec/test$)	**Time** ($1\mu sec/10^6 tests$)
32 bit	35.8 min.	2.15 msec.
40 bit	6.4 days	550 msec
56 bit	1140 years	10 hours
64 bit	500000 years	107 days
128 bit	$5 * 10^{24}$ years	$5 * 10^{18}$ years

- Baseline: 2^{112} provides a reasonable security level!

Monoalphabetic Substitution Cipher

- Instead of the shift used in the Ceaser Cipher, encryption performs an arbitrary permutation of the alphabet (decryption uses the inverse permutation);
- E.g.:

A	B	C	.....	W	Y	Z
R	X	K	.....	B	I	F

- Much larger key-space ($26! \approx 17.5 * 10^{24}$)

…should we trust the security of this cipher?

…an attack

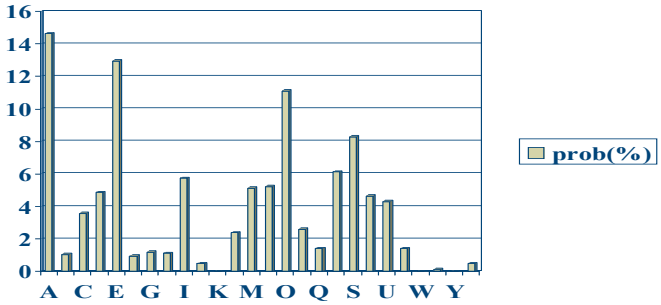
- Consider the following ciphertext:

```
FPGFBNBVPKFBDMSBEMDMGUCDKDGUGDMUSPMMDBEFLEFEQDCPPGIDEXDCBKPMDHKPMPFQBUGPSUGHKEGPF
QBMPXPKSESEBSURBHKBHBMEQBFUFSDSBGHKPPFCECPHDKQPDHDKQPFDBADVEDFDCDDCEZPKLDZEDGMPPM
NDKPMDGDVPEMPDNUPFQDVDMPCPZGEFUQBMCPUGMEOPFSEBHPFBMBFBDUNPCDPXSEQDSDBCBBUQBFBCPMU
KNEKDGUGDMHPKMBFDNPFMCPBKENPGBIMSUKDSBGJUPGPFQKPQEVPSBFSEOEDIUOBMPGKPMQDUKDFQPMPX
SPFQKESBMPMMPMQKDZEDGUGDHPKNUFQDQPKKEVPOBJUPPJUPVDEMUSPCPKKPMHBFCEOAPMJUPFDBMDIED
PPOPMBOADKDGHDKDBHKDQBSBGEFJUEPQDSDB
```

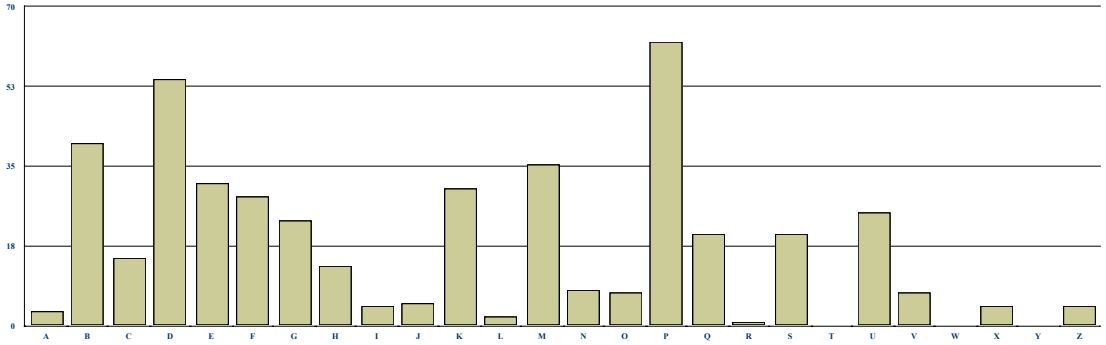
- …known to be a Portuguese text.

how should we proceed?

- Letter frequency of Portuguese:



- Histogram of letters in the ciphertext:



- Suggests matching “A”; “E”; “O” with “P”; “B”; “D”
- Pairs or triples of letters can also be considered, such as “os”; “es”; “que”; “nao”; …
- Several occurrences of the pairs “PM”, “PF”, “MP”, and “JUP” suggest the following partial decryption: {…; F:M; …}

```
ME-MO-O-E-MOAS-O-SAS-U-A-A-U-ASU-ESSAO-M--M--A-EE--A--A-O-ESA--ESEM-OU-E-U--
---EM-OSE-E-----O-U-O--O-OS--OMUM-A-O---EEM---E-A--EA-A--EMAO-A--AMA-AA---E-
-A--A-SEES-A-ESA-A-E-SEA-UEM-A-ASE-E---MU-OS-EU-S--EM--O-EMOSOMOAU-E-AE----
A-AO-OOU-OMO-ESU----A-U-AS-E-SOMA-EMS-EO---E-O-S-U-A-O-QUE-EM--E---E-OM----
A-U-OSE--ES-AU-AM-ESE--EM----OSESSES--A--A-U-A-E--UM-A-E----E-OQUEEQUE-A-SU-
E-E--ES-OM----ESQUEMAOSA--AEE-ESO--A-A--A-AO--A-O-O--MQU-E-A-AO
```

- …which, on closer inspection, does not seem to make much sense.… :-(

- We need to backtrack on some of our hypotheses…
- Reassigning the decryption of F into N…
- …we get something a lot more promising:

```
NE-NO-O-ERNOAS-O-SAS-U-ARA-U-ASU-ESSAO-N--N-TA-EE--A--A-ORESA-RESENTOU-E-U--R--
ENTOSE-ER----O-U-O-RO-OS-TONUN-A-O--REEN---E-ARTEA-ARTENAO-A--ANA-AA---ER-A--A-
SEES-ARESA-A-E-SEA-UENTA-ASE-E---NUTOS-EU-S--EN--O-ENOSONOAU-E-AE---TA-AO-
OOUTONO-ESUR--RA-U-AS-ERSONA-ENS-EOR--E-O-S-URA-O-QUE-ENTRET--E-ON----A-U-OSE-
RESTAURANTESE--ENTR--OSESSESTRA--A-U-A-ER-UNTATERR--E-OQUEEQUE-A-SU-E-ERRES-ON---
-ESQUENAOSA--AEE-ESO--ARA--ARAO-RATO-O--NQU-ETA-AO
```

- which rapidly leads to the whole plaintext

```
NEMNOGOVERNOASCOISASMUDARAMUMASUCESSAOINFINITADEEMBAIXADORESAPRESENTOUMECUMPRIMEN
TOSEXERCICIOCUJOPROPOSITONUNCACOMPREENDIDEPARTEAPARTENAOHAVIANADAADIZERFAZIAMSEES
GARESAMAVEISEAGUENTAVASEDEZMINUTOSDEUMSILENCIOPENOSONOAUGEDAEXCITACAODOOUTONODESU
RGIRAMUMASPERSONAGENSDEORIGEMOBSCURACOMQUEMENTRETIVECONCILIABULOSEMRESTAURANTESEX
CENTRICOSESSESTRAZIAMUMAPERGUNTATERRIVELOQUEEQUEVAISUCEDERRESPONDILHESQUENAOSABIA
EELESOLHARAMPARAOPRATOCOMINQUIETACAO
```

Vigenère Cipher

(polyalphabetic substitution)

- Attributed to *Blaise Vigenère* (16th century). Known as "*le chiffre indéchiffrable*".
- It interleaves multiple Caesar ciphers.
- Broken in the 19th century by *Charles Babbage* and *Friedrich Kasiski*.
- Description:
  - key is a password — each letter corresponds to a single Caesar cipher key (A=0; B=1; …);
  - Encryption: apply the Caesar cipher with each character of the key in sequence, starting over when no more key characters are available.
- Observations:
  - same letter is not always encrypted in the same way (add some hurdles to the frequency analysis)
  - … but if the plaintext is much larger than the key, patterns in the plaintext shall be reflected in the ciphertext.
  - The cryptanalysis techniques that have been developed exhibit already some degree of sophistication.

Transposition Cipher

- Encryption permutes character positions (instead of changing them…)
- E.g.
  - consider a permutation [2; 1; 3] (key)
  - to encrypt “AindaOutraCifra”, write it on a matrix…

2	1	3
A	I	N
D	A	O
U	T	R
A	C	I
F	R	A

  - …and read the columns (with header [1; 2; 3])
  - resulting ciphertext: “IATCRADUAFNORIA”
- Observation: frequency analysis is now useless!

Cipher Composition

- Having seen different simple ciphers…
- Is it sensible to construct more intricate ciphers, which are more secure, by combining multiple simpler ciphers?

- It depends!
  - it may happen that it doesn't add any value (e.g. combining two ciphers by substitution)
  - but there are combination patterns that can be advantageous (e.g. by interleaving substitutions with permutations — aka *SP-networks*).

One-Time Pad

(Vernam Cipher - 1917)

- Generalises Vigenère cipher with:
  1. key size is, at least, the plaintext size;
  2. key is fully random;
  3. key is only used in a single encryption.
- Normally described as operating on a binary alphabet: encryption/decryption is the *exclusive-or* (xor) with the corresponding key bit.

$$C_i = T_i \oplus K_i \qquad M_i = C_i \oplus K_i$$

- Shown to be secure (**information-theoretically secure**) by *Claude Shannon — "knowing the ciphertext does not reduce the uncertainty about the plaintext"*.
- Key generation and distribution make this cipher unfeasible in realistic scenarios.

Conclusion

- In cryptanalysis, all available information is used, including:
  - partial information about the transmitted message;
  - previous cipher's use (e.g. messages encrypted with the same key);
  - and possible weaknesses in the cipher's use (e.g. deficiencies in the choice of keys, etc.).
- Although unconditionally secure techniques exist, they often have such strict requirements that they become impractical.
- Most (all?) cryptographic techniques used today are based on **computational security**, which limits the adversary's computational capabilities.
- The size of a cipher's key should be determined based on the desired level of security (e.g. 2^{112}), while taking into account the amount of key size that is “consumed” by known cryptanalysis techniques.

Cryptography & Security

(Cryptographic) Security Properties

- We have mainly focused on ciphers — a cryptographic primitive whose goal is to ensure *confidentiality*.
- But cryptography is used today to provide guarantees for a wide range of security properties:
  - **confidentiality** — content of the message is only known to the legitimate parties;
  - **integrity** — the recipient would "reject" messages that have been tampered;
  - **autenticity** — ensures the "origin" of the message to the recipient;
  - **non-repudiation** — the “origin” of the message cannot deny it;
  - **anonymity** — no information available regarding the “origin” of the message;
  - **identification** — establish the “identity” of a party;
  - …

Cryptographic Services and Protocols

- Usually, one is interested in a combination of these properties (e.g. a secure channel between two parties aims to guarantee confidentiality, authenticity and integrity).
- On the other hand, some of these properties do not directly follow from a single cryptographic technique, but rather from a combination of techniques.
- Leading to what is known as **cryptographic protocols** — specifications for message exchanges that rely on cryptography to achieve a desired end.
- The security of those protocols relies not only on the security of the underlying cryptographic techniques, but also on subtle interactions between them.

Cryptography & Security

*The security of a system using cryptography **is not just** the security of the underlying cryptographic techniques.*

- We can distinguish (at least) the following levels when establishing the security of a system that uses cryptography:
  - cryptographic scheme;
  - protocol;
  - implementation (coding);
  - usage.

A breach at any of these levels jeopardises the security of the entire system!

Applied Cryptography

Systems and Information Security
Informatics Engineering (3rd year, 2nd sem.)

José Bacelar Almeida

Applied Cryptography

Symmetric Crypto

Asymmetric Crypto

Applications

Roadmap

- Ciphers
  - Stream ciphers
  - Block ciphers
- One-Way Functions
  - Cryptographic Hash Functions
  - Message Authentication Codes (MAC)
  - Key-Derivation Functions (KDF)
- Key-Management

Stream Ciphers

Recall OTP (one-time pad)

- Known to be (information-theoretically) secure

- Inherent issues:

  1. keys cannot be reused

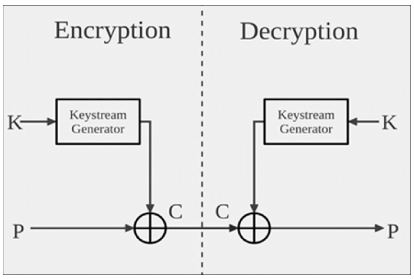
  2. truly random key with size greater than the plaintext

  3. does not promote diffusion — information on the structure of the message can be used to manipulate it (bit swapping).

- Difficulties associated with key-generation and key-distribution render the OTP mostly unusable in real-world applications.

Stream-Ciphers

- **Basic idea**: approximate OTP using a key stream generator which generates an arbitrary length keystream from a short, fixed-length key.

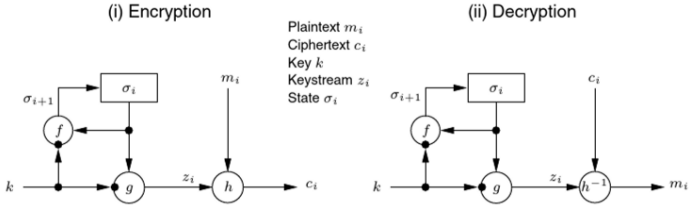


- The generation of the keystream must be reproducible (deterministic) — a finite state machine.
- Therefore, the sequence must necessarily be **cyclic**. The **period** is the length of the sequence before it repeats..

Criteria for the design of Stream Ciphers

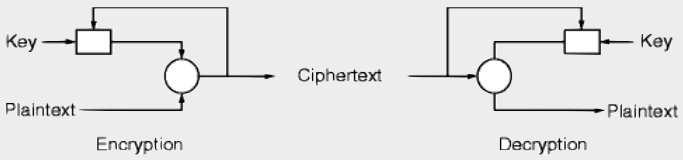
- Period should be as long as possible (always longer than the plaintext).
- Key sequence must be:
  - **pseudo-random**: statistic properties of the sequence are those of a truly-random sequence;
  - **unpredictable**: should be impossible to predict the “next bit”, after observing a given prefix of the sequence
- Other characteristics
  - synchronism:
    - **synchronous** — keystream is independent of the message;
    - **self-synchronising** — able to recover synchronism when bits of the ciphertext are lost .
  - error-propagation: the impact of transmission errors on decrypted messages
  - …

Synchronous Stream Ciphers



- Keystream is independent of the message;
- Loss or insertion of bits in the cryptogram results in loss of synchronisation.
- Transmission errors (bit swapping) only affect the corresponding position of the original message.
- The key might affect:
  - The next-state function *f* — **Output-Feedbak Mode**.
  - The output function *g* — **Counter Mode**.
  - Both...

Self-synchronising Stream Ciphers



- Next-bit is computed from the last n bits of the ciphertext (and key);
- A IV (initialisation vector) is used to initiate the process;
- In case of a transmission error, synchronisation is restored once the flipped bit is no longer used for the next bit computation.
- Possible problem: vulnerable to replay attacks.

Cryptographic-Secure Pseudo-Random Number Generation

- *Golomb*'s Randomness Postulates:
    1. The difference in the number of 1s and 0s must tend towards zero;
    2. The expected number of sub-sequences of repeated symbols (runs) with length *l* is given by $r(l)=r/2^l$;
    3. The auto-correlation must be a constant value for any deviation other than *0 (mod p)*.
- Cryptographic Security:

    The generated sequence must be indistinguishable from a random sequence for any Probabilistic Polynomial Time (PPT) adversary.

Key reuse and NONCEs

- The above description of sequential ciphers inherits a problem from OTP: **key reuse**
    - encrypting different messages with the same key shall lead to the same generated keystream.
- This problem is generally overcome by using NONCEs — abbreviation of **Number used only ONCE**.
    - A number that should never be repeated;
    - but not required to be kept secret (i.e. it can be sent along with the cryptogram).
- In practice, the Nonce is typically a sequence of randomly generated bytes used during encryption and sent along with the cryptogram.

Some Examples

- Stream ciphers are used in several well-known applications:
  - **A5** (**A5/2**), used in the GSM standard;
  - **E0**, used in the Bluetooth protocol;
  - **CSS** (Content Scramble System), used to protect DVD discs.

  (Note: All of the above examples are known to offer weak security guarantees.)
- **RC4** (ArcFour) — a cipher designed for efficient software implementation, which was once widely used but **is now considered broken**.
- **ChaCha20** — a modern (and secure) stream cipher.

Block Ciphers