# Introduction to Cryptology: confidentiality, integrity, authenticity 2015/12 - Yaoundé, Cameroun 

Damien Robert

Équipe LFANT, Inria Bordeaux Sud-Ouest
Institut de Mathématiques de Bordeaux
Équipe MACISA, Laboratoire International de Recherche en Informatique et Mathématiques Appliquées

université
${ }^{\text {de }}$ BORDEAUX


## Cryptology

Cryptology = Cryptography + Cryptanalysis Usage:

- SSL/TLS, ssh, gpg
- GSM, Wifi, Bluetooth
- Credit Card, Transport card, Passport


## Remark

Cryptology $\subset$ Security

## Public Canal

Alice communicates with Bob through a public canal. Eve does passive attacks on this canal (spying) and Charlie does active attacks.

Active attacks:

- Usurpation of identity;
- Altering data;
- Repudiation
- Replay, repetition
- Man in the middle
- Delay, Destruction
- Confidentiality
- Integrity
- Authenticity
- Key generation
- Disponibility
- Non repudiation
- Confidentiality Symmetric encryption, Asymmetric encryption
- Integrity Cryptographic hash functions
- Authenticity Signature, MAC
- Key generation Randomness
- Disponibility
- Non repudiation


## Primitive

- Without key: hash, random generator
- With key
- symmetric
- MAC
- Encryption: stream, block
- asymmetric: number theory, algebraic geometry, codes, lattices...
- signature
- Encryption


## Confidentiality, Authenticity

- Confidentiality: $E:\{0,1\}^{n} \rightarrow\{0,1\}^{n}$ a permutation; $D=E^{-1}$. Encryption: $m \mapsto c=E(m)$, Decryption: $c \mapsto m=D(c)$.
- Authenticity: $c=A(m)$. Alice sends ( $m, v$ ). Bob receives ( $m^{\prime}, v^{\prime}$ ). Verification: $V\left(m^{\prime}, v^{\prime}\right)=$ OK, NOT OK.
- $E$ and $D$ needs to be secret;
- So no external validation of security possible;
- And transmitting the algorithmes in painful;
- Kerchoff: parametrizes the algorithms by a key $K$;
- $E:\{0,1\}^{n} \times\{0,1\}^{k} \rightarrow\{0,1\}^{n}$ such that $E(\cdot, K)$ is a permutation for all $K$;
- $E$ can be public, the only secret is the secret key $K$.
- Public
- Authenticated
- Confidential
- Authenticated + Confidential


## Goals

Using a authenticated and/or confidential channel, construct an authenticated and/or confidential channels inside a public channel.

The goal if of course to use the preexisting authenticated/confidential channel as little as possible, and do everything else in the authenticated/confidential channel constructed inside the public channel.

## Transforming a public channel into a confidential/authenticated channel

- Authenticated $\Rightarrow$ Authenticated;
- Authenticated+Confidential $\Rightarrow$ Authenticated
- Authenticated $\Rightarrow$ Authenticated
- Authenticated+Confidential $\Rightarrow$ Confidential
- Authenticated $\Rightarrow$ Confidential
- Authenticated $\Rightarrow$ Authenticated+Confidential $\Rightarrow$ Authenticated+Confidential


## Example 1: integrity

Alice sends $m$ through the public channel, and $h(m)$ through the integrity channel, where $h:\{0,1\}^{*} \rightarrow\{0,1\}^{256}$ is a cryptographic hash function.

## Security:

- Preimage resistance;
- Second-Preimage resistance;
- Collision resistance.


## Example 2: authenticity (symmetric)

A secret key $K$ is generated (randomisation) and sent to Alice and Bob through a authenticated and confidential channel. Alice sends ( $m, \operatorname{MAC}(m, K)$ ) through the public channel. Bob verify via VERIF $(m, K)$. The MAC is a hash function parametrized by $K$.

Security: from several couples ( $M, C$ )

- Can't retrieve $K$;
- Can't generate a new ( $M^{\prime}, C^{\prime}$ );
- Can't distinguish the distribution $C$ from a uniform distribution.


## Example 2: authenticity (asymmetric)

A couple ( $K_{S}, K_{P}$ ) is generated by Alice and she sends the public key $K_{P}$ to Bob via an authenticated channel. Alice sends $\left(m, \operatorname{SIGN}\left(m, K_{S}\right)\right)$ through the public channel, and Bob verify via VERIF $\left(m, K_{P}\right)$.

Security: from several couples $(M, C)$ and $K_{P}$

- Can't retrieve $K_{S}$;
- Can't generate a new $\left(M^{\prime}, C^{\prime}\right)$;
- Can't distinguish the distribution $C$ from a uniform distribution.

Signature vs MAC:

- Public verification
- Non repudiation


## Example 3: confidentiality (symmetric)

A secret key $K$ is generated (randomisation) and sent to Alice and Bob through a authenticated and confidential channel. Alice sends $c=E(m, K)$ through the public channel, Bob decrypts via $m=D(c, K)$.

Security: from several ciphers $C$, several couples $(M, C=E(M, K)$ ) (chosen plain text attack) and several couples ( $C, M=D(M, K)$ ) (chosen cipher text attack)

- Can't retrieve $K$;
- Can't find $M^{\prime}$ from a new $C^{\prime}$;
- Can't distinguish the distribution $C$ from a uniform distribution.


## Example 3: confidentiality (asymmetric)

A couple ( $K_{S}, K_{P}$ ) is generated by Bob and he sends the public key $K_{P}$ to Alice via an authenticated channel. Alice sends $c=E\left(m, K_{P}\right)$ through the public channel, and Bob decrypt via $m=D\left(c, K_{S}\right)$.

Security: from $K_{P}$, several ciphers $C$, several couples $(M, C=E(M, K))$ (chosen plain text attack) and several couples ( $C, M=D(M, K)$ ) (chosen cipher text attack)

- Can't retrieve $K_{S}$;
- Can't find $M^{\prime}$ from a new $C^{\prime}$;
- Can't distinguish the distribution $C$ from a uniform distribution.


## Asymmetric Vs symmetric

- $N$ persons $\Rightarrow N$ keys rather than $N^{2}$;
- Does not need a confidential channel;
- Much slower.
$\Rightarrow$ Use an asymmetric cipher to send a symmetric secret key and switch to the symmetric channel to increase speed.
- Only cipher known
- Chosen plain text: CPA, Adaptative chosen plain text: CPA2
- Chosen cipher text: CCA, Adaptative chosen cipher text: CCA2
$\Rightarrow$ Invert the function: OW (One Wayness)
$\Rightarrow$ Indistiguinbility: IND. Detect an encryption of 1 from an encryption of 0 with probability $>0.5+\varepsilon$. IND $=$ Semantic security.
Ultimate goal: IND-CCA2 cryptosystem.


## Attacks:

- Black box or structural analysis;
- Side channels.


## Security parameters

- $2^{40}$ : One minute on a standard computer;
- $2^{60}$ : One year on a standard computer;
- $2^{80}$ : One year with $10^{6}$ cores at $5 \mathrm{GHz}=$ NSA?
- $2^{128}$ : security goal.


## Remark

One may want to take higher security parameters 192 bits or 256 bits for very long term security and for protection against potential attacks. For instance quantum computers can divide by 2 the security of some problems (and completely kill others like factorisation or the discrete logarithm problem).

## One way function

- Very important in cryptology, needs strong properties
- One way function $x \mapsto f(x)$ easy, but $y \mapsto f^{-1}(y)$ hard.


## Example <br> Multiplication, exponentiation in $\left(\mathbb{Z} / p \mathbb{Z}^{*}, \times\right.$ ).

In asymmetric key cryptography, use of trapdoor one way function: a secret trap can allows to compute $f^{-1}$.

## Example

If $N=p q, x \mapsto x^{2}$ is a trapdoor one way function.

- Origin: hash table to speed up lookup;
- $h:\{0,1\}^{*} \rightarrow\{0,1\}^{n}$;
- Very important in cryptology, needs strong properties:
- No preimage
- No second preimage
- No collision


## Example

- A checksum is not an hash function
- $x \mapsto x^{2} \bmod p q$ is one way but not a cryptographic hash function.
- MD5 (Broken), SHA-1 (Almost broken), SHA-2, SHA-3

Let $h$ be an hash function with $n$ bits of output; $N=2^{n}$.

- Given $y \in\{0,1\}^{n}$, finding $y$ at random requires $\Theta\left(2^{n}\right)$ tries;
- What about a collision: $x_{1} \neq x_{2} \mid h\left(x_{1}\right)=h\left(x_{2}\right)$ ?
- After $k$ tries, the probability of not finding a collision is

$$
p(k)=(1-1 / N)(1-2 / N) \ldots(1-k / N)
$$

- So we have the inequalities (in fact it is an order of equivalence)

$$
\begin{gathered}
\log p(k)=\sum_{i=1}^{k} \log (1-i / N) \leqslant-\sum_{i=1}^{k} i / N \leqslant-k^{2} / 2 N \\
p(k) \leqslant \exp \left(-k^{2} / 2 N\right)
\end{gathered}
$$

- So if $k=\Theta(\sqrt{N}), p(k)$ is small and the probability of collision is high;
- To get 128 bits of security we need $n=256$
- Compression function $f:\{0,1\}^{2 n} \rightarrow\{0,1\}^{n}$;
- Merkle-Damgard (IV=Input Vector)

$$
h\left(m_{1}, \ldots, m_{k}\right)=f\left(f\left(\ldots f\left(f\left(I V, m_{1}\right), m_{2}\right), \ldots, m_{k}\right)\right.
$$

- If $m$ is not of length a multiple of $n, \operatorname{pad} n$ with its length


## Theorem

If $f$ is collision resistant, then $h$ too.

## Remark

- A small weakness in $f$ can lead to a big weakness in $h$ (MD5);
- From $h(m)$, we can compute $h\left(m \| m_{0}\right)$ without knowing $m$.

Constructing $f$ : mixing boolean operations, addition (non bit-linear), into multiple rounds with some magic constants.

- Alice can publish $h(m)$ to prove later that she selected $m$;
- If $m$ is small (YES/NO), necessity to pad with a random sequence. "The hash of my CB card is fege5fa6d82a071422a576064doed49a7266ccb49390037c255e6e7baa8d4535" is a bad idea. Better idea: publish the hash of (CB card + long random sequence of characters).


## Example

- Head or tails by phone
- Zero-Knowledge via coloring graphs
- Stocking a password: pass, $h$ (pass), $h$ (login || pass), $h$ (salt || pass). Question: What is the best method?
- Integrity of a software download (Exercice: why would a signature be better?)
- $h_{K}(m)$ were $K$ is a secret key between Alice and Bob to prove the identity of the emitter
- Security: Eve can not produce ( $m^{\prime}, h_{K}\left(m^{\prime}\right)$ );
- $\operatorname{HMAC}(K, m)=H\left(K \oplus c_{1}\left\|H\left(K \oplus c_{2}\right)\right\| m\right)$; proven secure
- $H(M \| K)$ or $H(K \| M)$ is a bad idea if $h$ is constructed via Merkle-Damgard (MD5, SHA-1, SHA-2) but should be ok if $h$ use a sponge function (SHA-3).
- $h_{K}$ could also be a block cipher
- In fact a bloc cipher $E$ can define a hash function via $h_{i}=E\left(h_{i-1}, m_{i}\right) \oplus h_{i-1}$ (Davies-Meyer) but we want a faster hash function.


## Randomness generator

- Vital in a cryptosystem
- Sony: random = constant
- Debian ssh: random = date
- RSA key: a lot of common primes in public modulus
- Standard randomness: good statistic properties
- Linear congruence: $x_{n+1}=a x_{n}+b \bmod M$ very fast but some statistic bias
- Cryptographic randomness: needs much stronger properties
- Can't predict the next bit from the observed ones
- Broken for linear congruences
- True alea = non compressable (Kolmogorov)
- By definition an algorithm can't generate a true alea
$\Rightarrow$ Pseudorandom generator.
- Construction: a small seed (true alea) used by the pseudorandom generator
- Hash function = Compress state; PRNG = Expand state.
- $s$ internal state: $s=f(s)$ (update internal state), $x=g(s)$ (output next random bit)


## Remark

- Finite number of internal states $2^{n} \Rightarrow$ the PRNG will loop
- Birthday paradox: A "random" update of the seed loop in time $\sqrt{\text { internal states }}$
- Arithmetic PNRG can force a loop of $2^{n}$
- Bad idea for cryptography
- We need a true alea to initialize the seed
- Use physical input: input/output, mouse mouvements, IP packets...
$\Rightarrow$ In linux, /dev/random collects the entropy and outputs a random sequence until the entropy is $0 \Rightarrow$ blocks waiting for new entropy
- /dev/urandom uses the entropy inside a PRNG to output a random sequence which never blocks
- Problem: in early boot, urandom may output a sequence while the seed had not enough entropy yet;
- In an idle machine not a lot of entropy; even worse for virtual machine without help from the container;
- Possible solution: with a good PRNG, we just need an initial seed of true 256 bits of entropy; keep the current state across reboots.
- Quantity of information: if $p(x=1)=0.99$ and $p(x=0)=0.01$, observing $x=0$ is much more useful than $x=1$;
- Quantity of information: $Q I(m=x)=\log _{2}(1 / p(x))$
- The entropy is the average value of the $Q I$ :

$$
e=-\sum p_{i} \log _{2} p_{i}
$$

- $n$ bits of entropy $\approx$ information that needs $n$ bits to be encoded.


## Example

- $x \in\{0,1 \ldots, 15\}$ uniformly: 4 bits of entropy
- $p_{A}=0.5, p_{B}=0.25, p_{C}=0.25$. $e=1 / 2+2 / 4+2 / 4=3 / 2$. Encode $A$ with $0, B$ with 10 and $C$ with $11 \Rightarrow 3 / 2$ bits on average to encode the message.
- CESAR: translate letter in the alphabet by the same amount
- Alphabetical substitution
- VIGENERE: CESAR depending on the position of the letter:

CRYPTOGRAPHYINGABON SECRETSECRETSECRETS UVAGXHYVCGLRARIRFHF

- The messages correspond to word in a language, they are not uniform;
- If the ciphers are still non uniform $\Rightarrow$ statistical attacks
- Alphabetical substitution: most frequent letters, vowels are linked with many other letters;
- Index of correlation: split the message into several lines. Probability that one letter is the same as the letter below (ie probability that two random letters are the same);
- Uniform messages: index of correlation is $1 / 26$;
- Far from the case in French: $10-15 \%$;
- Vigenere: if we split the messages into blocks of length $k$ and find an index of correlation similar to the French one then high probability than the length of the secret is a divisor of $k$ and we are back to CESAR.
- Keyboard Plug $P$ (several disjoint transposition)
- Several rotors $R_{i}$ (each rotor is a permutation): 3 then 4
- Reflector M: 13 couples for the 26 letters $A \hookrightarrow D, B \longleftrightarrow M, \ldots$
- $E(m)=P^{-1} R_{1}^{-1} R_{2}^{-1} R_{3}^{-1} M R_{3} R_{2} R_{1} P$
- After each output, $R_{1}$ makes a turn; if $R_{1}$ has made a full
- The secret state is given by the position of the plugboard and the initial position of the rotors.
- Feature: $E^{2}=$ Id so decryption use the same initial state as encryption;
- Security problem: for all letter $x, E(x) \neq x$. Big statistical drawback;
- Cryptanalysis of Enigma (Poland then England+USA). A bombe explores a lot of Enigma position, using statistical analysis to greatly speed up the process.
- Keyboard Plug $P$ (several disjoint transposition)
- Several rotors $R_{i}$ (each rotor is a permutation): 3 then 4
- Reflector M: 13 couples for the 26 letters $A \hookrightarrow D, B \longleftrightarrow M, \ldots$
- $E(m)=P^{-1} R_{1}^{-1} R_{2}^{-1} R_{3}^{-1} M R_{3} R_{2} R_{1} P$
- After each output, $R_{1}$ makes a turn; if $R_{1}$ has made a full
- The secret state is given by the position of the plugboard and the initial position of the rotors.
- Feature: $E^{2}=$ Id so decryption use the same initial state as encryption;
- Security problem: for all letter $x, E(x) \neq x$. Big statistical drawback;
- Cryptanalysis of Enigma (Poland then England+USA). A bombe explores a lot of Enigma position, using statistical analysis to greatly speed up the process.
- Linearity: solve big linear systems
- Algebraic attacks: solve big multivariate algebraic systems
- Differential attacks: let $E_{k}\left(m \oplus \Delta_{m}\right)=c \oplus \Delta_{c}$ and study the distribution of $\Delta_{c}$.
- Vernam cipher (One time pad) $c_{i}=m_{i} \oplus k_{i}$
- Shannon: unconditionally secure if $k=k_{1} \ldots k_{n}$ is uniform random (Proof: distribution of $m \oplus$ uniform distribution = uniform distribution);
- Not convenient: key of same length as the message
- Reusing key (or part of the key) is catastrophic: if $c_{1}=m_{1} \oplus k$ and $c_{2}=m_{2} \oplus k$ then $c_{1} \oplus c_{2}=m_{1} \oplus m_{2}$; this reveals a lot of information;
- Unconditonal security is too strong, we only care about computational security.
- Simulate the One Time Pad by using a PRNG parametrized by a secret key $k$;
- Internal state: $s_{i}$. Update: $s_{i+1}=f\left(s_{i}, K\right)$. Output $x_{i}=g\left(s_{i}, K\right)$.
- Encryption/Decryption: $c_{i}=m_{i} \oplus x_{i}$.


## Remark

Problems of synchronisation. Autosynchronising stream ciphers: use the last $t$ ciphers as the state: $x_{i}=g\left(c_{i-t}, \ldots, c_{i-1}, K\right)$. If there is an error of transmission, this corrupt the decryption for only $t$ bits.

## Linear Feedback Shift Register (LFSR)

A LFSR has $L$ cells.

- Output $x_{0}$;
- Shift: $x_{i}=x_{i+1}$;
- Feedback: $x_{L-1}=x_{i_{1}} \oplus x_{i_{2}} \oplus \cdots \oplus x_{i_{k}}$.


## Definition

The retroaction polynomial is $P(x)=x^{L}+\sum x_{i_{k}} x^{k}$.
The LFSR is uniquely determined by its initial value and its retroaction polynomial.

## Linear algebra

- The state of the vector $X=\left(x_{0}, \ldots, x_{L_{1}}\right)$ in the LFSR is linear
- For instance if $P(x)=x^{4}+x+1$, then at step $i+1, X_{i+1}=M X_{i}$ where

$$
M=\left(\begin{array}{llll}
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1
\end{array}\right) \in \mathbb{F}_{2}
$$

- The characteristic polynomial of $M$ is $P(x)$.
- The LFSR will loop when $M^{k}=$ Id
- This is the order of $X$ in $\mathbb{F}_{2}[X] / P(x)$ (if $P$ is the minimal polynomial).
- If $P(x)$ is irreducible then $K=\mathbb{F}_{2}[X] / P(X)=\mathbb{F}_{2}(x)$ is a field of degree $L$. The order of $x$ in $K$ divides $2^{L}-1$.
- If $P$ is primitive the order is exactly $2^{L}-1$;
- If $P=\prod P_{i}$ is a product of distinct irreducible polynomials then $K=\prod \mathbb{F}_{2}[X] / P_{i}(X)$ is a product of fields (CRT) and the period divides $\prod 2^{\operatorname{deg} P_{i}}-1<2^{L}-1$.
- The highest period is given by primitives polynomials.


## Theorem

There is $\frac{\varphi\left(2^{L}-1\right)}{L}$ primitive polynomials of degree $L$ in $F_{2}[x]$.
In particular if $2^{L}-1$ is prime (a Mersenne prime) then there is $\left(2^{L}-2\right) / L$ primitive polynomials of degree $L$ and all irreducible polynomials are primitive.

## Proof.

The splitting field of an irreducible polynomial of degree $L$ is always $\mathbb{F}_{2 L}$ since the absolute Galois group is procyclic. There is $\varphi\left(2^{L}-1\right)$ generators of the multiplicative group $\mathbb{F}_{2^{L}}$. The Galois group splits this group into $\varphi\left(2^{L}-1\right) / L$ orbits (since the Frobenius is of order $L$ ), each orbit corresponds to a primitive polynomial.

- An LFSR can have a high period;
- But the output is linear, from $2 L$ terms one can recover its retroaction polynomial (Berlekamp Masse)


## Proof.

There exists a fraction $P_{0} / P_{1}$ whose formal sum $\sum x_{i} X^{i}$ corresponds to the bit output by the LFSR. The Euclidean algorithm between $\sum_{i=0}^{2 L-1} x_{i} X^{i}$ and $X^{2 L}$ recovers this fraction (as the continued fraction algorithm recovers the rational fraction $p / q$ from its decimal development).

- In practice combine several LFSR with a non linear filter function
- A5/1 (GSM) combines 3 LFSR; but the filter function is weak $\Rightarrow$ attacks if enough data is gathered.
- $c=E_{K}(m), m=D_{K}(c)$ where $m$ is a block of $n$ bits and $K$ is block of $k$ bits;
- There is $\left(2^{n}\right)$ ! bijections and $2^{k}$ possible keys; so we can have $k>m$.
- If the block is too small $(n=8)$ dictionary attacks;
- AES works with blocks of 128 bits but has three level of security: 128 , 192 and 256 bits (which corresponds to 10,12 and 14 rounds).
- Security: Observing ( $m, c$ ) should reveal no information on $K$ or allows to generate ( $m^{\prime}, c^{\prime}$ ).
- Related keys: changing one bit of $K$ should completely change the ( $m, c$ ).
- Several rounds: $m=L_{0} \| R_{0}$.
- $L_{i+1}=R_{i}, R_{i+1}=L_{i} \oplus F_{K_{i}}\left(R_{i}\right)\left(K_{i}\right.$ is a key derived from $K$ for round $\left.i\right)$
- This is always inversible, even if $F_{K_{i}}$ is not injective!
- Decrypting: $R_{i}=L_{i+1}, L_{i}=R_{i+1} \oplus F_{K_{i}}\left(L_{i+1}\right)$.
- Used by DES: Feistel scheme with 16 rounds.
- Blocks of 64 bits, key of 56 bits;
- Good for the time (1976);
- Key size too low now.
- Triple DES used instead (now superseded by AES): $E_{K_{1}, K_{2}}=\mathrm{DES}_{K_{1}} \circ \mathrm{DES}_{K_{2}}^{-1} \circ \mathrm{DES}_{K_{1}}$;
- Key length of Triple DES is 112 bits.
- Some plain text attacks $\Rightarrow$ effective security of 80 bits.


## Exercice

(2) Why not simply use $E_{K_{1}, K_{2}}=\mathrm{DES}_{K_{2}} \circ \mathrm{DES}_{K_{1}}$ ?
(2) Why the $\mathrm{DES}_{K_{2}}^{-1}$ in the middle?

- Selection by NIST in 2001;
- Blocks of size 128, Keys of size 128, 192, 256;
- Several rounds $(10,12,14)$ parametrized by subkeys;
- One round: 128 bits = 16 bytes, organized in a $4 \times 4$ square;
(2) SubBytes: inversion in $\mathbb{F}_{2^{8}}=\mathbb{F}_{2}[x] / x^{8}+x^{4}+x^{3}+x+1$
(2) ShiftRows: rows are shifted (by a different value)
(3) MixColumns: linear
(4) AddRoundKey: XOR with the derived keys.


## Electronic Code Book (ECB)

- $c_{i}=E_{K}\left(m_{i}\right)$;


## Example (From Wikipedia)

Name + Salaries encrypted through ECB with blocks of 2 characters. Jack salary is $105000 €$ by year, and the encrypted data is Q92DFPVXC9IO. The other encrypted data are TOAV6RFPY5VXC9, YPFGFPDFDFIO, Q9AXFPC9IOIO, ACED4TFPVXIOIO, UTJSDGFPRTAVIO What is the salary of Jane, Jack's boss?

## Cipher-block chaining (CBC)

- Initialisation: $c_{0}=I V$ (input/initialisation vector)
- $c_{i}=E_{K}\left(c_{i-1} \oplus m_{i}\right)$;
- $m_{i}=c_{i-1} \oplus D_{K}\left(c_{i}\right)$;
- Randomizing the IV $\Rightarrow$ same plain text to different cipher text.


## Counter

- $I_{0}=I V$
- $c_{j}=m_{j} \oplus E_{K}\left(I_{j}\right) ; I_{j+1}=I_{j}+1$,
- This is actually a stream cipher!


## Output feedback (OFB)

- $I_{0}=I V$
- $c_{j}=m_{j} \oplus I_{j} ; I_{j+1}=E_{K}\left(I_{j}\right)$
- A block cipher can also be used as a MAC: the last cipher block is the MAC (needs a good operational mode).
- MAC then Encrypt? (SSL); Encrypt and MAC? (SSH); Encrypt then MAC?
- Encrypt then MAC is secure; MAC then Encrypt has a lot of problem (decryption oracle); Encrypt and MAC has theoretical problems (For instance $\mathrm{MAC}=E \oplus m$ ) but no strong practical problems.
- Authenticity+Integrity: HMAC, Poly1305, Galois Message Authentication Code (GMAC);
- Confidentiality+Authenticity+Integrity: GCM (Galois Counter Mode), CCM (Counter Mode + CBC-Mac)
- Block ciphers: AES
- Stream ciphers: Salsa20 (and the variant Chacha20), also used in the BLAKE hash function, ESTREAM.


## Challenge / Answer

- Bob chooses a random $r$, computes $x=h(r)$ and sends the challenge $\left(x, E_{K_{P}}(r)\right)$ to Alice;
- Alice decrypt to find $r$, checks that $x=h(r)$ and sends the answer $r$ to Bob;
- Question: Why use a hash function here and not just send $E_{K_{P}}(r)$ ?


## Signature

- Bob sends a random message $r$ to Alice;
- Alice appends random noise to $r$ and signs this.
- Question: Why the random noise?
- Even with asymmetric cryptography, we still need an authenticated channel to transmit the public key $K_{P}$;
- Web of trust (decentralized): I trust the persons trusted by the persons I trust. Used by gpg.
- PKI (centralized): public key signed by an organism via a certificate. Verification done recursively until we find a root certificate. Used by TLS/SSL: 166 root certificates on my computer.
- Certificate for $n$ persons: $n$ certificates? 1 certificate using a binary hash tree: recursively if the node $N$ has two children $C_{1}, C_{2}$ then $h(N)=h\left(C_{1} \| C_{2}\right)$. We only need to verify the authenticity of the root node $R$; verification of a node $N$ only uses the path between $N$ and $R$ $\Rightarrow O(\log n)$.
- TLS Key exchange + Authentication algorithms: RSA, DHE-RSA, DHE-DSS, ECDH-ECDSA, ECDHE-ECDSA, ECDH-RSA, ECDHE-RSA
- TLS Ciphers: AES-CBC, AES-CCM, AES-GCM, Chacha20-Poly1305
- SSH Authentication: id_dsa, id_rsa, id_ecdsa, id_ed25519
- SSH Key exchange algorithms: curve25519-sha256@libssh.org, ecdh-sha2-nistp256, ecdh-sha2-nistp384, ecdh-sha2-nistp521, diffie-hellman-group-exchange-sha256, diffie-hellman-group-exchange-sha1, diffie-hellman-group14-sha1, diffie-hellman-group1-sha1
- SSH Ciphers: aes128-ctr, aes192-ctr, aes256-ctr, arcfour256, arcfour128, aes128-gcm@openssh.com, aes256-gcm@openssh.com, chacha20-poly1305@openssh.com, aes128-cbc, 3des-cbc, blowfish-cbc, cast128-cbc, aes192-cbc, aes256-cbc, arcfour


## Bibliography

Novels

- Dan Brown, Digital Fortress as an exercice to find the numerous technical mistakes about cryptography in the novel
- Neal Stephenson, Cryptonomicon, where the hero use the Solitaire encryption algorithm by Schneier which just needs a deck of card.
Historical
- David Kahn, The Codebreakers;
- Simon Singh, The Code Book (Histoire des Codes Secrets);
- Jacques Stern, La Science du Secret;

Reference

- Steven D. Galbraith, Mathematics of Public Key Cryptography.
- Jeffrey Hoffstein, Jill Pipher et Joseph H. Silverman, An Introduction to Mathematical Cryptography;
- Antoine Joux, Algorithmic Cryptanalysis;
- Alfred J. Menezes, Paul C. van Oorschot et Scott A. Vanstone, Handbook of Applied Cryptography
http://www.cacr.math.uwaterloo.ca/hac/;
- Serge Vaudenay, A Classical Introduction to Cryptography;
- Bruce Schneier, Applied Cryptography;
- Douglas R. Stinson, Cryptography: Theory and Practice.

