Introdução a Criptografia

Diego F. Aranha

Instituto de Computação
UNICAMP

Definições

Definição clássica

Etimologicamente, criptografia é a **arte da escrita secreta**.

Definição moderna

Criptografia é a arte/ciência/engenharia que estuda técnicas para fornecimento de serviços de segurança, como sigilo, autenticação de origem, anonimato, integridade e irretratabilidade, primordialmente em sistemas computacionais.

Criptoanálise

Refere-se ao conjunto de técnicas para **analisar** métodos criptográficos.

$$\text{Criptologia} = \text{Criptografia} + \text{Criptoanálise}$$

Terminologia

Conjuntos:

- Alfabeto de definição $\mathcal{A}$;
- Espaço de mensagens $\mathcal{M}$;
- Espaço de criptogramas $\mathcal{C}$;
- Espaço de chaves $\mathcal{K}$.

Algoritmos:

- Bijeção $E_e : \mathcal{M} \to \mathcal{C}$ parametrizada por chave $e \in \mathcal{K}$;
- Bijeção $D_d : \mathcal{C} \to \mathcal{M}$ parametrizada por chave $d \in \mathcal{K}$;

Sistema criptográfico

Um sistema criptográfico é dado por
$\{\{E_e : e \in \mathcal{K}\}, \{D_d : d \in \mathcal{K}\} | \forall e \in \mathcal{K}, \exists d \in \mathcal{K}, D_d = E_e^{-1}\}$.

Consistência: $\forall m \in \mathcal{M}, D_d(E_e(m)) = m$.

Terminologia

Entidades:

- Participantes;
- Emissor, Receptor;
- Adversário.

Canais:

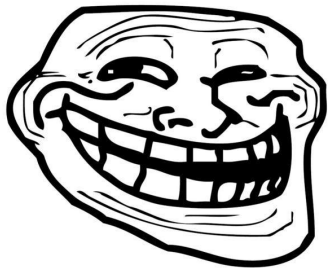
- Inseguro;
- Seguro;
- Fisicamente seguro;

Segurança

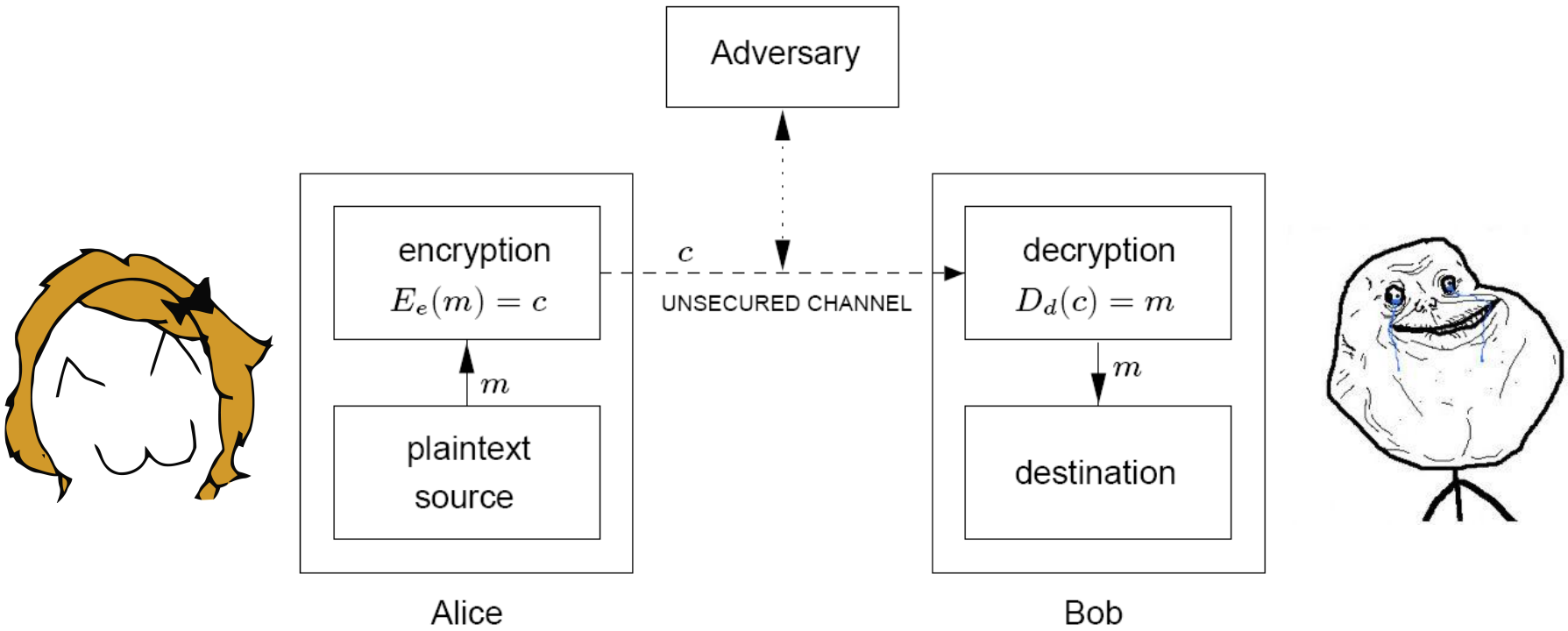
Um sistema criptográfico é **quebrável** se um adversário pode recuperar sistematicamente mensagens a partir de criptogramas sem o conhecimento de (e, d) em um tempo **razoável**.

Importante: Quando o tempo é razoável?

Sistema criptográfico



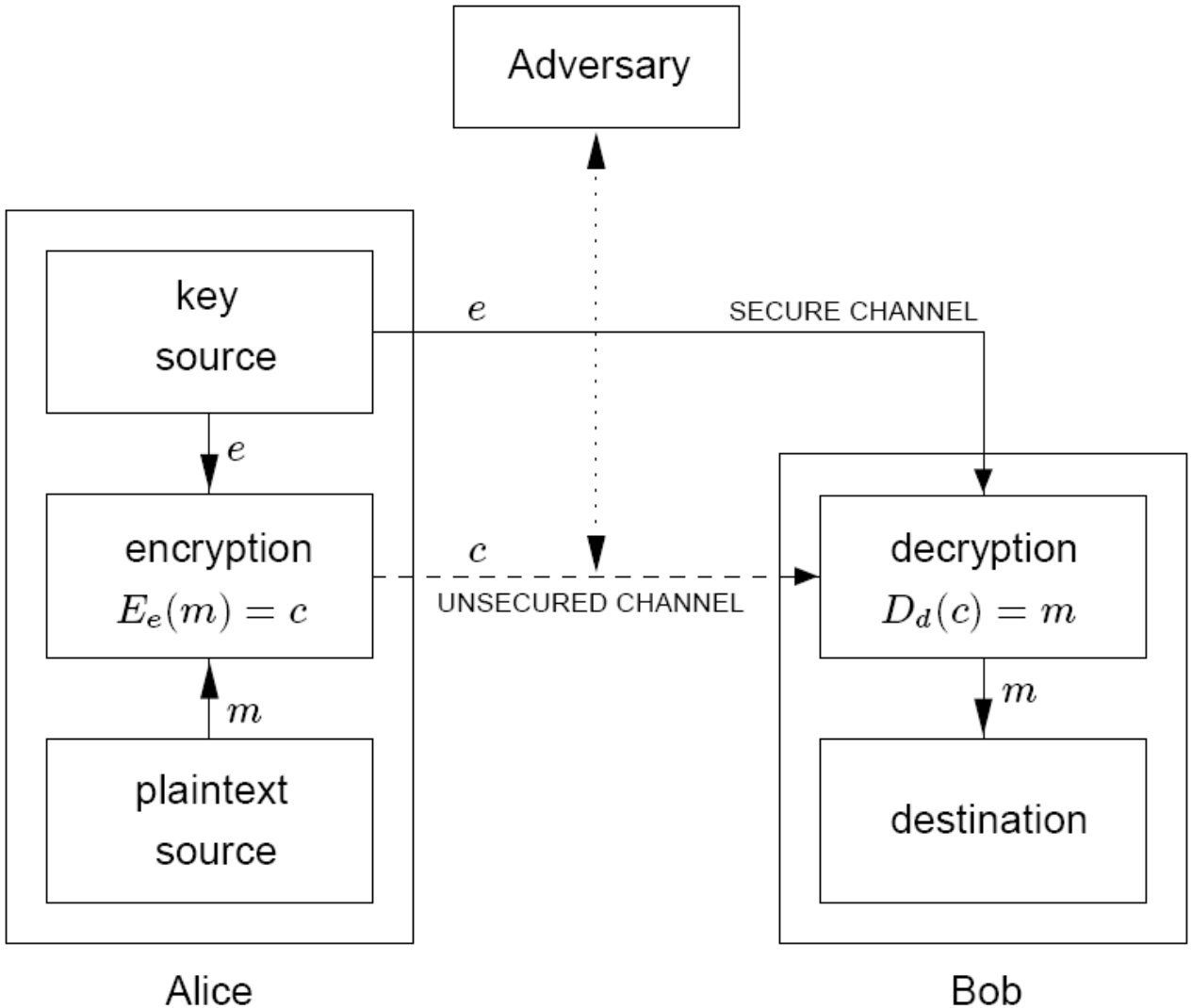
problem?



Classificação: simétrico, assimétrico.

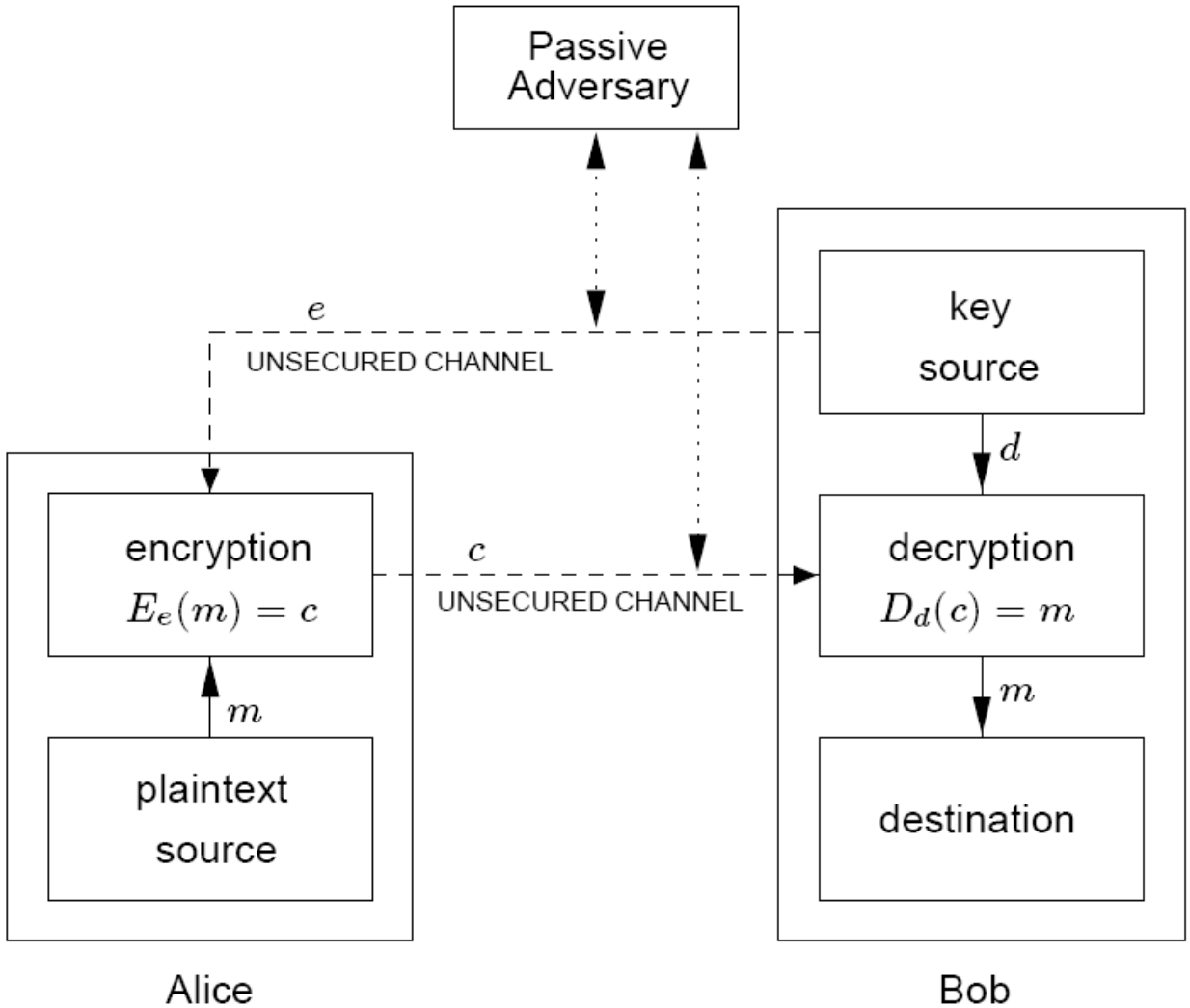
Sistema simétrico

Se é possível derivar d a partir de e em tempo polinomial:



Sistema assimétrico

Se não é possível derivar d a partir de e em tempo polinomial:



Requisitos para sistemas criptográficos

Princípios de Kerckhoffs (*La Cryptographie Militaire*, 1883):

- O sistema deve ser inquebrável na teoria ou prática;
- A divulgação do sistema não deve comprometê-lo;
- A chave deve ser memorizada sem anotações;
- O criptograma deve ser transmissível via telégrafo;
- O sistema deve ser simples e eficiente.

Adaptações:

- O sistema deve ser **seguro**;
- A segurança deve residir nas chaves, e não no algoritmo;
- A chave deve ter tamanho polinomial no nível de segurança;
- O criptograma deve ter tamanho polinomial na mensagem;
- As bijeções devem ser calculadas em tempo polinomial.

Segurança de sistemas criptográficos

Segurança incondicional:

- Criptograma não revela nenhuma informação do texto claro;
- Segredo perfeito.

Segurança por complexidade teórica:

- Adversário tem poder computacional polinomial;
- Modelo de ataque para um certo nível de segurança;
- Demonstração sob premissas plausíveis.

Segurança de sistemas criptográficos

Segurança demonstrável:

- Redução de um problema difícil para a quebra do sistema;
- Premissas nem sempre realistas.

Segurança computacional:

- Custo do melhor ataque conhecido excede poder do adversário hipotético.

Segurança heurística:

- Análise argumentativa baseada na resistência a ataques já conhecidos.

Na prática

Um sistema criptográfico é considerado seguro se o melhor ataque conhecido não é mais eficiente do que a **busca exaustiva** no espaço de chaves.

Ataques a sistemas criptográficos

Ataque de criptograma:

- Adversário tenta obter chave a partir do criptograma apenas.

Ataque de texto claro conhecido:

- Adversário possui texto claro e criptogramas correspondentes.

Ataque de texto claro escolhido:

- Adversário escolhe o texto claro a ser cifrado e recebe criptograma correspondente.

Ataque de criptograma escolhido:

- Adversário pode escolher a decifração de criptogramas.

Cifras de substituição monoalfabética

Definição

Uma **substituição monoalfabética** $E_\pi : \mathcal{M} \to \mathcal{C}$ é uma regra para substituir cada caractere m_i da mensagem m por $\pi(m_i)$, onde π define uma permutação no alfabeto de definição.

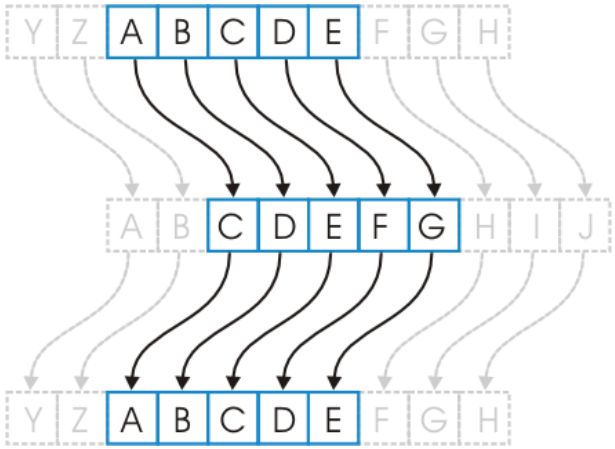
Ou seja:

- A chave é a permutação $\pi : \mathcal{A} \to \mathcal{A}$;
- A função de cifração é $E_\pi(m) = (\pi(m_1), \pi(m_2), \ldots, \pi(m_{|m|})$;
- A função de decifração é $D_\pi(c) = (\pi^{-1}(c_1), \pi^{-1}(c_2), \ldots, \pi^{-1}(c_{|c|})$

Observações:

- O espaço de chaves tem tamanho $(|\mathcal{A}|!)$;
- No caso geral, a chave tem tamanho $|\mathcal{A}|$.

Cifra de César (César, 58 A.C.)



Definição

É um caso especial de substituição monoalfabética onde cada caractere é substituído pelo caractere k posições à frente.

Ou seja:

- A chave é o número k;
- A permutação é dada por $\pi(m_i) = (m_i + k) \bmod |\mathcal{A}|$;

Observações:

- Qual o tamanho do espaço de chaves?

Cifras de transposição

Definição

Uma **transposição** $E_\theta : \mathcal{M} \to \mathcal{C}$ troca a posição i de cada caractere m_i da mensagem m por $\theta(i)$, onde θ define uma permutação no conjunto $\{1, 2, \ldots, |m|\}$.

Ou seja:

- A chave é a permutação $\theta : \{1, 2, \ldots, |m|\} \to \{1, 2, \ldots, |m|\}$;
- A função de cifração é $E_\theta(m) = (m_{\theta(1)}, m_{\theta(2)}, \ldots, m_{\theta(|m|)})$;
- A função de decifração é $D_\theta(c) = (c_{\theta^{-1}(1)}, c_{\theta^{-1}(2)}, \ldots, c_{\theta^{-1}(|m|)})$;

Observações:

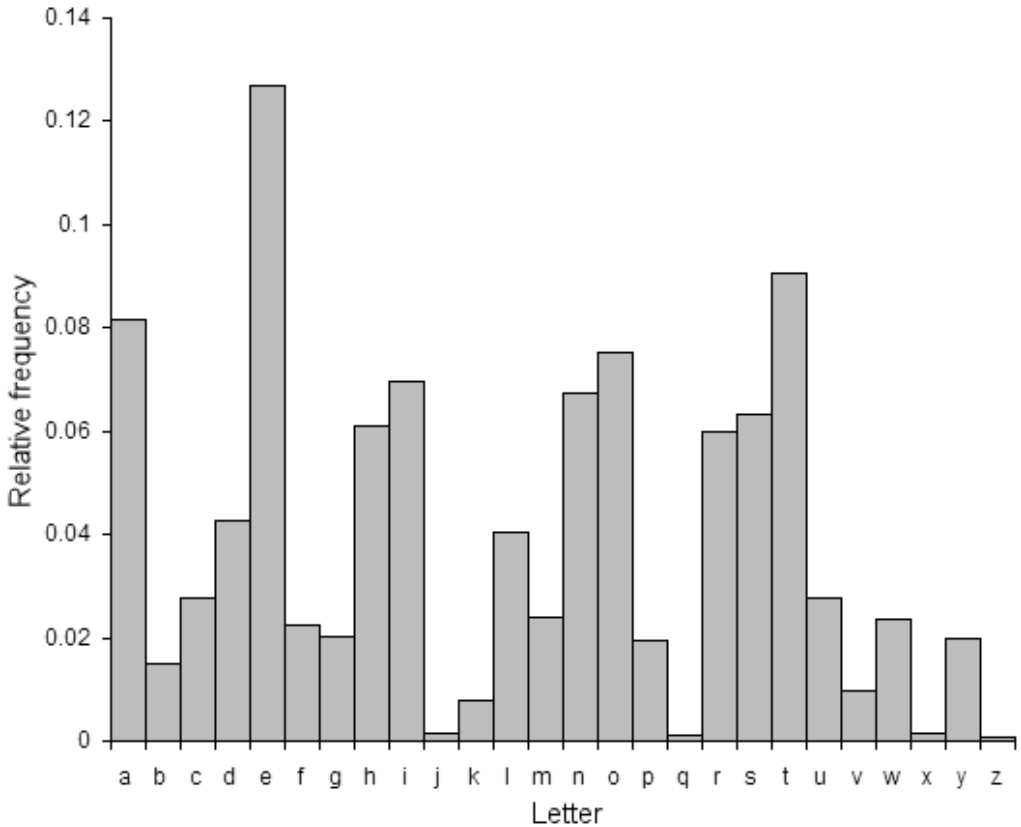
- O espaço de chaves tem tamanho $(|m|!)$;
- No caso geral, a chave tem tamanho $|m|$.

Scytale (Esparta, 500 A.C.)

Criptoanálise por análise de freqüência (Al-Kindi, 800 D.C.)

Al-Kindi escreve "Manual para decifração de mensagens cifradas":

- Transposição preserva as freqüências exatas dos caracteres no texto original;
- Substituição monoalfabética permuta as freqüências dos caracteres.



Cifra de substituição polialfabética

Definição

Uma **substituição polialfabética** mapeia conjuntos disjuntos de caracteres com permutações π_i distintas.

Ou seja:

- A chave é o conjunto de permutações $\Pi = (\pi_1, \pi_2, \ldots, \pi_t)$;
- A função de cifração é
  $E_\Pi(m) = (\pi_1(m_1), \pi_2(m_2), \ldots, \pi_{1+|m| \bmod t}(m_{|m|}))$;
- A função de decifração é análoga.

Observações:

- A freqüência dos símbolos é distorcida!

Cifra de Vigenère (Vigenère, 1586)

Entitulada "le chiffre indéchiffrable", por permanecer intratável por quase 300 anos!

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Cifras de fluxo (Vernam, 1919)

Definição

Uma **cifra de fluxo** é uma cifra de substituição polialfabética com o número de permutações idêntico ao tamanho da mensagem de entrada.

Ou seja:

- A chave é uma cadeia de caracteres $(k_1, k_2, \ldots, k_{|m|})$;
- Cada caractere da chave define uma permutação distinta;
- A função de cifração pode ser vista como uma "soma" caractere à caractere entre a chave e a mensagem.
  Exemplo:

```
Chave:        ABCDEF GH IJKLM NOP QRSTUVWXYZ12
Mensagem:     CRYPTO IS SHORT FOR CRYPTOGRAPHY
Criptograma:  CSASTP KV SIQUT GQU DFMATPIUAQJB
```

One-time pad (Vernam, 1925)

Definição

Um **one-time pad** é uma cifra de fluxo onde a chave é selecionada aleatoriamente e nunca é repetida.

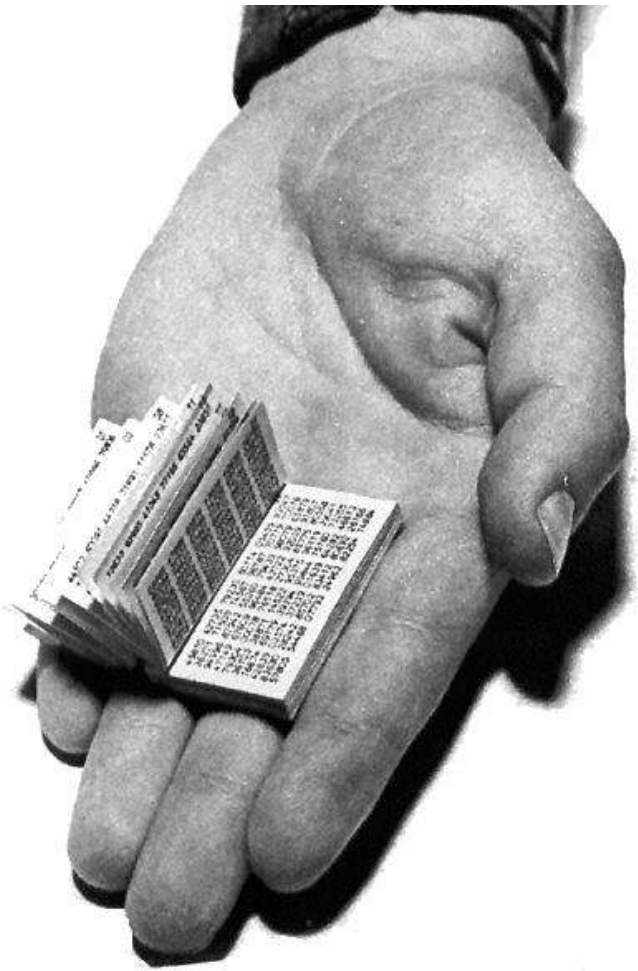
Talvez a maior contribuição dada à criptografia!

Desvantagens:

- Gerar dados verdadeiramente aleatórios é difícil;
- Distribuir chave de tamanho arbitrário é difícil.

Observação: Principal forma conhecida de segurança incondicional!

One-time pad (Vernam, 1925)



Produto de cifras (Shannon, 1949)

Shannon observou que:

- Composição de cifras já era usada por intuição;
- Cifras de substituição e transposição não são seguras;
- Cifras de substituição desligam o criptograma da chave;
- Cifras de transposição espalham redundância da mensagem.

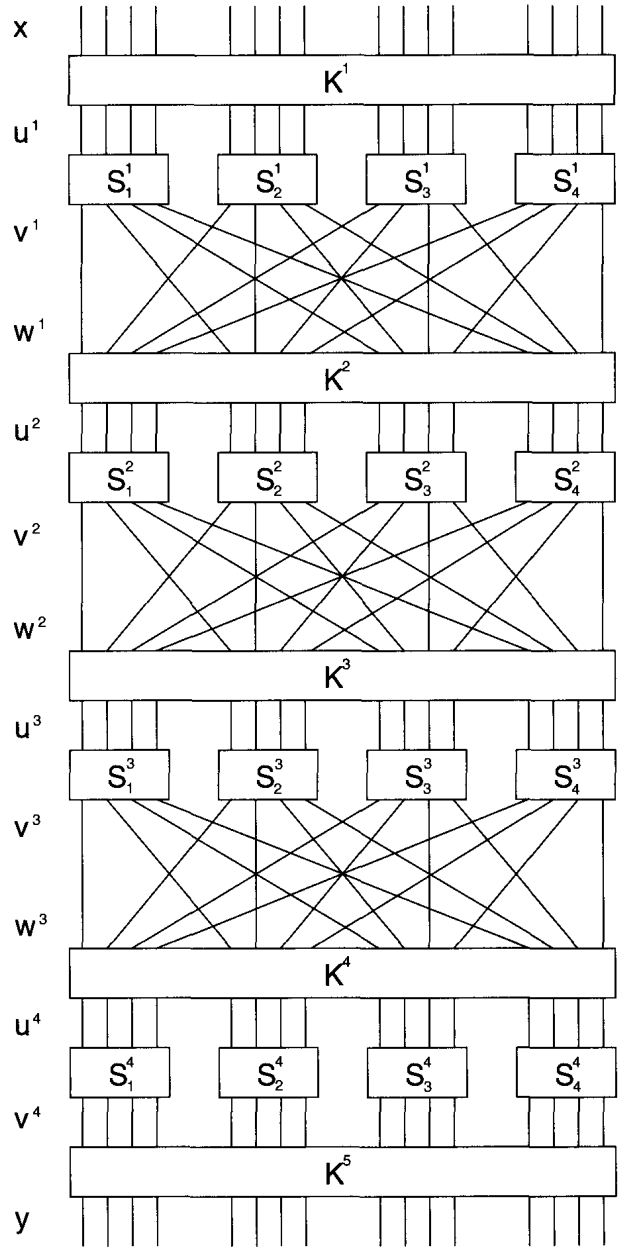
Conclusão: composições dessas cifras podem ser mais seguras!

Definição

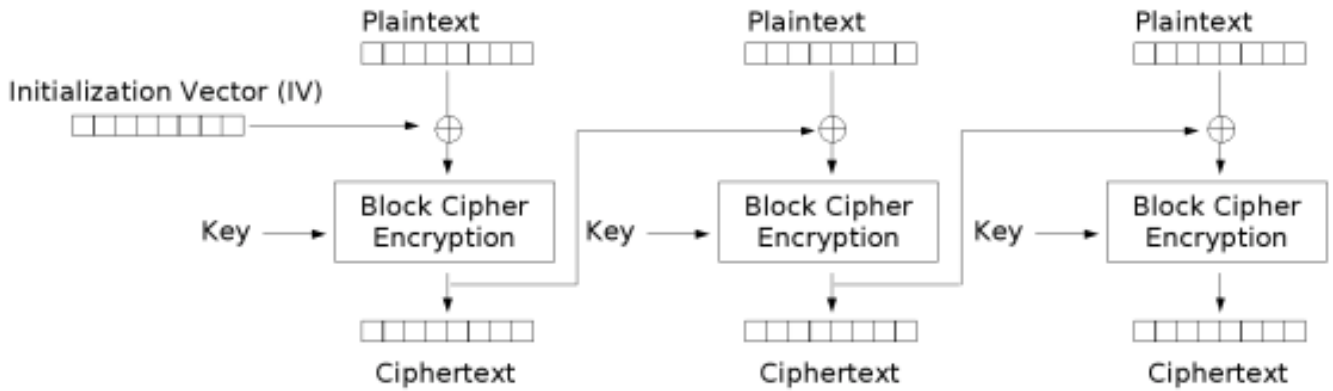
Um **produto de cifras** é a composição de t funções de cifração $E_{k_1}, E_{k_2}, \ldots, E_{k_t}$, onde cada função E_{k_i} é uma substituição ou transposição.

Cifras de substituição adicionam **confusão** e cifras de transposição adicionam **difusão** ao processo de cifração.

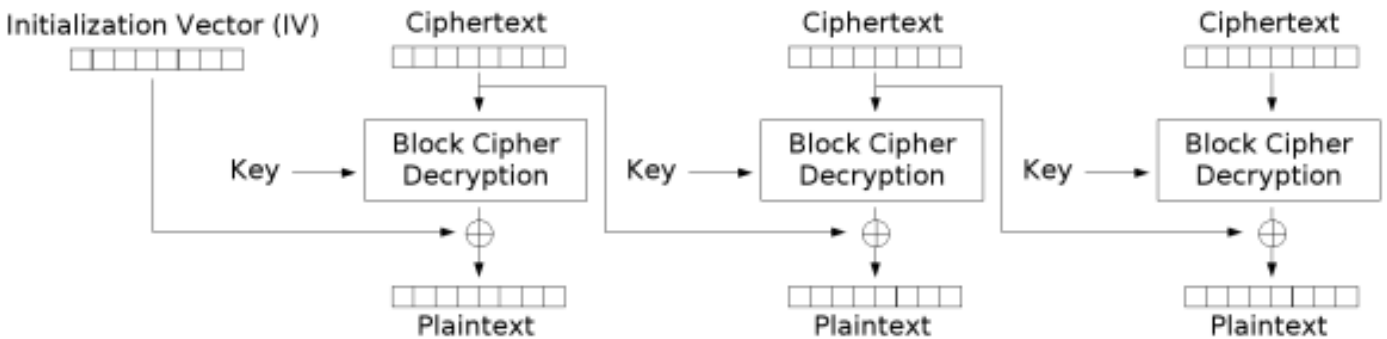
Cifras de bloco modernas



Cipher Block Chaining (CBC)

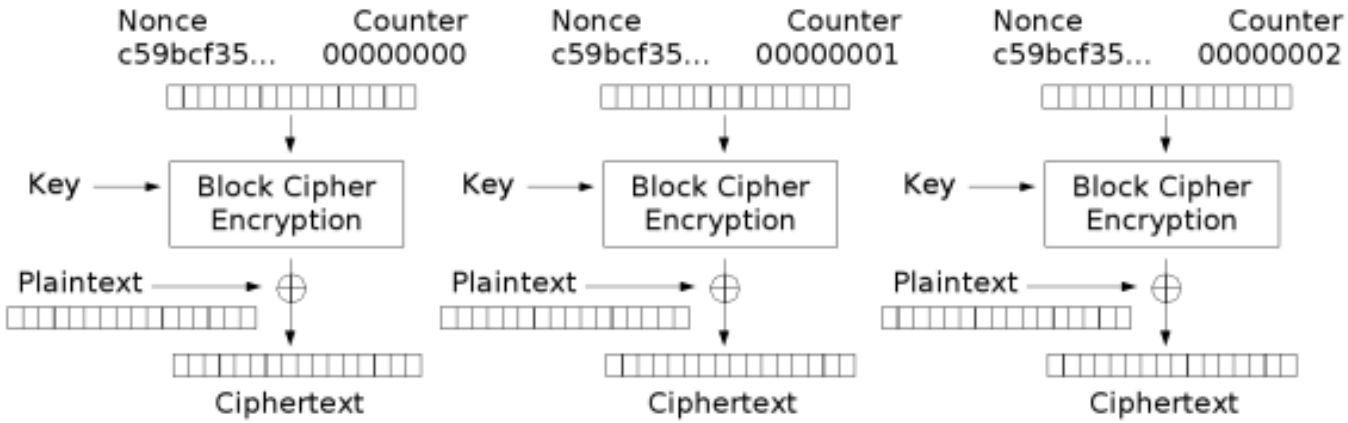


Cipher Block Chaining (CBC) mode encryption

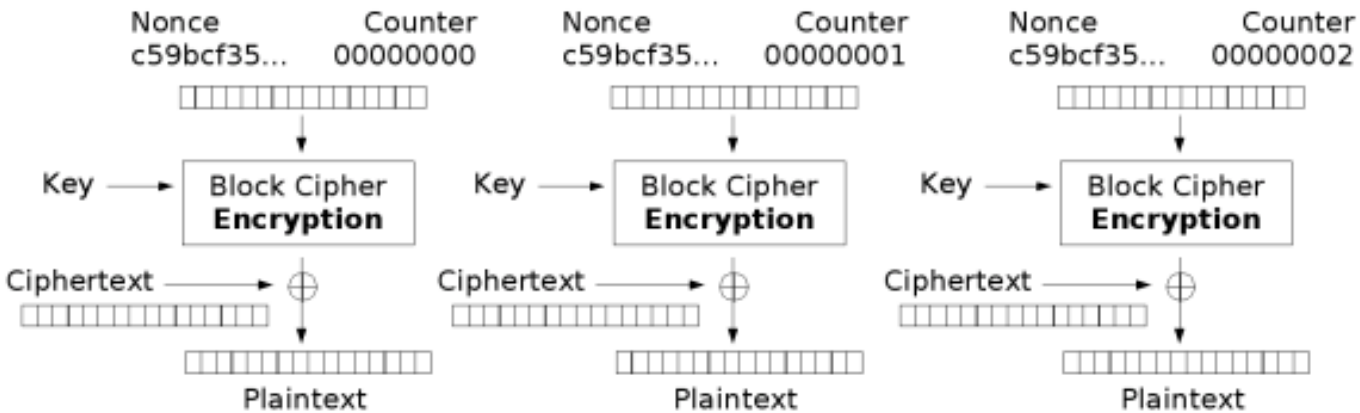


Cipher Block Chaining (CBC) mode decryption

Counter Mode (CTR)



Counter (CTR) mode encryption



Counter (CTR) mode decryption

Código de Autenticação de Mensagem (MAC)

Definição

É um autenticador de mensagem produzido a partir de criptografia simétrica (cifra ou função de resumo criptográfico). É normalmente utilizado para se construir modos de operação de *cifração autenticada*.

Importante: Por que não é assinatura digital?

Modos de cifração autenticada:

- *Counter with Cipher Block Chaining (CCM)*: CTR + CBC-MAC;
- *Galois Counter Mode (GCM)*.

Código de Autenticação de Mensagem (MAC)

Definição

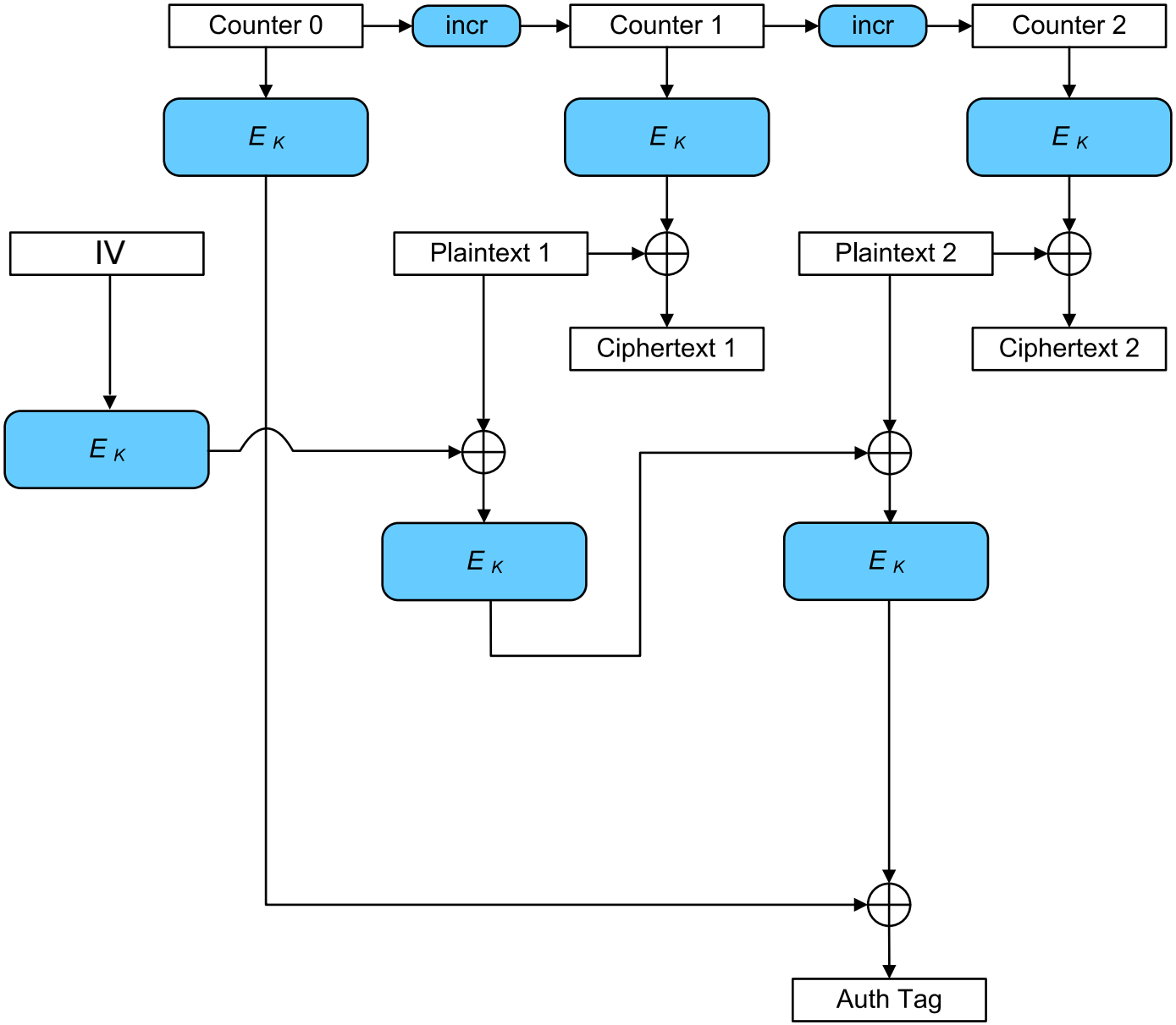
É um autenticador de mensagem produzido a partir de criptografia simétrica (cifra ou função de resumo criptográfico). É normalmente utilizado para se construir modos de operação de *cifração autenticada*.

Importante: No mínimo duas entidades possuem chave secreta!

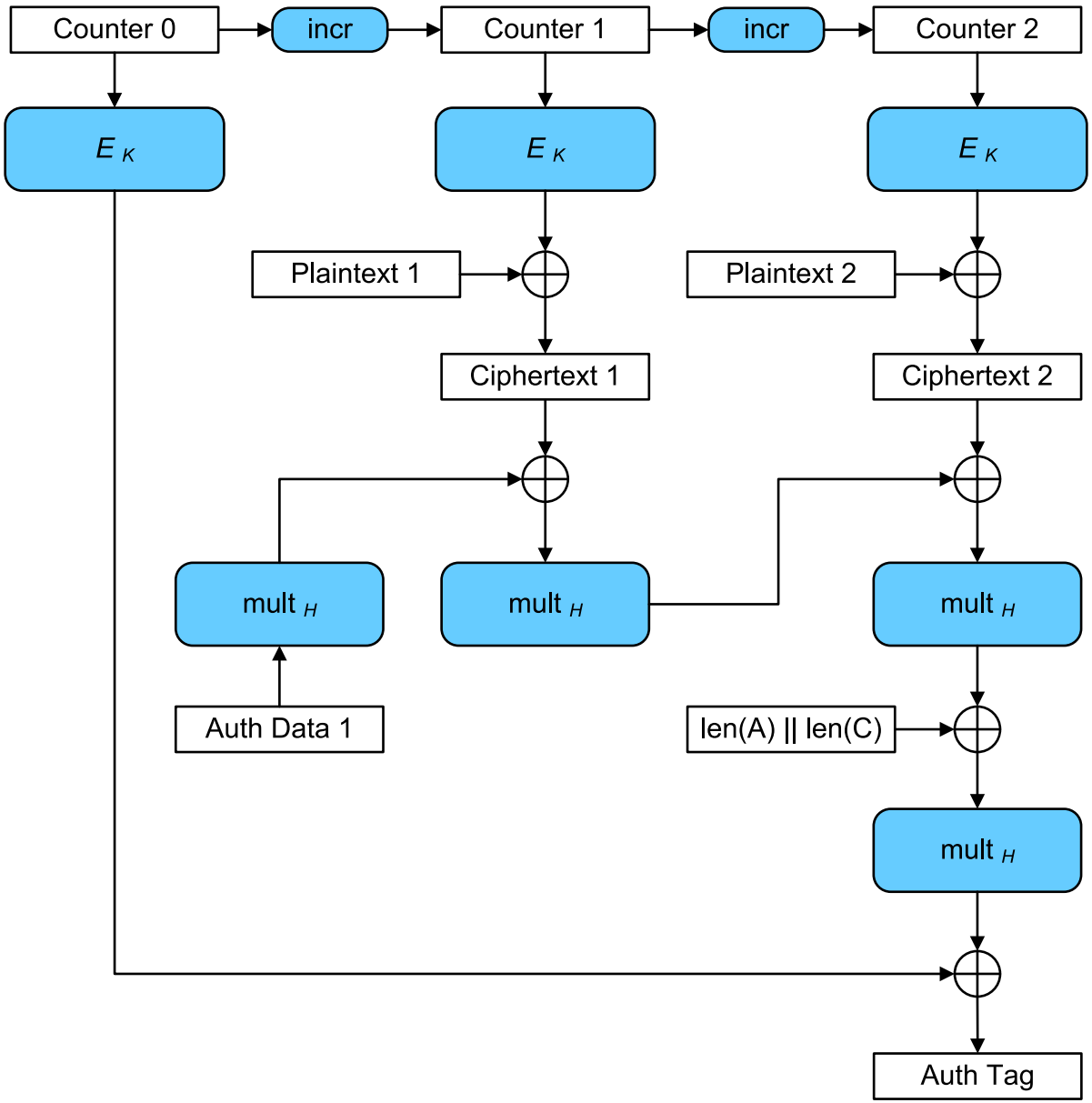
Modos de cifração autenticada:

- *Counter with Cipher Block Chaining (CCM)*: CTR + CBC-MAC;
- *Galois Counter Mode (GCM)*.

Counter with Cipher Block Chaining (CCM)



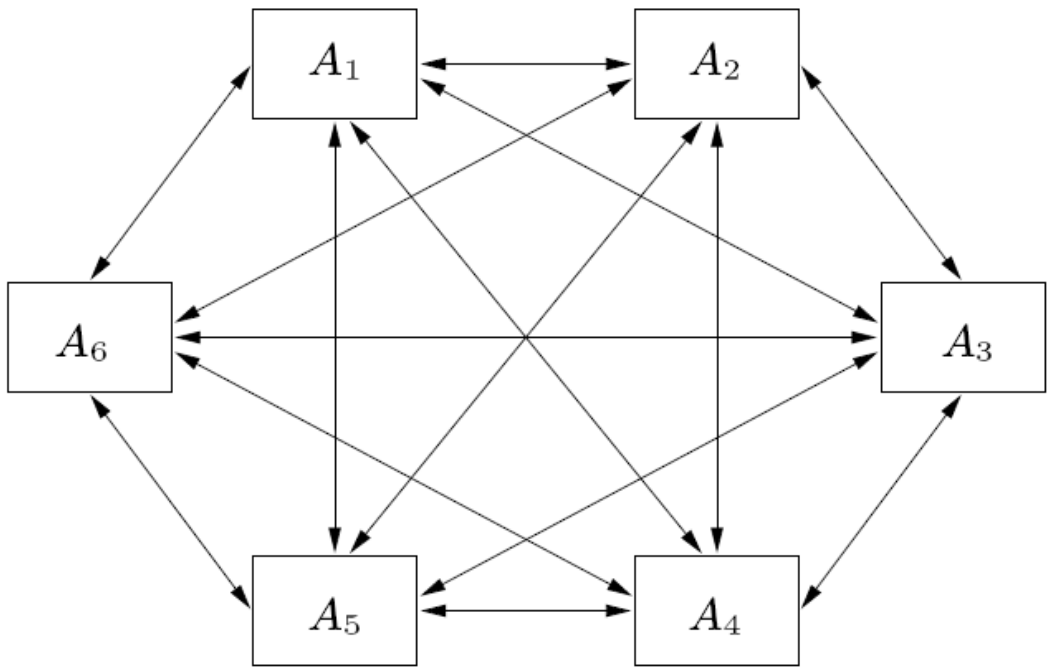
Galois Counter Mode (GCM)



Problema da distribuição de chaves

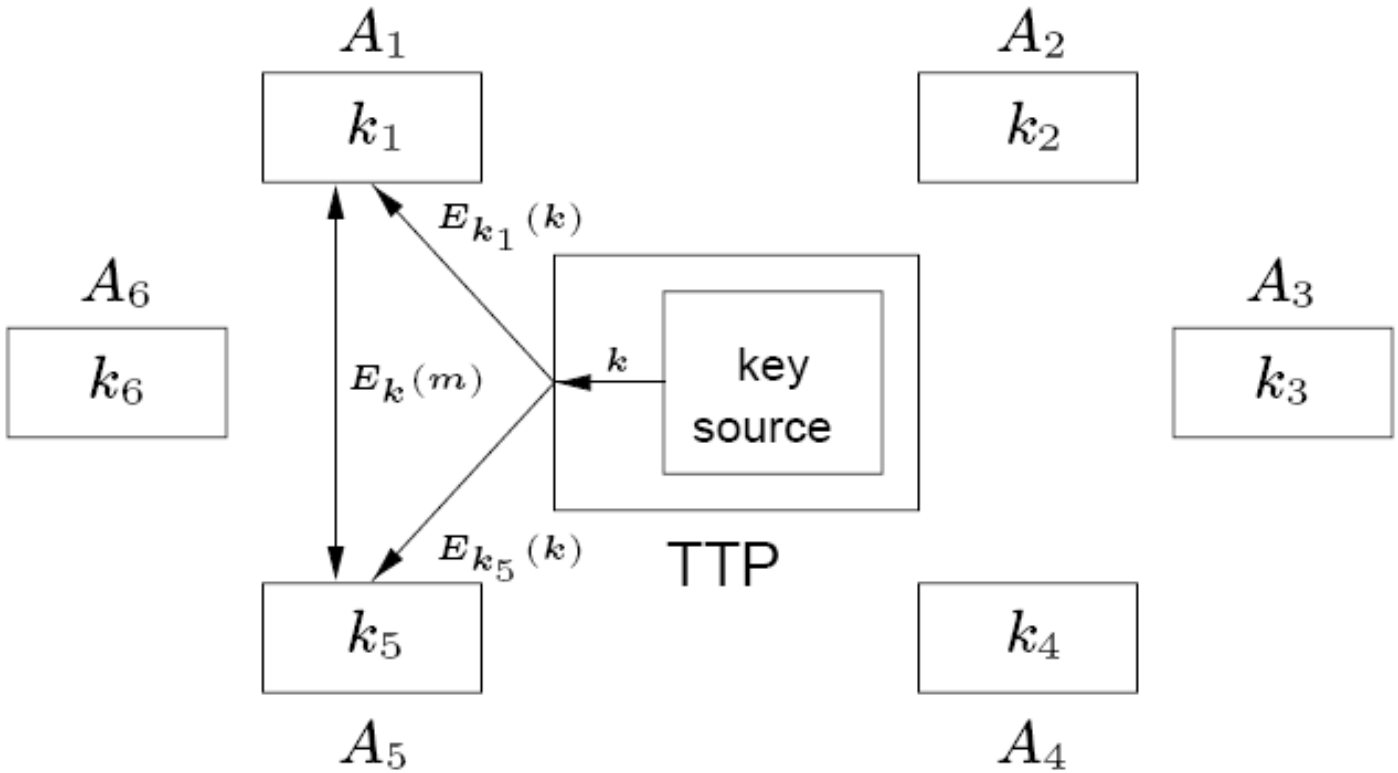
Cifras simétricas exigem o compartilhamento de segredo.

Como estabelecer chaves compartilhadas com todos os usuários que se deseja comunicar?



Problema da distribuição de chaves

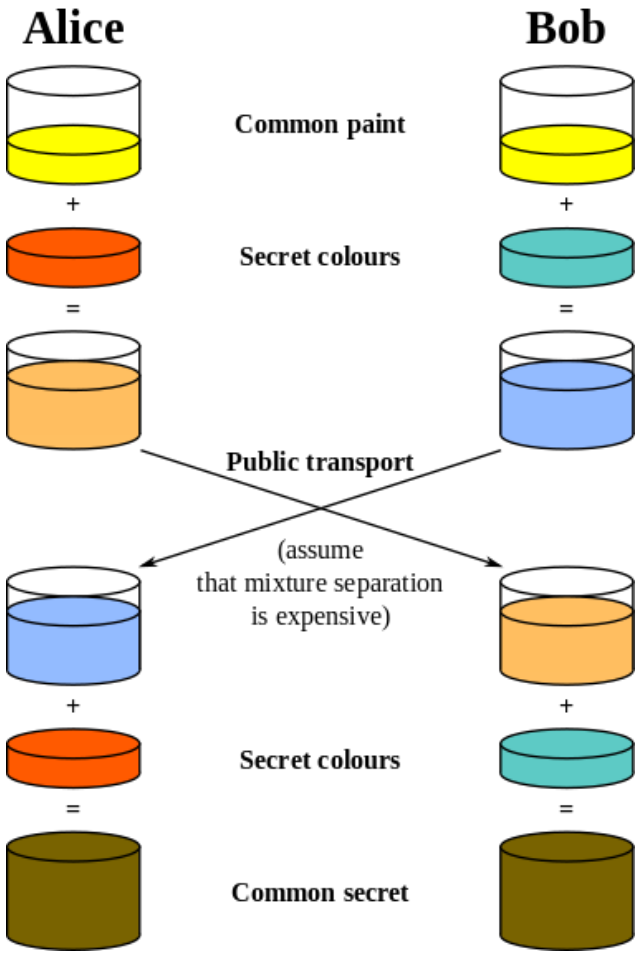
Podemos confiar em uma entidade que produz chaves efêmeras.



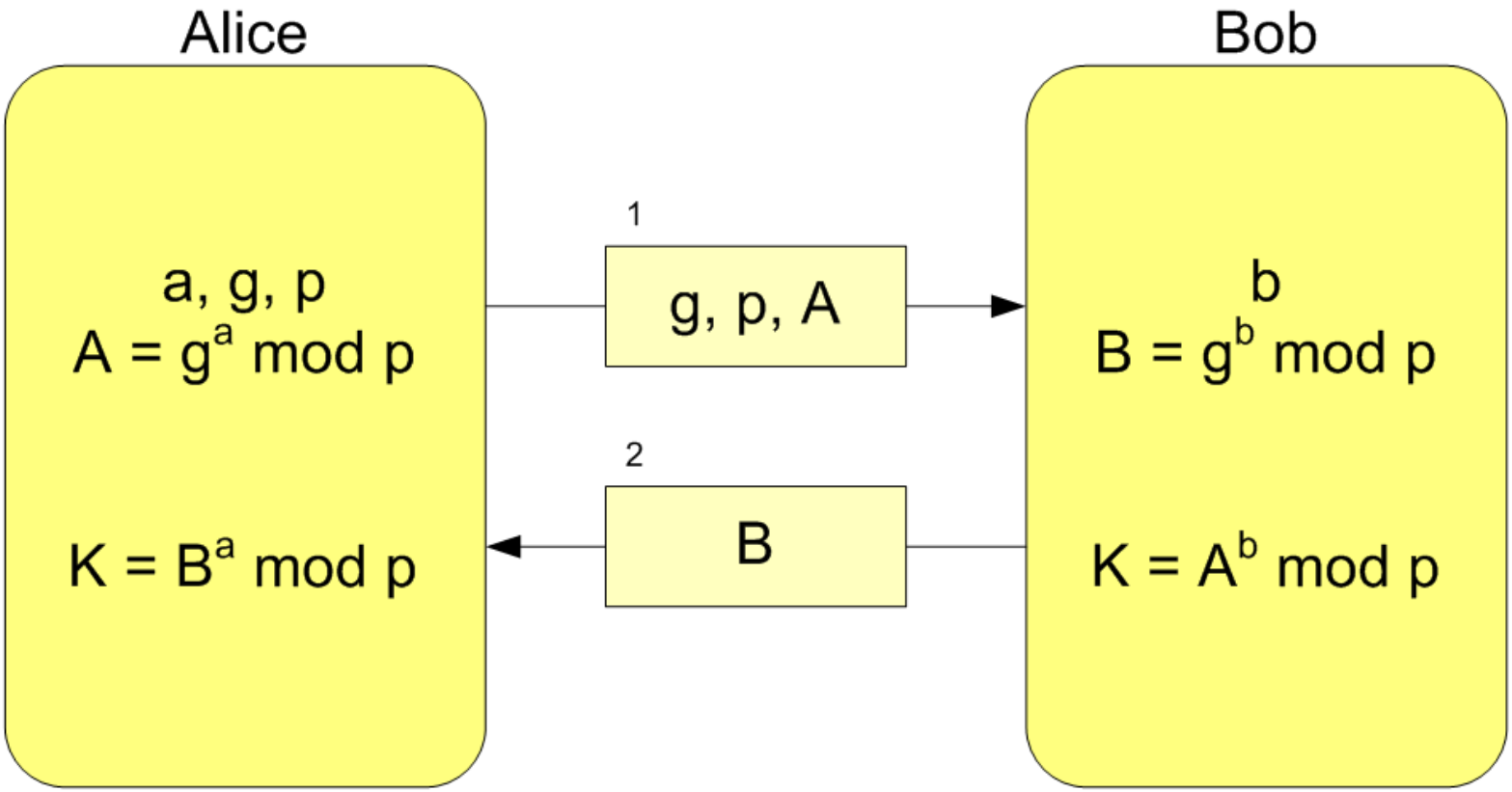
Problema: Mas e se não tivermos em quem confiar?

Criptografia assimétrica (Diffie, Hellman, 1976)

Talvez a segunda maior contribuição dada à criptografia!



Criptografia assimétrica (Diffie, Hellman, 1976)



$K = A^b \bmod p = (g^a \bmod p)^b \bmod p = g^{ab} \bmod p = (g^b \bmod p)^a \bmod p = B^a \bmod p$

Premissa: Recuperar x a partir de $g^x \bmod p$ é difícil!

Criptografia assimétrica (Rivest,Shamir,Adleman, 1977)

Primeira realização de criptografia assimétrica que permite cifração e assinatura!

Geração de chaves:

1. Escolher dois primos grandes p e q;
2. Calcular o módulo $n = pq$ e $\phi(n) = (p-1)(q-1)$;
3. Escolher o inteiro $1 < e < \phi(n)$ como a chave pública;
4. Calcular a chave privada $d = e^{-1} \bmod \phi(n)$.

Cifração: Calcular $c = m^e \bmod n$.

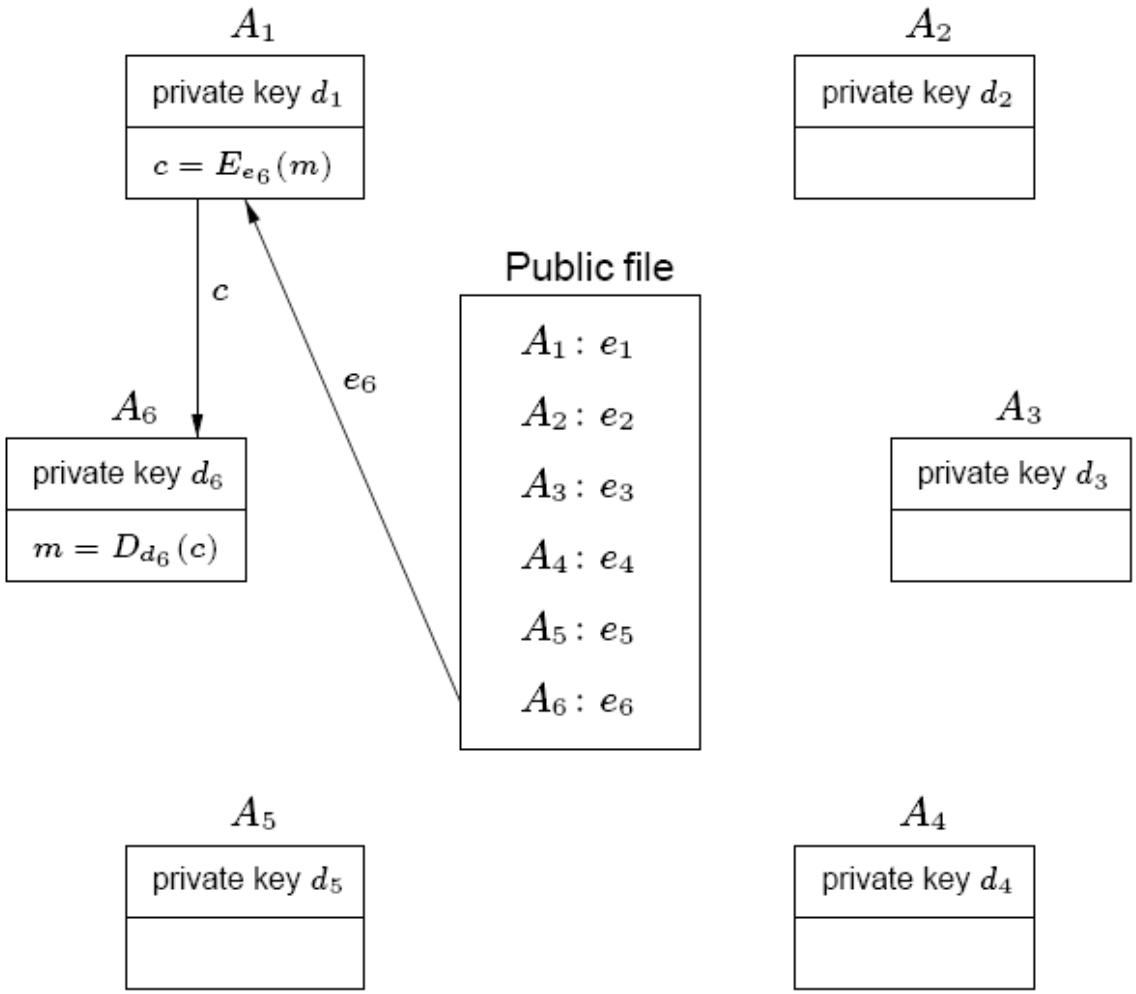
Decifração: Calcular $m = c^d \bmod n$.

Consistência: $c^d \equiv (m^e)^d \equiv m^{ed} \equiv m \bmod n$.

Premissa: Fatorar n em p e q é difícil!

Problema da distribuição de chaves

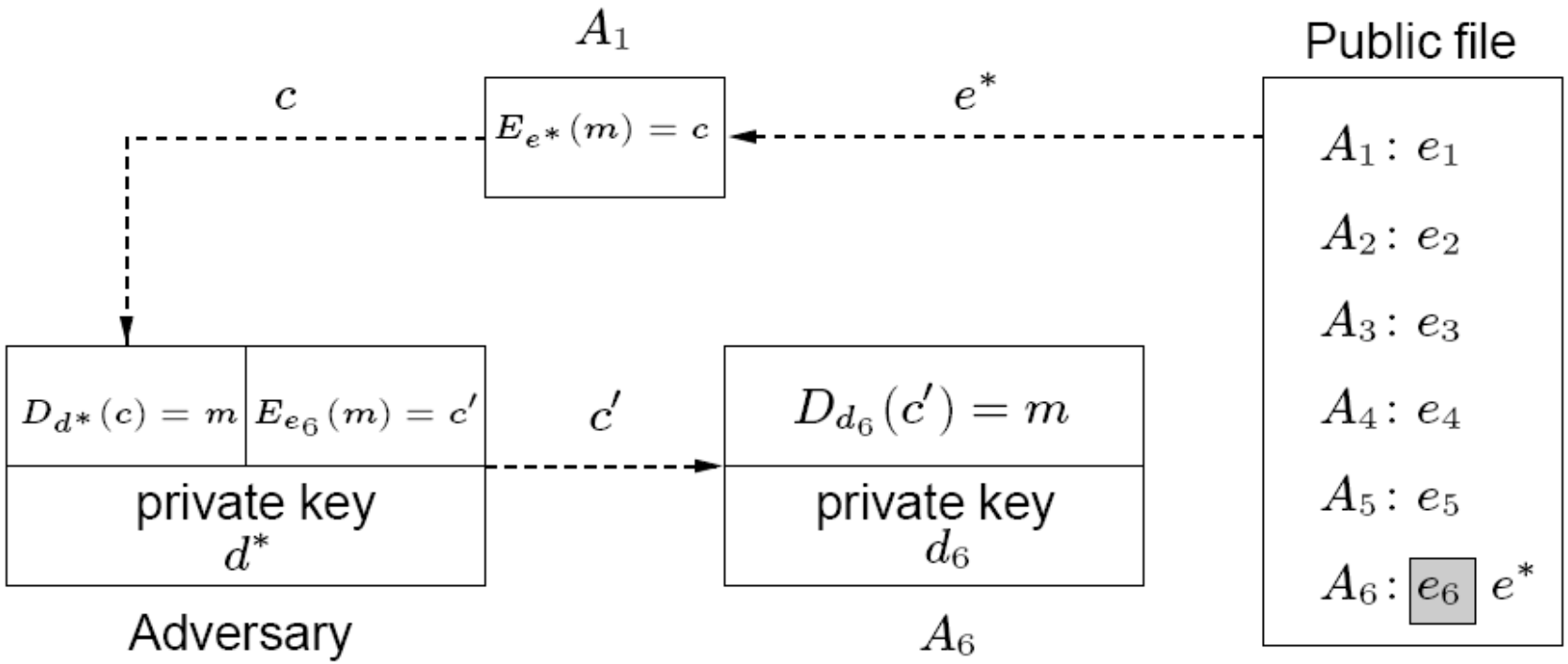
Criptografia assimétrica resolve o problema!



Problema: O repositório precisa ser confiado!

Problema da autenticação de chaves públicas

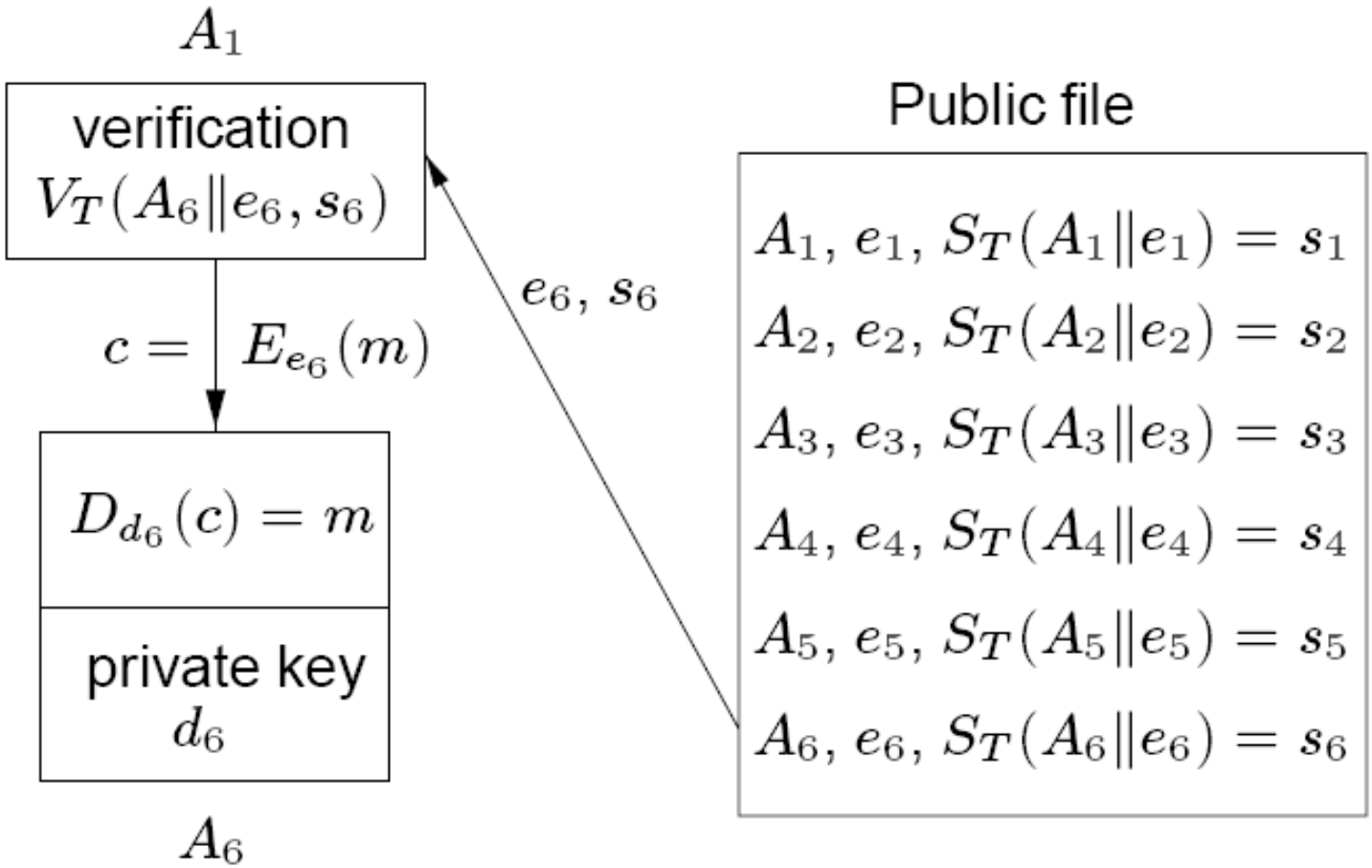
Criptografia assimétrica cria um novo problema!



Problema: Precisamos novamente de alguém pra confiar!

Certificados de titularidade (Kohnfelder, 1979)

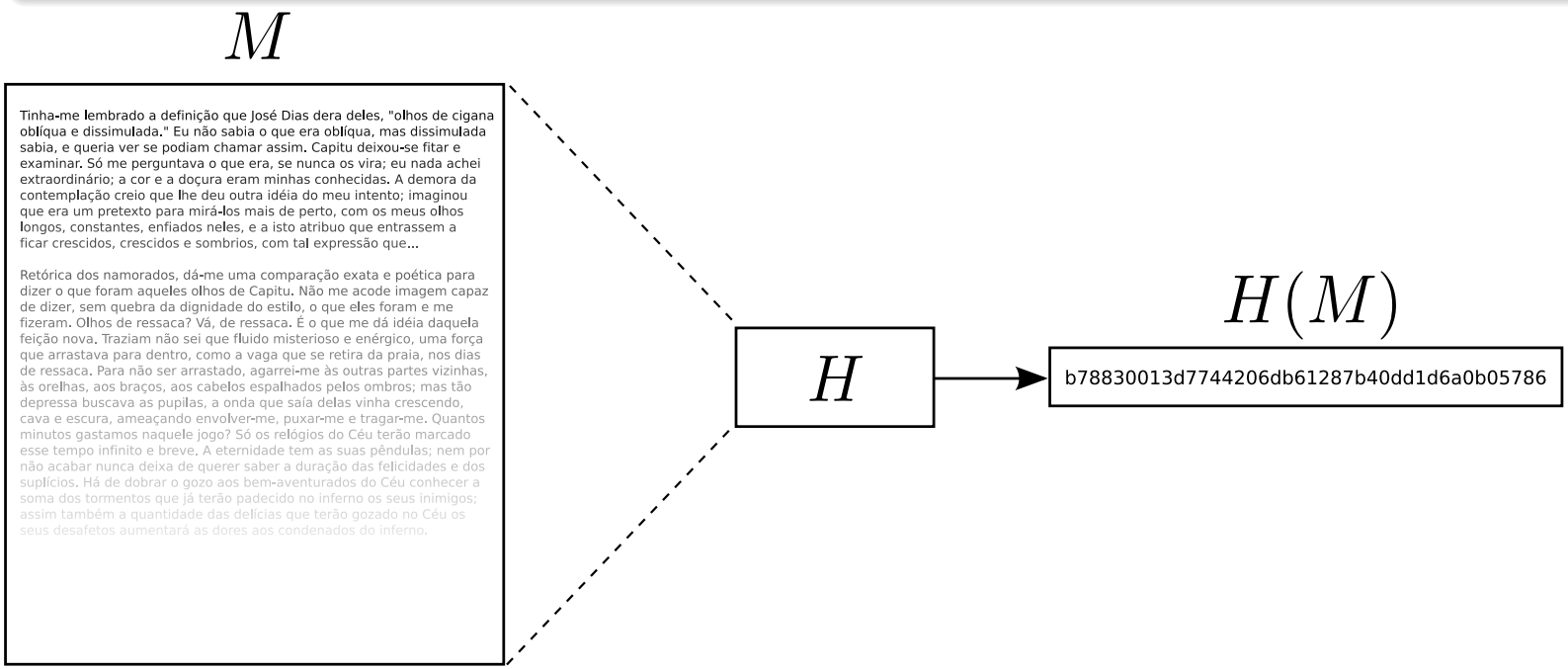
Solução: Autoridade confiada autentica chaves públicas!



Funções de *hash* criptográficas

Definição

Uma **função de hash criptográfica** é uma função computacionalmente eficiente que mapeia cadeias arbitrárias de *bits* em cadeias de tamanho fixo.



Funções de *hash* criptográficas

Uma função de *hash* criptográfica h deve ser:

- **Resistente à inversão**: difícil recuperar m a partir de $h(m)$;
- **Livre de colisão**: difícil encontrar m' com $h(m) = h(m')$;
- **Resistente à colisão**: difícil encontrar m e m' com $h(m) = h(m')$.

Funções de *hash* criptográficas

Uma função de *hash* criptográfica h deve ser:

- **Resistente à inversão**: difícil recuperar m a partir de $h(m)$;
- **Livre de colisão**: difícil encontrar m' com $h(m) = h(m')$;
- **Resistente à colisão**: difícil encontrar m e m' com $h(m) = h(m')$.

Diferentes aplicações:

- Armazenamento de senhas (armazenar $h(s)$ ao invés de s).
- Derivação de chaves ($k = h(g^{xy} \bmod p)$, $k_i = h(k_{i-1})$).
- Verificação de integridade ($y = h(x)$).
- Assinaturas digitais (sign $h(m)$ instead of just m).
- Message Authentication Codes (MACs) ($y = h_K(x)$).

Importante: Pelo Paradoxo do Aniversário, segurança de $n/2$ *bits* para resumo com n *bits*!

Assinaturas digitais

Definição

Uma **assinatura digital** é uma técnica criptográfica para adicionar um autenticador *irretratável* publicamente verificável a uma mensagem.

Conjuntos:

- Espaço de mensagens $\mathcal{M}$;
- Espaço de assinaturas $\mathcal{S}$;
- Espaço de chaves $\mathcal{K}$.

Algoritmos:

- Função de assinatura $S_A(m) = D_d(h(m)) = s$;
- Função de verificação $V_A(s) = 1$ sse $E_e(s) = h(m)$.

Assinatura digital RSA (Rivest,Shamir,Adleman, 1977)

Geração de chaves:

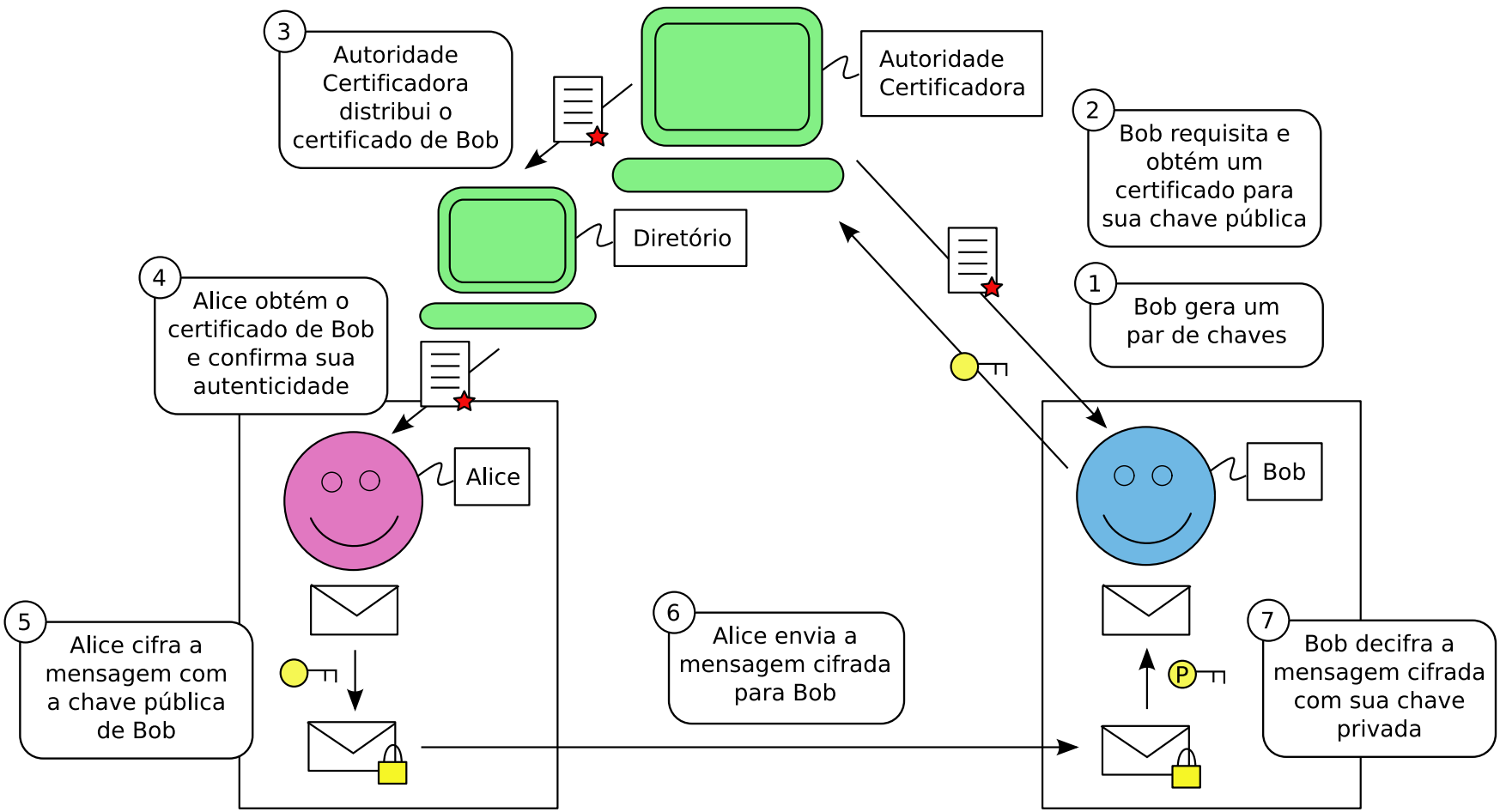
1. Escolher dois primos grandes p e q;
2. Calcular o módulo $n = pq$ e $\phi(n) = (p-1)(q-1)$;
3. Escolher o inteiro $1 < e < \phi(n)$ como a chave pública;
4. Calcular a chave privada $d = e^{-1} \bmod \phi(n)$.

Assinatura: Calcular $s = h(m)^d \bmod n$.

Verificação: Calcular $h' = s^e \bmod n$ e verificar se $h' = h(m)$.

Premissa: Fatorar n em p e q é difícil!

Infra-estruturas de chave pública



Comparação

Vantagens da criptografia simétrica:

- Alto desempenho;
- Chaves mais curtas;
- Tem sido explorada há mais tempo.

Desvantagens da criptografia simétrica:

- Compartilhamento de segredo;
- Problema da distribuição de chaves;
- Impossível fornecer irretratabilidade.

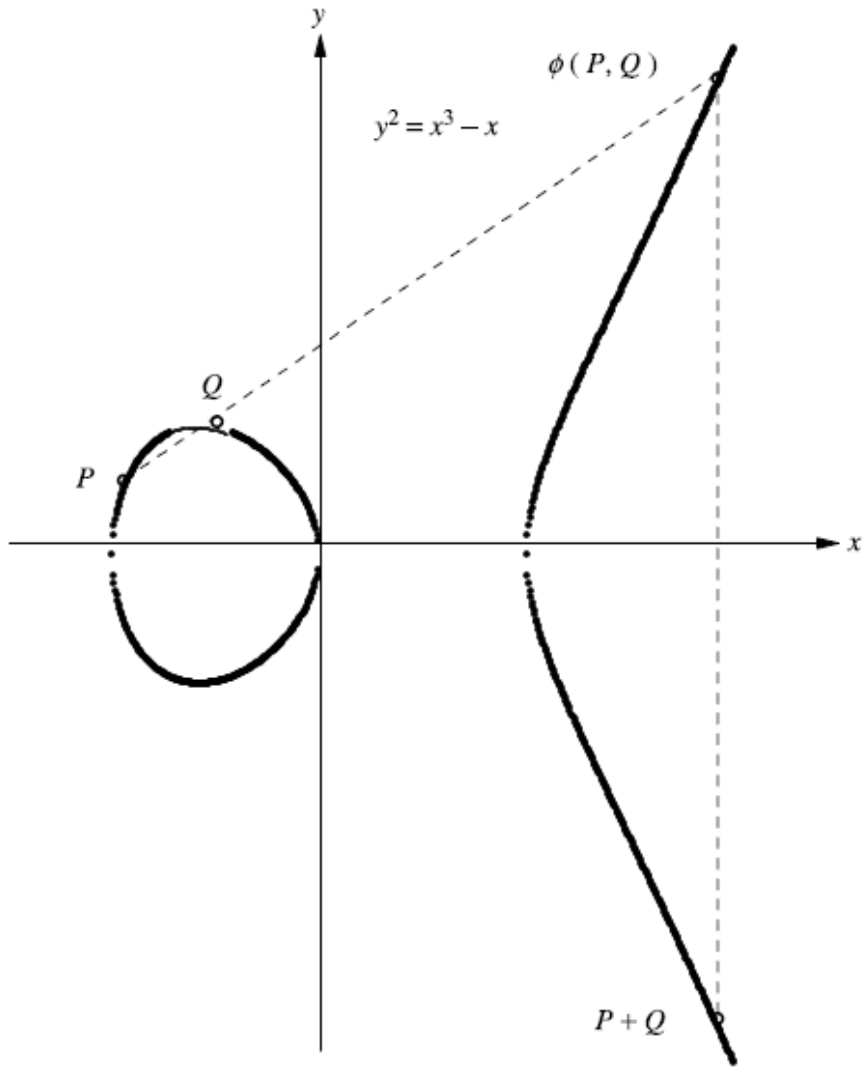
Vantagens da criptografia assimétrica:

- Apenas a chave privada precisa ser secreta;
- Autoridade confiada com menor poder;
- Assinaturas digitais eficientes.

Desvantagens da criptografia assimétrica:

- Baixo desempenho;
- Chaves longas;
- Baseia-se na dificuldade aparente de problemas.

Criptografia de curvas elípticas (Koblitz, 1986)



Premissa: Calcular k a partir de kP é difícil!

Gabarito de algoritmos

1. **Cifra de bloco:** AES em modo CBC
2. **Cifra de fluxo:** AES em modo CTR
3. **Autenticador:** HMAC
4. Cifração autenticada: AES em modo CCM/GCM
5. **Função de hash:** SHA-2 ou SHA-3
6. **Cifração assimétrica:** RSA PKCS #1 v2.1
7. **ASsinatura Digital:** RSA PKCS #1 v2.1
8. **Acordo de chaves:** Curve25519