Crittografia Moderna

Prof. Paolo D'Arco

Introduzione

Dal vocabolario Treccani:

"Crittografia: l'insieme delle teorie e delle tecniche (manuali, meccaniche o elettroniche) che permettono di **cifrare** un testo in chiaro …"

Dal Concise Oxford English Dictionary:

"Cryptography: the **art** of **writing or solving codes**"

Definizioni storicamente corrette ma … non catturano l'ampiezza attuale!



Introduzione: oltre la “confidenzialità”

- Meccanismi per assicurare l’**integrità** dei dati
  - esser certi che i dati trasmessi o memorizzati non siano stati modificati

- Meccanismi per assicurare l’**autenticità** dei dati e delle parti che entrano in gioco in un protocollo
  - esser certi dell’origine dei dati stessi
  - esser certi della parte con cui stiamo comunicando

e, ancora …



Introduzione: oltre la “confidenzialità”

- Protocolli per funzionalità generali e funzionalità specifiche:
  - Schemi per lo cambio di chiavi crittografiche
  - Schemi per la condivisione di segreti
  - Sistemi di prova a conoscenza zero
  - Calcolo sicuro di funzioni multi-party (n var. – n parti)
    - E-voting
    - Giochi on-line
  - Database distribuiti di transazioni autenticate
    - Crittomonete, notaio digitale, smart contract …



Crittografia moderna

"Lo studio delle tecniche matematiche utili a proteggere l'informazione digitale, i sistemi di elaborazione e le computazioni distribuite da attacchi avversari"

- Tra gli anni 80' e 90': transizione da "arte" a "scienza"



Natura della transizione

Insieme di strumenti euristici che garantiscono comunicazioni segrete a militari e diplomatici



Scienza che aiuta a proteggere i sistemi digitali usati da persone ordinarie sparse su tutto il globo terrestre



Scopo del corso

Presentare i principi alla base della Crittografia Moderna.

Al termine dovreste essere in grado di:

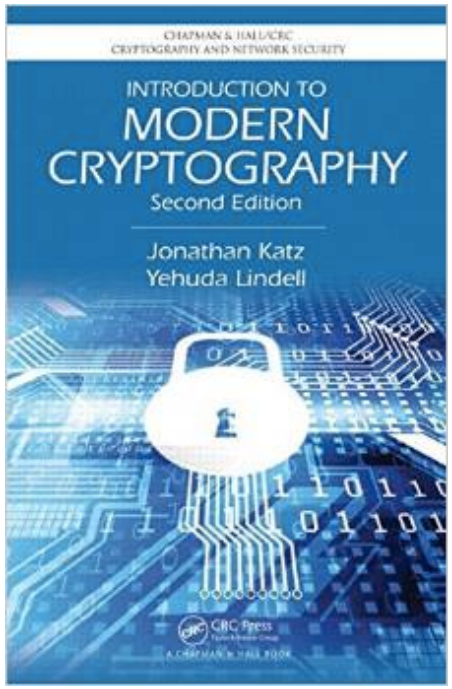
- **apprezzare** le **garanzie** che le primitive crittografiche offrono in termini di sicurezza
- **conoscere** ed avere una discreta familiarità con le **costruzioni standard**
- **effettuare valutazioni di base** di primitive e protocolli crittografici

Approccio: studio rigoroso e strutturato



Testo di riferimento principale

J. Katz and J. Lindell, **Introduction to Modern Cryptography,** CRC Press, 2015



Più letture specifiche



Crittografia Classica: qualche esempio

- Ci serve per motivare l'approccio più rigoroso che applicheremo nel seguito

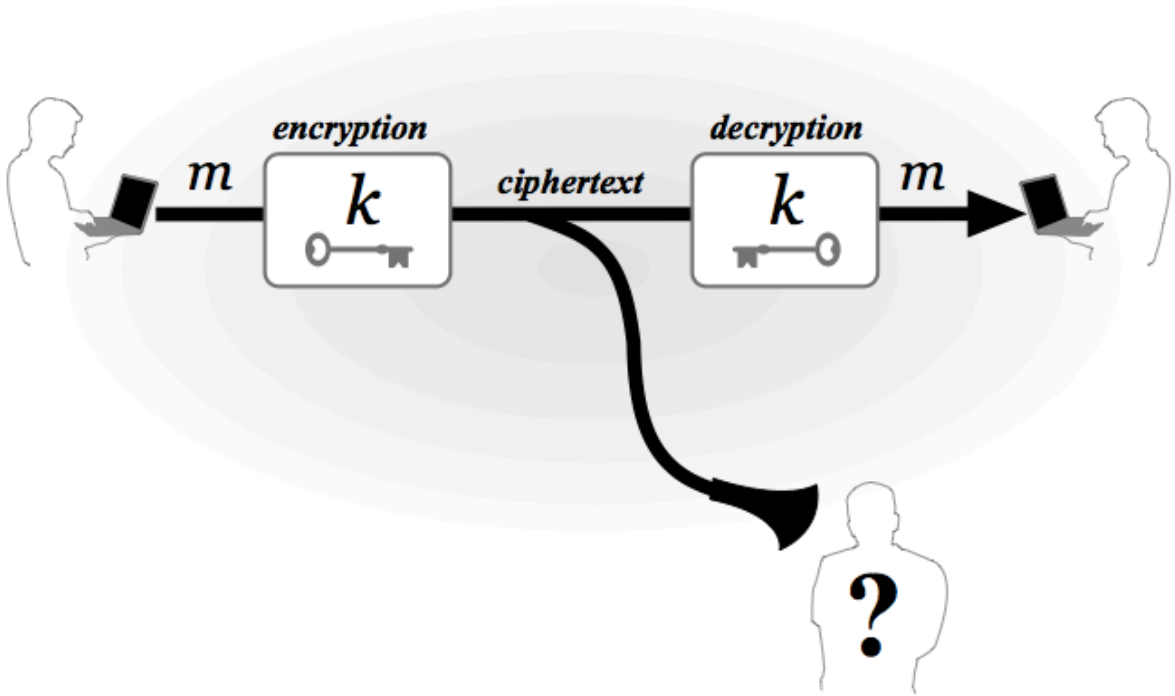
Obiettivo classico: progetto ed uso di codici (anche detti cifrari)

"due parti comunicano segretamente in presenza di un ascoltatore che può monitorare tutte le comunicazioni tra di loro"

Codici, in termini moderni: **schemi di cifratura**



Schemi di cifratura



Basato sulla segretezza della chiave condivisa dalle due parti comunicanti



Schemi di cifratura

A chiave simmetrica (symmetric key setting) = **stessa chiave** per cifrare e decifrare
A chiave asimmetrica (asymmetric key setting / public key setting = **chiavi diverse**

Due applicazioni canoniche per lo scenario a chiave simmetrica:

- le parti sono separate "**nello spazio**"
- una parte comunica con se stessa "**nel tempo**"

Sintassi che useremo per riferirci agli oggetti di uno schema:

M	:	spazio dei messaggi leciti
C	:	spazio dei cifrati possibili
Gen	:	algoritmo per generare le chiavi
Enc	:	algoritmo di cifratura
Dec	:	algoritmo di decifratura



Chiavi e principio di Kerckhoffs

Osservazione: un avversario, disponendo della chiave K e **Dec**, decifra!

Pertanto, la chiave deve essere segreta. E circa l'algoritmo **Dec** di decifratura?

August Kerckhoffs, fine 19esimo secolo

"Il metodo di cifratura non è richiesto che sia segreto e dovrebbe essere in grado di cadere nelle mani del nemico senza creare problemi."



La sicurezza non dovrebbe basarsi sulla segretezza degli algoritmi di cifratura, ma soltanto sulla segretezza della chiave.



Perché?

Argomenti a favore del principio di Kerckhoffs:

- è più facile per le parti proteggere solo K
- è più facile cambiare una chiave che un intero schema
- rende possibile la **standardizzazione**, che è utile per garantire compatibilità ed il pubblico scrutinio

Oggi il principio di Kerckhoffs è **universalmente accettato** e la pratica della cosiddetta "security by obscurity" – che poggia sulla segretezza anche dei metodi - sempre più desueta.



Schemi “storici”

Consideriamo ora alcuni cifrari al fine di:

- mostrare le debolezze di un approccio “ad hoc”, motivando l’importanza dell’approccio moderno strutturato

- rendere chiaro che approcci semplicistici difficilmente hanno successo



Cifrario di Cesare

"De Vita Caesarum", 110 a. C.

Cifratura: ogni lettera del messaggio in chiaro viene sostituita dalla lettera si trova tra posti in avanti nell'alfabeto

$$A \rightarrow D, \quad B \rightarrow E, \; C \rightarrow F, \quad \ldots, \quad Z \rightarrow C$$

Decifratura: operazione inversa (lettera che precede di 3 posti)

Osservazioni: lo schema di cifratura è fisso. Non c'è chiave segreta!

- una variante – ROT 13 – è usata ancora oggi in alcuni forum on-line per disturbare chi vuole capire al volo cosa le parti si dicono
- invece di 3, i posti in avanti sono 13

Il cifrario di Cesare è un caso particolare dello Shift Cipher



Shift cipher

Alfabeto Inglese mappato su {0, 1, 2, …, 25}



Messaggio $m = m_1 m_2 \ldots m_l$, con m_i in {0,1, 2, …, 25}

$\mathbf{Enc}_K(m_1 m_2 \ldots m_l) = c_1 c_2 \ldots c_l$, con $c_i = [(m_i + k) \bmod 26]$ per i=1, …, l

dove [a mod N] denota il resto della divisione per N e risulta 0 <= [a mod N] < N

L'associazione di [**a** mod N] ad **a** prende il nome di "riduzione modulo N".

$\mathbf{Dec}_K(c_1 c_2 \ldots c_l) = m_1 m_2 \ldots m_l$, con $m_i = [(c_i - k) \bmod 26]$ per i=1, …, l

È sicuro lo Shift cipher?



Ricerca esaustiva

Dato un cifrato, ci sono soltanto 26 possibili chiavi. Basta provarle tutte!

Un attacco che richiede di provare ogni possibile chiave viene detto un attacco per "**ricerca esaustiva**" o anche a "**forza bruta**".

Condizione necessaria (ma non sufficiente) affinché un cifrario sia sicuro è che lo spazio delle chiavi sia sufficientemente grande (sufficient key-space principle)

"*Qualsiasi schema di cifratura sicuro deve avere uno spazio delle chiavi sufficientemente grande da rendere un attacco per ricerca esaustiva impraticabile.*"

Tipicamente deve contenere almeno 2^{70} elementi al giorno d'oggi.



Cifrari per sostituzione monoalfabetica

"Shift cipher" → l'associazione carattere in chiaro/carattere cifrato è uno spostamento fissato

"Mono-alfabetico" → associazione arbitraria

Spazio delle chiavi = { tutte le possibili permutazioni dell'alfabeto}

alfabeto in chiaro

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

alfabeto cifrante



Cifrari per sostituzione monoalfabetica

Messaggio in chiaro = tell him about me tellhimaboutme
Messaggio cifrato = GDOO KVC XEFLG CD GDOOKVCXEFLGCD

Sostituzione monoalfabetica → La chiave specifica una sostituzione FISSA per ogni carattere del messaggio in chiaro

Lo stesso carattere viene cifrato sempre allo stesso modo

| Spazio delle chiavi | = 26! = 26•25•24 • … •2•1 ≈ 2^{88}

Un attacco a forza bruta è impraticabile! Sfortunatamente, **non è sufficiente** …



Analisi delle frequenze

Il cifrario è facile da "rompere".

Supponiamo che il messaggio in chiaro sia testo inglese corretto. Il cifrario può essere attaccato usando **caratteristiche statistiche** della lingua inglese.

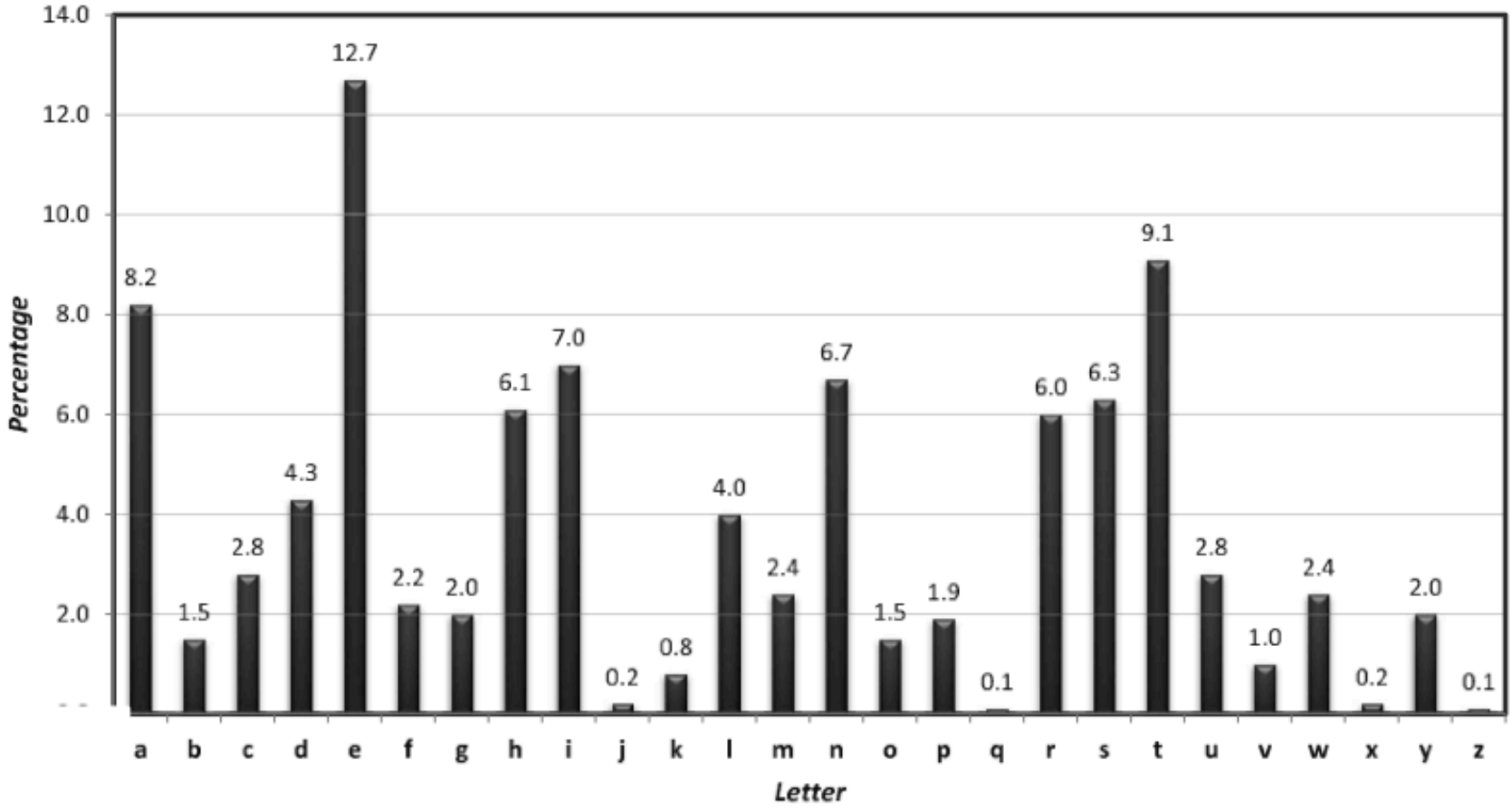
L'attacco si basa su due elementi:

1. Per ogni chiave, l'associazione di ciascuna lettera **è fissa**, per cui se "D" viene associata "e", allora **ogni** occorrenza di "e" nel messaggio in chiaro diventa "D" nel cifrato

2. La distribuzione delle frequenze delle lettere dell'alfabeto Inglese è nota.



Frequenze delle lettere dell'Inglese



Attacco tramite analisi delle frequenze

L'attacco opera **tabulando le frequenze dei caratteri presenti nel cifrato**.

Queste frequenze sono poi comparate alle frequenze dell'Inglese

In questo modo:

- parte dell'associazione può essere inferita
- ipotesi corrette ed altre conoscenze della lingua Inglese possono aiutare a ricostruirla (e.g., dipendenze tipo "the", "qu", etc)

Conclusione: spazio delle chiavi grande ma … cifrario debole!



Provate!

Per rendervi conto dell'applicabilità dell'attacco e del suo funzionamento, provate a decifrare l'esempio che segue (usando la tabella delle frequenze della lingua inglese)

JGRMQOYGHMVBJWRWQFPWHGFFDQGFPFZRKBEEBJIZQQOCIBZKLFAFGQVFZFWWE
OGWOPFGFHWOLPHLRLOLFDMFGQWBLWBWQOLKFWBYLBLYLFSFLJGRMQBOLWJVFP
FWQVHQWFFPQOQVFPQOCFPOGFWFJIGFQVHLHLROQVFGWJVFPFOLFHGQVQVFILE
OGQILHQFQGIQVVOSFAFGBWQVHQWIJVWJVFPFWHGFIWIHZZRQGBABHZQOCGFHX



Shift cipher: un attacco più efficiente

Usiamo la distribuzione delle frequenze per calcolare la chiave segreta.

Sia $0 \leq p_i \leq 1$ la frequenza della i-esima lettera dell'alfabeto inglese.

Risulta:

$$\sum_{i=0}^{25} p_i^2 \approx 0.065$$

Si consideri il cifrato dato.

Sia $0 \leq q_i \leq 1$ la frequenza della i-esima lettera nel cifrato, cioè:

$0 \leq q_i$ = (num. di occorrenze i-esima lettera nel cifrato / lungh. cifrato) ≤ 1

Se la chiave segreta scelta è k, allora si ha che:



Shift cipher: un attacco più efficiente

$p_i \longrightarrow q_{i+k}$ per tutti gli indici i=0, …, 25

i-esima lettera (i+k)-esima lettera nel cifrato

sostituita con

Per ogni valore di j = 0, 1, …, 25, calcoliamo $I_j \overset{def}{=} \sum_{i=0}^{25} p_i \cdot q_{i+j}$

Ci aspettiamo: $I_k \approx 0.065$

$I_j, j \neq k$ diverso da 0.065



Shift cipher: un attacco più efficiente

L'attacco restituisce come ipotesi per k il valore di j per cui I_j è più vicino a 0.065

Osservazioni sull'attacco:

- è efficiente
- è facile da automatizzare
- non è richiesta alcuna "analisi del significato" (attacco precedente manuale)



Il cifrario di Vigenère

Realizza uno "shift poli-alfabetico"

Osservazioni:

• gli attacchi statistici contro i cifrari mono-alfabetici sono possibili poiché la chiave definisce un'associazione FISSA, lettera per lettera, da applicare al testo in chiaro

• nei cifrari poli-alfabetici la chiave definisce una sostituzione da applicare **a blocchi** di caratteri del messaggio in chiaro

e.g., ab → DZ, ac → TY

non è associata ad un singolo carattere nel cifrato

In questo modo essenzialmente la distribuzione delle frequenze viene "appianata"



Il cifrario di Vigenère

Il cifrario di Vigenere usa semplicemente "più istanze indipendenti dello shift cipher".

chiave → stringa di caratteri

Plaintext:	tellhimaboutme
Key (repeated):	cafecafecafeca
Ciphertext:	VEQPJIREDOZXOE

Ricordo che:
caratteri = numeri
shift = + mod 26

shift cipher con chiave "**e**": 4°, 8°, 12°, …
shift cipher con chiave "**f**": 3°, 7°, 11°, …
shift cipher con chiave "**a**": 2°, 6°, 10°, …
shift cipher con chiave "**c**": 1°, 5°, 9°, …



Il cifrario di Vigenère

Se la chiave è sufficientemente lunga, rompere il cifrario sembra impresa ostica.

È stato ritenuto inattaccabile per centinaia di anni.



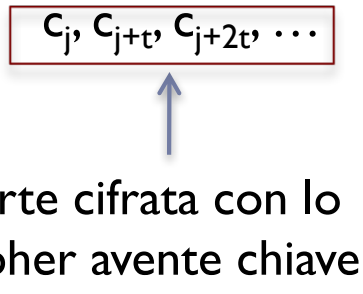


Rompere il cifrario di Vigenère

Per iniziare, si **assuma** che la lunghezza (detta periodo) della chiave k sia nota.

Sia $k = k_1 k_2 \ldots k_t$, dove k_j è un carattere della chiave.

Si divida il cifrato $c = c_1 c_2 \ldots$ in t parti. Precisamente, per $j = 1, \ldots, t$



Si calcoli la chiave k_j, per $j = 1, \ldots, t$, usando l'attacco efficiente contro lo shift cipher descritto in precedenza.



Test di Kasiski

Facile, vero? E se invece la lunghezza della chiave non è nota? Che si fa?

1. Se è noto un massimo alla lunghezza, potremmo tentare tutte le ipotesi, una per una, i.e., t=1, t=2, t=3, …

2. Usare il metodo di Kasiski
   - cerca di individuare "pattern ripetuti"

Plaintext:	the	man	and	the	woman	retrieved	the	letter	from	the	post	office
Key:	bea	dsb	ead	sbe	adsbe	adsbeadsb	ead	sbeads	bead	sbe	adsb	eadsbe
Ciphertext:	ULE	PSO	ENG	LII	WREBR	RHLSMEYWE	XHH	DFXTHJ	GVOP	LII	PRKU	SFIADI



due volte



Test di Kasiski

La distanza tra i pattern ripetuti, assumendo che non sia un evento accidentale, deve essere un **multiplo** del periodo della chiave

Il **massimo comune divisore** delle distanze tra i pattern ripetuti restituirà il periodo (o un multiplo di esso)



limite superiore al periodo

Tuttavia, esiste un altro approccio al calcolo del periodo, maggiormente strutturato e facile da automatizzare.

È il metodo dell'indice di coincidenza.



Indice di coincidenza

Se il periodo del cifrato $c = c_1 c_2 \ldots$ è t, allora

$$c_1, c_{1+t}, c_{1+2t}, c_{1+3t}, \cdots$$

è la prima sequenza che risulta dalla cifratura ottenuta usando il "primo shift cipher".

Come già osservato

le frequenze dei caratteri in **questa** sequenza

identiche a

le frequenze dei caratteri dell'Inglese in qualche ordine shiftato

$p_i \longrightarrow q_{i+j}$

se lo shift usato è j



Indice di coincidenza

Ciò significa che $q_0 q_1 \ldots q_{25}$ è semplicemente $p_0 p_1 \ldots p_{25}$ "shiftata" di j posizioni.
Pertanto, risulta

$$\sum_{i=0}^{25} q_i^2 \approx \sum_{i=0}^{25} p_i^2 \approx 0.065$$

Disponiamo quindi, di un modo elegante per determinare la lunghezza t della chiave.
Per w=1,2,3, … consideriamo la sequenza

$$c_1, c_{1+w}, c_{1+2w}, c_{1+3w} \cdots$$

e calcoliamo i valori $q_0 q_1 \ldots q_{25}$, e il valore

$$S_w = \sum_{i=0}^{25} q_i^2$$



Indice di coincidenza

Ci aspettiamo che, quando w = t, risulta $S_w \approx 0.065$.

D'altra parte, se w ≠ t, i caratteri $c_1, c_{1+w}, c_{1+2w}, c_{1+3w}\ldots$ sono presenti approssimativamente con la stessa probabilità, prossima a quella uniforme.

Vale a dire, $q_i \approx 1/26$, per i = 0, …, 25. Da cui risulta

$$S_w = \sum_{i=0}^{25} q_i^2 = \sum_{i=0}^{25} \left(\frac{1}{26}\right)^2 \approx 0.038$$

Pertanto, il valore più piccolo di w per cui $S_w \approx 0.065$ è verosimilmente la lunghezza della chiave.

Una seconda sequenza $c_2, c_{2+w}, c_{2+2w}, c_{2+3w}\ldots$ può essere usate per confermare l'ipotesi



Conclusioni

- l'attacco è elegante ed efficiente
- richiede un cifrato lungo per la stima delle frequenze
- una chiave più lunga implica la necessità di un cifrato più lungo

Progettare cifrari sicuri non è impresa facile!



Crittografia moderna: pilastri

- Spostamento verso una "scienza"
- **Definizioni** rigorose di cosa "sicuro" significa
- **Assunzioni** circa la complessità algoritmica di certi problemi matematici
- **Dimostrazioni/prove** che una costruzione è sicura



Principio 1: definizioni rigorose

- Essenziali per la progettazione accurata, lo studio, la valutazione e l'uso di primitive crittografiche

"*Se non è chiaro cosa si vuole ottenere, come è possibile stabilire quando (e se) il risultato è stato ottenuto?*"



Principio 1: definizioni rigorose

- Permettono di valutare ciò che è stato costruito
  - la costruzione raggiunge le garanzie volute?
  - si può produrre evidenza?
  - esiste un controesempio?
  - sotto quali condizioni la costruzione raggiunge le garanzie?

- Permettono di comparare schemi
  - contesti diversi possono richiedere definizioni diverse
  - valutare trade-off tra schemi
  - un approccio modulare (rimpiazzabilità dei moduli)



Le definizioni non sono facili

- Due componenti
  - garanzie di sicurezza (**security goal**): da quali tipi di azioni di un attaccante lo schema protegge
  - un modello delle minacce (**threat model**): che potere ha l'attaccante

- Cosa dovrebbe garantire uno schema di cifratura sicuro?
  - Dovrebbe essere impossibile per un attaccante recuperare la chiave di cifratura? Lo schema

$$\mathbf{Enc}_K(m) = m$$

non fornisce alcuna informazione su K ma è chiaramente insicuro.



Le definizioni non sono facili

- Dovrebbe essere impossibile per un attaccante recuperare l'intero testo in chiaro dal cifrato?
  - Si consideri uno schema di cifratura che protegge un database di dati sensibili e rivela il 90% del suo contenuto.
  - Siamo soddisfatti che il 10% è invece protetto?

- Dovrebbe essere impossibile per un attaccante recuperare qualsiasi carattere del messaggio in chiaro dal cifrato?
  - Sembra buona, ma ancora insufficiente.
  - In un database di dati finanziari potrebbe rivelare se alcune transazioni hanno valori maggiori o minori di una certa soglia
  - E poi, come formalizzare il "recupero di un carattere"?
  - Inoltre, provare ad indovinare è un attacco da considerare?



Le definizioni non sono facili

- La definizione giusta dovrebbe:
  - escludere il rilascio di informazioni utili da parte del cifrato
  - chiarire cosa debba essere considerato un attacco

- Ciò che richiediamo in fondo è che:

Indipendentemente da qualsiasi informazione l'attaccante possa già avere, un cifrato non dovrebbe rilasciare ***nessuna informazione aggiuntiva*** circa il sottostante messaggio in chiaro



Le definizioni non sono facili

- La definizione non cerca di definire quale tipo di informazione circa il messaggio in chiaro sia "significativa"
  - Nessuna informazione aggiuntiva deve essere rilasciata
  - Lo schema di cifratura è utile in tutte le potenziali applicazioni

- Cosa manca ancora? Una precisa formulazione di
  - conoscenza a-priori dell'attaccante sul messaggio in chiaro
  - cosa significa esattamente "rilasciare informazione"



Le definizioni non sono facili

- Cosa dovrebbe prevedere il modello delle minacce?
  - specificare il potere dell'avversario, le sue abilità
  - non porre alcuna restrizione alle strategie d'attacco, cioè non fare alcuna assunzione su come usa le proprie abilità
- Nel contesto della cifratura, i modelli sono 4
- Ciphertext only
  - l'attaccante può solo osservare cifrati c, prodotti usando una chiave k
- Known-plaintext:
  - l'attaccante acquisisce coppie (m,c) in qualche modo, prodotte usando una chiave k



Le definizioni non sono facili

- Chosen-plaintext
  - l'attaccante acquisisce coppie (m,c), scegliendo i valori di m, prodotte usando la chiave k
- Chosen-ciphertext
  - l'attaccante acquisisce coppie (m,c), scegliendo i valori di c, prodotte usando la chiave k



Principio 2: assunzioni

- La maggior parte delle costruzioni crittografiche moderne non possono essere provate sicure "incondizionatamente"
  - richiederebbe la risoluzione di questioni della teoria della complessità che oggi non hanno ancora risposta (e.g., P ≠ NP ?)

- Le prove di sicurezza poggiano su **assunzioni** enunciate con *chiarezza* e *rigore matematico*



richiesto dalle prove ma anche per i motivi che seguono



Rigore matematico: permette

- Validazione delle assunzioni
  - enunciati che si "congettura" risultino veri
  - più sono studiati, maggiore è la confidenza che vi riponiamo
  - formulazioni imprecise ostacolano lo studio

- Comparazione di schemi
  - uno schema basato su un'assunzione più debole è preferibile ad uno schema basato su un'assunzione più forte
  - se due assunzioni non sono confrontabili, dovrebbe preferirsi lo schema basato sull'assunzione studiata di più



Rigore matematico: permette

- Comprensione delle assunzioni necessarie
  - se uno schema è basato su "blocchi" e un blocco viene rotto, possiamo verificare se il problema è nel blocco o nell'assunzione



Assumere che uno schema è sicuro?

- Quando uno schema ha resistito con successo ad attacchi per molti anni, può essere ragionevole
- In generale questo approccio non è **mai** da preferirsi

- Ragioni più specifiche:
  - un'assunzione scrutinata per diversi anni è preferibile ad una nuova, magari ad hoc
  - assunzioni semplici sono preferibili
  - assunzioni di basso livello possono essere usate in svariate costruzioni
  - la progettazione può essere modulare, blocchi sostituibili



Principio 3: prove

- Definizioni ed assunzioni permettono di fornire prove che una costruzione soddisfa una data definizione sotto le assunzioni specificate

- Le prove sono assicurazioni del fatto che nessun attaccante, **relativamente** alla definizione ed alle assunzioni, avrà successo

- Meglio di un approccio "euristico" e non strutturato



non basato su principi chiari



Terminologia: prove …

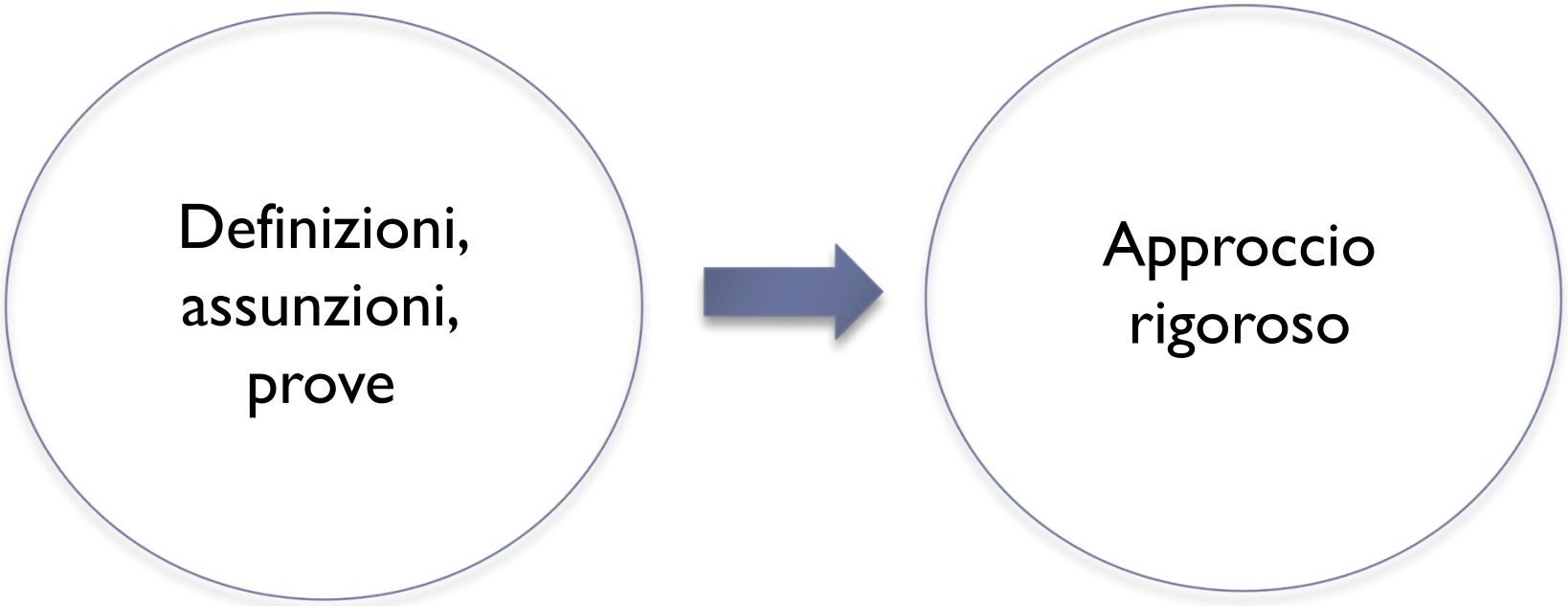
- Useremo entrambi i termini "dimostrazione" e "prova"
- Una locuzione più precisa ma più lunga sarebbe **riduzione di sicurezza (security reduction)**
  - gli enunciati che dimostreremo saranno del tipo: "se le Assunzioni x, y … valgono, allora la Costruzione Π soddisfa la Definizione Z



specifica il security goal ed il threat model



Conclusioni: rigoroso vs ad hoc



… ma nel mondo reale soluzioni veloci sono spesso progettate seguendo un approccio ad hoc e valutazioni euristiche



“Mondo reale” e “mondo delle prove”

- Che relazione c’è tra i due mondi?

- Occorre **non sopravvalutare** cosa una prova offre
  - le garanzie sono in relazione alla definizione considerata ed alle assunzioni utilizzate
  - sono un suggerimento all’avversario circa le “direzioni d’attacco” da **non** seguire …
  - l’efficacia di una prova dipende in **maniera cruciale** da quanto il mondo reale sia ben modellato dalla definizione



Conclusione

L'approccio delle riduzioni di sicurezza non conclude sicuramente l'eterna battaglia tra attaccanti e difensori, ma sposta sicuramente l'ago della bilancia dalla parte dei difensori



Shannon: transizione verso scienza

- Ci occuperemo di schemi di cifratura perfettamente sicuri

  ↓ confidenzialità / riservatezza

- Avversari di potere computazionale illimitato
- Useremo definizioni matematiche precise e prove
- Nessuna assunzione!

  ↓

  Inerenti limitazioni → giustificano la necessità di assunzioni

- Useremo esperimenti ed algoritmi randomizzati



Come generare randomness?

- In principio, per pochi bit random, lanciando una moneta!
- Tecniche moderne procedono in due fasi
  - una discreta quantità di dati ad "alta entropia" viene raccolta



  - i dati vengono elaborati per produrre una sequenza di bit quasi indipendenti e distribuiti quasi uniformemente

- Per la prima fase, occorre una sorgente di dati imprevedibili



Sorgenti

- Input esterni
  - ritardo dei dati in transito in rete
  - tempi di accesso all'hard disk
  - misure sulla pressione dei tasti
  - misure sul movimento del mouse
  - …

dati "lontani" dalla distribuzione uniforme

- Fenomeni fisici
  - misure del rumore nei dispositivi elettronici
  - misure del decadimento radioattivo



Estrazione di bit casuali

- Il processo di "livellamento" dei dati ad alta entropia è non banale. Ne mostriamo uno.
  - Supponiamo di disporre di una sequenza dei risultati del lancio di una moneta difettosa, rappresentata in binario
  - La moneta dà:

    T=1 con probabilità p
    C=0 con probabilità (1-p)

  - I lanci sono indipendenti l'uno dall'altro
  - Consideriamo i bit a coppie
  - Se vediamo: 00 01 10 11
    scriviamo / 0 1 /



Estrazione di bit casuali

10 01 11 01 10 00 11 10 10 01

1 0 0 1 1 1 0

- la sequenza binaria calcolata è distribuita uniformemente
- Perchè?
  - La probabilità che una coppia produca 0 è: (1-p) p
  - La probabilità che una coppia produca 1 è: p (1-p)
- La generazione e l'uso di random bit è cruciale per la sicurezza dei protocolli crittografici
  - Generatori di numeri casuali devono essere progettati per uso crittografico.
  - "General-purpouse" non sono idonei.



Definizioni

- Schema di cifratura: (Gen, Enc, Dec), M tale che |M|>1

spazio dei messaggi

- Gen: algoritmo probabilistico di generazione delle chiavi, restituisce $k \in K$ in accordo ad una distribuzione di probabilità

spazio delle chiavi

- Enc: algoritmo probabilistico di cifratura, prende in input $m \in M, k \in K$ e restituisce un cifrato $c \in C$

spazio dei cifrati

- Dec: algoritmo deterministico di decifratura, prende in input $c \in C, k \in K$ e restituisce $m \in M$



Notazioni

- $k \leftarrow Gen()$
- $c \leftarrow Enc_k(m)$

output di alg. probabilistici

- $c := Enc_k(m)$
- $m := Dec_k(c)$

output di alg. deterministici

- Useremo la notazione $x \leftarrow S$ anche per denotare l'estrazione di un elemento da un insieme S in accordo alla distribuzione uniforme



Condizione di correttezza perfetta

- Per ogni $k \in K$, per ogni $m \in M$, e qualsiasi c $\leftarrow$ $Enc_k(m)$, risulta

$$Dec_k(c) = m, \text{ con probabilità } 1$$

Nei teoremi e nelle definizioni che seguono faremo riferimento alle distribuzioni di probabilità sugli insiemi K, M e C.

La distribuzione su K dipende da Gen()

→ assumeremo uniforme

La variabile casuale **K** denoterà l'output di Gen()



Variabili casuali

- La variabile casuale **M** denoterà il messaggio "in chiaro"

La Pr[**M** = m] è la probabilità che m debba essere cifrato



Non dipende dallo schema di cifratura. Riflette la frequenza dei differenti messaggi che le parti si inviano, così come l'incertezza che l'avversario Adv ha su di essi.

e.g., Adv può sapere che i messaggi sono solo due e che:
Pr[**M** = "attacca oggi"] = 0.7
Pr[**M** = "attacca domani"] = 0.3

- Le variabili casuali **K** ed **M** si assumono **indipendenti**



Variabili casuali

- Fissato uno schema di cifratura ed una distribuzione su M si determina una distribuzione sullo spazio dei cifrati C, ottenuta
  - da k generata tramite Gen()
  - scegliendo m in accordo alla distribuzione su M
  - calcolando $c \leftarrow Enc_k(m)$



- La variabile casuale **C** denoterà l'output del processo di cifratura
- Nota che $\mathbf{C} = Enc_{\mathbf{K}}(\mathbf{M})$





Esempio

- Consideriamo lo Shift Cipher

  $K = \{0,\ldots,25\}$, $Pr[\mathbf{K}=k] = 1/26$, per ogni $k \in K$

  Sia data su M la distribuzione: $Pr[\mathbf{M} = a] = 0.7$ e $Pr[\mathbf{M} = z] = 0.3$

  Qual è la probabilità che il cifrato sia B?

  Ci sono solo due possibilità: $m = a$ e $k = 1$, oppure $m = z$ e $k = 2$

  Poiché le variabili casuali **M** e **K** sono indipendenti, risultano:

  $$Pr[\mathbf{M} = a \wedge \mathbf{K} = 1] = Pr[\mathbf{M} = a] \bullet Pr[\mathbf{K} = 1] = 0.7 \bullet 1/26$$

  $$Pr[\mathbf{M} = z \wedge \mathbf{K} = 2] = Pr[\mathbf{M} = z] \bullet Pr[\mathbf{K} = 2] = 0.3 \bullet 1/26$$

  Pertanto,

  $$Pr[\mathbf{C} = B] = Pr[\mathbf{M} = a \wedge \mathbf{K} = 1] + Pr[\mathbf{M} = z \wedge \mathbf{K} = 2]$$
  $$= 0.7 \bullet 1/26 + 0.3 \bullet 1/26 = 1/26$$



Esempio

- Consideriamo lo Shift Cipher

Qual è invece la probabilità che il messaggio in chiaro sia “a”, dato il cifrato B?

Usando il teorema di Bayes, risulta:

$$\Pr[\mathbf{M} = a \mid \mathbf{C} = B] = \frac{\Pr[\mathbf{C} = B \mid \mathbf{M} = a] \cdot \Pr[\mathbf{M} = a]}{\Pr[\mathbf{C} = B]}$$

$$= \frac{\Pr[\mathbf{C} = B \mid \mathbf{M} = a] \cdot 0.7}{1/26}$$

Ma $\Pr[\mathbf{C} = B \mid \mathbf{M} = a] = 1/26$ perché, se $\mathbf{M} = a$, l’unico modo per ottenere $\mathbf{C} = B$ è che $\mathbf{K} = 1$ (che accade con prob. 1/26)

Pertanto, $\Pr[\mathbf{M} = a \mid \mathbf{C} = B] = \frac{1/26 \cdot 0.7}{1/26} = 0.7$



Ulteriore esempio

- Ancora lo Shift Cipher

- Sia data su M la distribuzione:

$\Pr[\mathbf{M} = \text{kim}] = 0.5, \quad \Pr[\mathbf{M} = \text{ann}] = 0.2 \text{ e } \quad \Pr[\mathbf{M} = \text{boo}] = 0.3$

Qual è la probabilità che $\mathbf{C} = \text{DQQ}$?

Ci sono soltanto due possibilità: $\mathbf{M} = \text{ann}$ e $\mathbf{K} = 3$, oppure $\mathbf{M} = \text{boo}$ e $\mathbf{K} = 2$

Pertanto,

$$\begin{aligned}\Pr[\mathbf{C} = \text{DQQ}] &= \Pr[\mathbf{M} = \text{ann} \wedge \mathbf{K} = 3] + \Pr[\mathbf{M} = \text{boo} \wedge \mathbf{K} = 2] \\ &= 0.2 \bullet 1/26 + 0.3 \bullet 1/26 = 0.5 \bullet 1/26 \\ &= 1/2 \bullet 1/26 \\ &= 1/52\end{aligned}$$



Ulteriore esempio

Similmente, usando il teorema di Bayes, risulta:

$$\Pr[\mathbf{M} = \text{ann} \mid \mathbf{C} = \text{DQQ}] = \frac{\Pr[\mathbf{C} = \text{DQQ} \mid \mathbf{M} = \text{ann}] \bullet \Pr[\mathbf{M} = \text{ann}]}{\Pr[\mathbf{C} = \text{DQQ}]}$$

$$= \frac{1/26 \bullet 0.2}{1/52} = 0.4$$

essendo $\Pr[\mathbf{C} = \text{DQQ} \mid \mathbf{M} = \text{ann}] = 1/26$, dato che solo $\mathbf{K} = 3$ produce $\mathbf{C} = \text{DQQ}$.



Segretezza perfetta

- Scenario: Adv conosce
  - la distribuzione di probabilità su M
  - lo schema di cifratura
  - non conosce la chiave k
  - può ascoltare/vedere la comunicazione

Per ottenere segretezza perfetta, l'osservazione del cifrato non dovrebbe fornire alcuna informazione ad Adv sul messaggio in chiaro. Cioè:

Il cifrato non rivela nulla



Segretezza perfetta

Definizione 2.3. Uno schema di cifratura (Gen, Enc, Dec) con spazio dei messaggi M è perfettamente segreto se

- per ogni distribuzione di probabilità su M
- per ogni messaggio $m \in M$
- e per ogni cifrato $c \in C$ per cui risulta Pr[**C**=c]>0, si ha:

$$\Pr[\mathbf{M} = m \mid \mathbf{C} = c] = \Pr[\mathbf{M} = m]$$



Definizione equivalente

- Possiamo fornire una definizione equivalente richiedendo che la distribuzione di probabilità dei cifrati **non dipenda** dal messaggio in chiaro

- Formalmente, per ogni $m, m' \in M$ e per ogni $c \in C$

$$\Pr[\mathrm{Enc}_k(m) = c] = \Pr[\mathrm{Enc}_k(m') = c]$$

dove le prob sono calcolate sulle possibili scelte

  - della chiave k
  - dei bit casuali utilizzati $\mathrm{Enc}_k()$



Definizione equivalente

- La precedente condizione implica
  - un cifrato non rivela alcuna informazione sul messaggio in chiaro
  - è impossibile distinguere una cifratura di m da una di m', poiché le distribuzioni sui cifrati sono le stesse

Lemma 2.4 Uno schema di cifratura (Gen, Enc, Dec) con spazio dei messaggi M è perfettamente segreto **se e solo se** per ogni $m,m' \in M$ ed ogni $c \in C$ risulta:

$$\Pr[\mathrm{Enc}_k(m) = c] = \Pr[\mathrm{Enc}_k(m') = c]$$



Dimostrazione

- Proviamo che, se la condizione vale, **allora** lo schema è perfettamente segreto. Fissiamo una distribuzione su M e sia c tale che $\Pr[\mathbf{C} = c] > 0$.
  - se $\Pr[\mathbf{M} = m] = 0$, allora banalmente

$$\Pr[\mathbf{M} = m \mid \mathbf{C} = c] = \frac{\Pr[\mathbf{C} = c \mid \mathbf{M} = m] \cdot \Pr[\mathbf{M} = m]}{\Pr[\mathbf{C} = c]} = 0$$

(il numeratore tende a 0)

$$\Pr[\mathbf{M} = m \mid \mathbf{C} = c] = \Pr[\mathbf{M} = m]$$

  - se $\Pr[\mathbf{M} = m] > 0$, invece, notiamo che

$$\Pr[\mathbf{C} = c \mid \mathbf{M} = m] = \Pr[\mathrm{Enc}_{\mathbf{K}}(\mathbf{M}) = c \mid \mathbf{M} = m] = \Pr[\mathrm{Enc}_{\mathbf{K}}(m) = c]$$

Sia $\delta_c = \Pr[\mathrm{Enc}_{\mathbf{K}}(m) = c]$. Se la condizione del lemma vale, per ogni m'



Dimostrazione

risulta:

$$\Pr[\mathrm{Enc}_{\mathbf{K}}(m') = c] = \Pr[\mathbf{C} = c \mid \mathbf{M} = m'] = \Pr[\mathbf{C} = c \mid \mathbf{M} = m] = \delta_c$$

Applicando il teorema di Bayes,

$$\Pr[\mathbf{M} = m \mid \mathbf{C} = c] = \frac{\Pr[\mathbf{C} = c \mid \mathbf{M} = m] \bullet \Pr[\mathbf{M} = m]}{\Pr[\mathbf{C} = c]}$$

$$= \frac{\Pr[\mathbf{C} = c \mid \mathbf{M} = m] \bullet \Pr[\mathbf{M} = m]}{\sum_{m' \in M} \Pr[C = c \mid M = m'] \cdot \Pr[M = m']}$$

$$= \frac{\delta_c \bullet \Pr[\mathbf{M} = m]}{\delta_c \bullet \sum_{m' \in M} \Pr[M = m']} = \Pr[\mathbf{M} = m]$$

$\sum_{m' \in M} \Pr[C = c \mid M = m'] \cdot \Pr[M = m'] \rightarrow \delta_c$

$\delta_c \bullet \sum_{m' \in M} \Pr[M = m'] \rightarrow 1$

Pertanto lo schema è perfettamente segreto

- L'altra direzione si prova in modo simile.

Perfetta indistinguibilità

- Possiamo dare ancora un'altra definizione di segretezza perfetta equivalente alle precedenti
  - basata su un esperimento che coinvolge Adv e un Challenger C
- Informalmente
  - Adv sceglie due messaggi in chiaro, m ed m'
  - C sceglie uniformemente a caso uno dei due e lo cifra usando una chiave k casuale
  - Il cifrato ottenuto c viene dato ad Adv
  - Adv dà in output una sua ipotesi su quale dei due è stato cifrato
  - Adv ha successo se la sua ipotesi è corretta



Perfetta indistinguibilità

- Uno schema di cifratura è perfettamente indistinguibile se nessun Adv può avere successo con probabilità > ½

- Osservazioni:
  - un Adv che ipotizza a caso ha successo con probabilità ½ richiediamo che Adv non disponga di una strategia migliore
  - non poniamo alcun limite al potere computazionale di Adv



Esperimento $\mathrm{Priv}K^{eav}_{A,\Pi}$

Sia Π = (Gen, Enc, Dec) uno schema di cifratura con spazio dei messaggi M.
Sia A un Adv (i.e., un algoritmo che mantiene uno stato)

$\mathrm{Priv}K^{eav}_{A,\Pi}$

- A dà in output una coppia di messaggi m_0, m_1
- C genera k tramite Gen() e sceglie uniformemente a caso il bit b in {0,1}
- C calcola c $\leftarrow$ $\mathrm{Enc}_K(m_b)$, detto cifrato di sfida, e lo passa ad A
- A dà in output un bit b'
- L'output dell'esperimento è 1 se b' = b, 0 altrimenti. Se l'output è 1, A ha successo.



Definizione

Definizione 2.5. Uno schema di cifratura (Gen, Enc, Dec) con spazio dei messaggi M è perfettamente indistinguibile se, per ogni A, risulta:

$$\Pr[\ \Pr iv K^{eav}_{A,\Pi} = 1] = 1/2$$

Risultato dell'esperimento

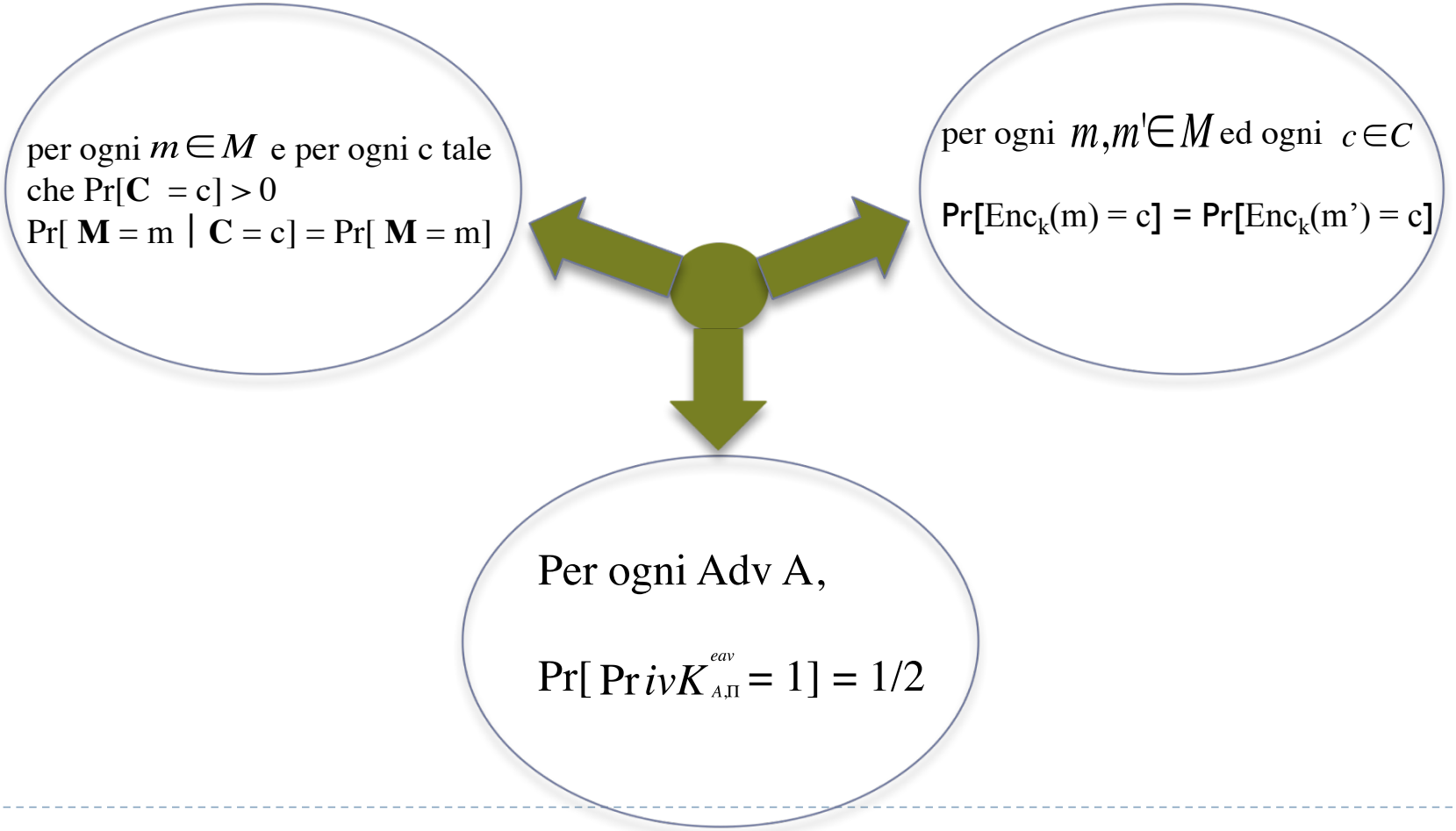
È possibile dimostrare che vale il seguente

Lemma 2.6. Uno schema di cifratura Π è perfettamente segreto **se e solo se** è perfettamente indistinguibile.



Conclusione

- Disponiamo di tre definizioni di segretezza perfetta equivalenti!



Esempio 2.7. Il cifrario di Vigenere, per certi parametri, non é perfettamente indistinguibile.

Consideriamo un cifrario di Vigenere per uno spazio di messaggi M di stringhe di due caratteri, ed in cui la lunghezza della chiave (periodo) é scelta uniformemente in $\{1, 2\}$.

Mostreremo un Adv A per cui $Pr[PrivK^{eav}_{A,\Pi} = 1] > \frac{1}{2}$.

Adv A

1. Dá in output $m_0 = aa$ ed $m_1 = ab$.
2. Dopo aver ricevuto il cifrato $c = c_1 c_2$
   - se $c_1 = c_2$ dá in output $b' = 0$;
   - altrimenti, dá in output $b' = 1$.

Perfetta Indistinguibilitá

Homework

Calcoliamo la probabilitá di successo di **A**.

$$\begin{aligned} Pr[PrivK^{eav}_{A,\Pi} = 1] &= \frac{1}{2} \cdot Pr[PrivK^{eav}_{A,\Pi} = 1 | b = 0] + \frac{1}{2} \cdot pr[PrivK^{eav}_{A,\Pi} = 1 | b = 1] \\ &= \frac{1}{2} \cdot Pr[A \text{ dá } 0 | b = 0] + \frac{1}{2} \cdot Pr[A \text{ dá } 1 | b = 1] \end{aligned}$$

Valutiamo i due termini separatamente:

$Pr[A \text{ dá } 0 | b = 0]$ solo se

- viene scelta una chiave di lunghezza 1 (prob. $\frac{1}{2}$)
- viene scelta una chiave di lunghezza 2 (prob. $\frac{1}{2}$) con due valori uguali (prob. $\frac{1}{26}$)

Pertanto,

$$Pr[A \text{ dá } 0 | b = 0] = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{26} \approx 0.52.$$

Perfetta Indistinguibilitá

Homework

D'altra parte, poiché $Pr[A \text{ dá } 0|b = 1]$ solo se

- viene scelta una chiave di lunghezza 2 (prob. $\frac{1}{2}$) ed il primo valore vale uno piú del secondo (prob. $\frac{1}{26}$)

Pertanto,

$$Pr[A \text{ dá } 1|b = 1] = 1 - Pr[A \text{ dá } 0|b = 1] = 1 - \frac{1}{2} \cdot \frac{1}{26} \approx 0.98.$$

Mettendo assieme le varie parti, risulta

$$\begin{aligned} Pr[PrivK^{eav}_{A,\Pi} = 1] &= \frac{1}{2} \cdot (\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{26}) + \frac{1}{2} \cdot (1 - \frac{1}{2} \cdot \frac{1}{26}) \\ &= \frac{1}{2} \cdot (\frac{1}{2} + 1) = \frac{3}{4} = 0.75 > 0.5 = \frac{1}{2} \end{aligned}$$

Quindi, lo schema **non** é perfettamente indistinguibile.

One-time Pad

Brevettato da Vernam nel 1917. Circa 25 anni piú tardi Shannon dimostró che é perfettamente segreto.

Costruzione 2.8

Sia $\ell > 0$ un intero. Siano $M = K = C = \{0,1\}^\ell$.

- Gen: sceglie $k \in \{0,1\}^\ell$ uniformemente a caso
- Enc: dati $k \in \{0,1\}^\ell$ ed $m \in \{0,1\}^\ell$, dá in output il cifrato

$$c := k \oplus m$$

- Dec: dati $k \in \{0,1\}^\ell$ e $c \in \{0,1\}^\ell$, dá in output il messaggio

$$m := k \oplus c$$

É facile verificare che:

$$\forall k, \forall m \text{ risulta } Dec_k(Enc_k(m)) = (k \oplus (k \oplus m)) = m.$$

One-time Pad

Homework

Teorema 2.9. Lo schema di cifratura one-time pad é perfettamente segreto.

Dim. Prima di tutto, calcoliamo $Pr[\mathbf{C} = c|\mathbf{M} = m']$ per un arbitrario $c \in C$ ed $m' \in M$. Risulta:

$$\begin{aligned} Pr[\mathbf{C} = c|\mathbf{M} = m'] &= Pr[Enc_{\mathbf{K}}(m') = c] \\ &= Pr[m' \oplus \mathbf{K} = c] \\ &= Pr[\mathbf{K} = m' \oplus c] = 2^{-\ell}, \end{aligned}$$

poiché k é una chiave scelta uniformemente a caso in $\{0,1\}^{\ell}$.
Per ogni $c \in C$, abbiamo:

$$\begin{aligned} Pr[\mathbf{C} = c] &= \sum_{m' \in M} Pr[\mathbf{C} = c|\mathbf{M} = m'] \cdot Pr[\mathbf{M} = m'] \\ &= \sum_{m' \in M} 2^{-\ell} \cdot Pr[\mathbf{M} = m'] = 2^{-\ell} \cdot \sum_{m' \in M} Pr[\mathbf{M} = m'] \\ &= 2^{-\ell} \cdot 1 = 2^{-\ell}, \end{aligned}$$

Homework

dove la somma é calcolata su tutti gli $m' \in M$ tali che $Pr[\mathbf{M} = m'] > 0$.

Applicando allora il Teorema di Bayes, otteniamo:

$$\begin{aligned} Pr[\mathbf{M} = m | \mathbf{C} = c] &= \frac{Pr[\mathbf{C} = c | \mathbf{M} = m] \cdot Pr[\mathbf{M} = m]}{Pr[\mathbf{C} = c]} \\ &= \frac{2^{-\ell} \cdot Pr[\mathbf{M} = m]}{2^{-\ell}} \\ &= Pr[\mathbf{M} = m]. \end{aligned}$$

Pertanto, lo schema di cifratura one-time pad é perfettamente segreto.

Limitazioni della segretezza perfetta

Si noti che, nello schema one-time pad

- la chiave é tanto lunga quanto il messaggio che si intende cifrare
- é sicuro per un uso soltanto
  - per esempio, dati $c = m \oplus k$ e $c' = m' \oplus k$, risulta

$$c \oplus c' = (m \oplus k) \oplus (m' \oplus k) = m \oplus m',$$

    ovvero un Adv puó calcolare la differenza tra i due messaggi (molta informazione)!
- L'esempio é sufficiente per dire che il one-time pad **non** é perfettamente segreto per qualsiasi nozione di segretezza perfetta per messaggi multipli.

Purtroppo i limiti del one-time pad sono limiti *intrinseci* alla segretezza perfetta.

Limitazioni della segretezza perfetta

Faremo vedere che ogni schema perfettamente segreto deve avere uno spazio delle chiavi *almeno* tanto grande quanto lo spazio dei messaggi.

Da cui, discende che :

- se in uno schema perfettamente segreto tutte le chiavi sono della stessa lunghezza
- e se lo spazio dei messaggi consiste di tutte le stringhe di una data lunghezza

$$\Downarrow$$

la chiave é almeno tanto lunga quanto il messaggio

$$\Downarrow$$

lo schema di cifratura one-time pad é ottimale rispetto alla lunghezza della chiave.

Risultati di Shannon

Teorema 2.10. Se (Gen, Enc, Dec) é uno schema di cifratura perfettamente segreto con spazio dei messaggi M e spazio delle chiavi K, allora

$$|K| \geq |M|.$$

Dim. Mostriamo che, se $|K| < |M|$, lo schema non puó essere perfettamente segreto. Sia $|K| < |M|$. Sia $\mathbf{M}$ distribuita uniformemente e sia $c \in C$ tale che $Pr[\mathbf{C} = c] > 0$. Definiamo

$$M(c) \stackrel{def}{=} \{m | m = Dec_k(c), \text{ per qualche } k \in K\}.$$

Chiaramente $|M(c)| \leq |K|$. Se $|K| < |M|$, allora $\exists m' \in M$ tale che $m' \notin M(c)$. Ma allora risulta:

$$Pr[\mathbf{M} = m' | \mathbf{C} = c] = 0 \neq Pr[\mathbf{M} = m'].$$

Pertanto lo schema non é perfettamente segreto. Quindi deve essere $|K| \geq |M|$.

Teorema di Shannon

É uno strumento utile per provare la segretezza perfetta di uno schema di cifratura.

Teorema 2.11. Sia (Gen, Enc, Dec) uno schema di cifratura con spazio dei messaggi M per cui $|M| = |K| = |C|$. Lo schema é perfettamente segreto **se e solo se**:

1. Gen sceglie ogni chiave $k \in K$ con probabilitá uguale a $\frac{1}{|K|}$
2. per ogni $m \in M$ ed ogni $c \in C$, esiste un'unica chiave $k \in K$ tale che $Enc_k(m) = c$.

Dim. Si consulti il libro di testo.