CyberChallengeIT 2019
Università degli studi di Pisa

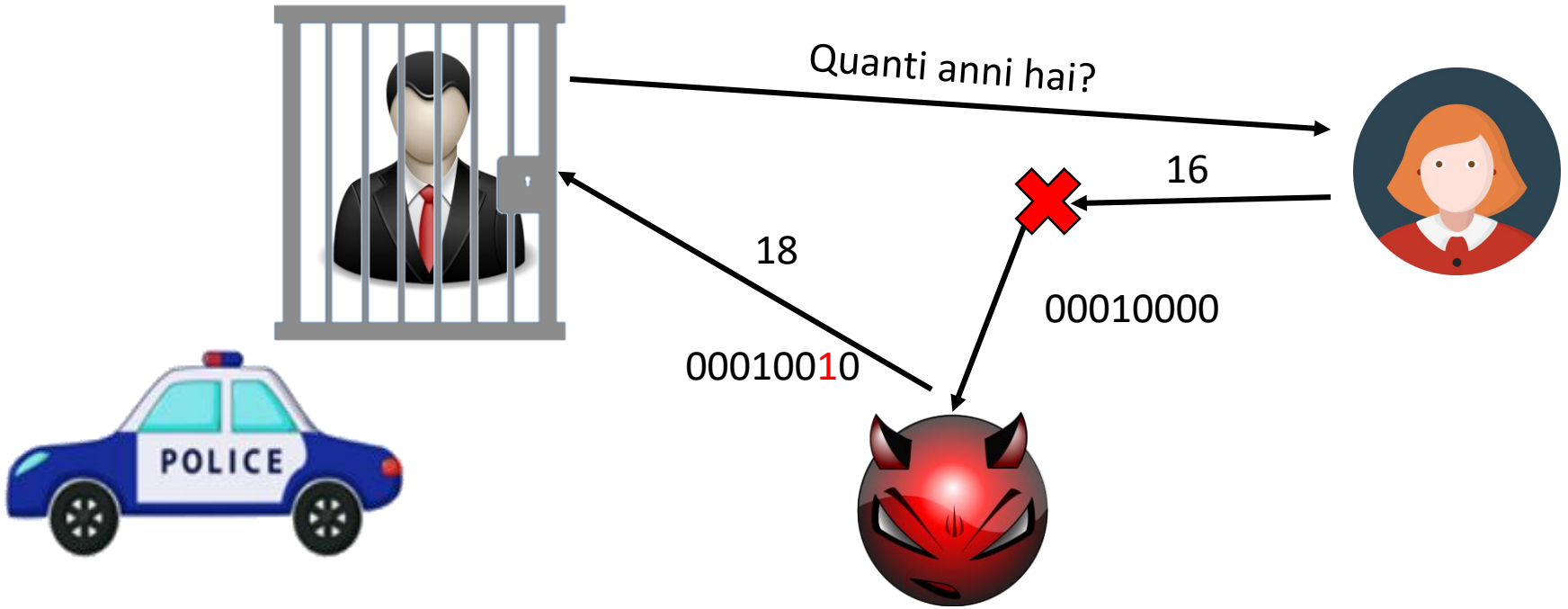
Funzioni Hash e Steganografia

Definizioni

Autenticazione sul dato: l'atto di verificare che un'entità A abbia generato il messaggio M. → Per esempio, una firma del professore sul libretto universitario, accanto al voto dato all'esame.

Integrità: proprietà di un messaggio M che garantisce l'assenza di modifiche dopo aver generato e mandato tale dato. → Per esempio un sigillo di cera posto dai nobili medievali su un messaggio da recapitare.

Integrità



Integrità

Anche cifrando, non sarebbe cambiato molto!
k chiave, M messaggio, E/D algoritmi, T testo cifrato

Alice: E(k,M) = T
Eva: T → perturbazione → T'
Bob: D(k,T')=M' → M' ≠ M

E Bob non ha alcun modo per accorgersi del cambiamento!

Funzioni Hash

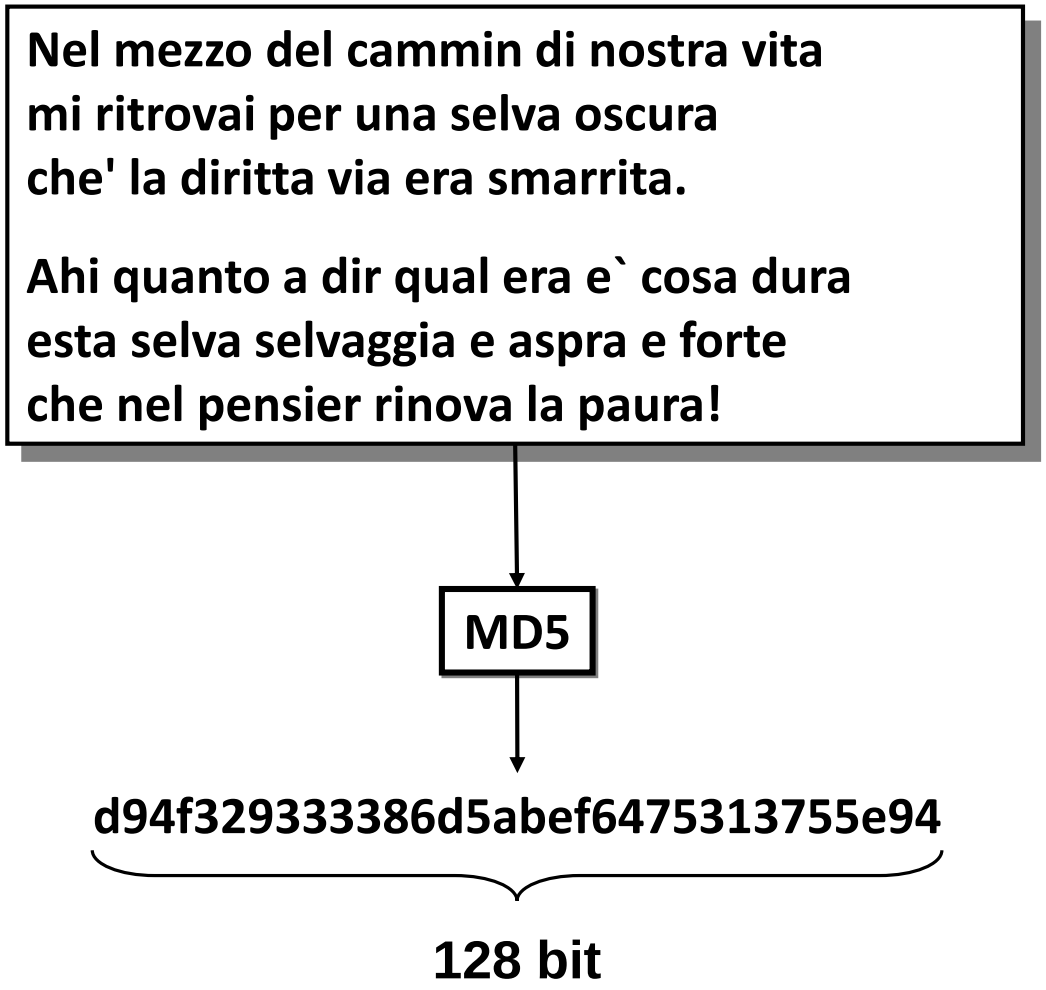
Per verificare l'integrità del dato si calcola un **digest**.
Il digest è una sorta di «riassunto» digitale: a seconda della **funzione hash** utilizzata, il digest cambia.

Una funzione Hash è solitamente rappresentata così:

Dgt = H(M)

Dove Dgt è il digest, M il messaggio, e H la funzione hash.

Esempio di Hash



Digest

Solitamente il digest, una volta decisa la funzione hash da utilizzare, ha dimensione fissa indipendentemente dal messaggio mandato in «pasto» alla funzione hash.

Le funzioni hash più famose sono:

- MD2, MD4, MD5 (vecchie e deboli, dgt 128 bit)
- SHA1 (rotto nel 2017, dgt 160 bit)
- SHA2 family (buona sicurezza, dgt 224-512)

Collisione (Aliasing)

La creazione di un digest implica una perdita di informazioni!

Idealmente, sarebbe opportuno che dati due messaggi M1 ed M2 succeda che H(M1) ≠ H(M2)… e di solito è così!

Per come sono costruite le funzioni Hash, però, dato M esiste **sicuramente** M' tale che H(M) = H(M').

Più digest si calcolano, più probabilità c'è di trovare una collisione!

Proprietà Funzioni Hash

Non invertibilità (one-wayness): una funzione hash deve essere impossibile (o molto difficile) da invertire. Dato H, H(M) deve essere difficile risalire ad M.

Resistenza alle collisioni (debole): dato H, H(M) deve essere computazionalmente difficile trovare X tale che H(M) = H(X).

Resistenza alle collisioni (forte): dato H, deve essere computazionalmente difficile trovare X e X' tali che H(X) = H(X').

Classificazione Funzioni Hash

One Way Hash Function (OWHF)
Offre one wayness e resistenza debole alle collisioni. È infatti detta anche weak one-way hash function.

Collision Resistant Hash Function (CRHF)
Offre resistenza forte alle collisioni e (solitamente) one wayness. È infatti detta anche strong one-way hash function.

Funzioni Hash con chiave

Data una funzione Hash H, una chiave kh, ed un messaggio M:
dgt1 = H(kh_1,M)
dgt2 = H(kh_2,M)
Il digest dgt cambia al variare di M e kh.
Nonostante M rimanga uguale nei due calcoli, dgt1 e dgt2 sono diversi.
Se Bob ed Alice vogliono comunicare in maniera autenticata, possono usare un segreto condiviso (kh) per farlo.

Comunicazione con funzioni hash

Dati: kh chiave hash, k chiave simmetrica, E/D algoritmi, M messaggio, T il testo cifrato, H funzione hash.

Alice: genera M e calcola S=H(M,kh)
Alice: calcola T= E(k,M||S)
Bob: riceve T e ottiene M||S = D(k,T)
Bob: calcola H(M,kh) e confronta H(M,kh)==S

Questo utilizzo è **SICURO** tanto quanto lo schema di cifratura simmetrica utilizzato.

Attaccare una funzione hash

Un attacco alle funzioni hash ha successo se si riesce a trovare una collisione. Ci sono due tipi di attacchi:

Falsificazione Esistenziale (Existential forgery)
L'attaccante genera una coppia messaggio-hash [x, H(kh,x)], ma non ha controllo su x.

Falsificazione Selettiva (Selective forgery)
L'attaccante genera una coppia messaggio-hash [x, H(kh,x)], ed ha il controllo completo o parziale sul messaggio.

Paradosso del Compleanno

In una stanza con 23 persone…

La probabilità che almeno una persona sia nata il 25 dicembre è:
Pr(25Dic) = 1- (364/365)^23 = 0.061 →6,1%

La probabilità che almeno due persone condividano il compleanno:
Pr(collisione) = 1 – (364!/342!)/(365^22) = 0,507→50,7%

Paradosso del Compleanno

Possiamo applicare la stessa logica ad un attacco ad una funzione hash (falsificazione esistenziale) abbattendo il costo dell'attacco rispetto ad un tradizionale guessing attack (brute force). Con un digest di N bit…

Guessing attack → 2^{N}

Birthday attack → $2^{(N/2)}$

Questo vuol dire che i bit di sicurezza di una hash function sicuramente saranno minori o uguali della metà dei bit del digest.

Comunicazione con funzioni hash

Dati: kh chiave hash, k chiave simmetrica, E/D algoritmi, M messaggio, C il testo cifrato, H funzione hash.

Alice: genera M e calcola S=H(kh,M)
Alice: calcola T=E(k,S) e manda M||T
Bob: riceve M||T e ricava S=D(k,T)
Bob: calcola H(kh,M) e verifica H(kh,M)==S

Questo utilizzo è **NON SICURO** perché fornisce all'attaccante M in chiaro, che potrebbe essere utilizzato come «oracolo».

Comunicazione con funzioni hash

Dati: k chiave simmetrica, E/D algoritmi, M messaggio, C il testo cifrato, H funzione hash.

Alice: genera M e calcola S=H(kh,M)
Alice: calcola T=E(k,M) e manda T||S
Bob: riceve T||S e ricava M=D(k,T)
Bob: calcola H(kh,M) e verifica H(kh,M)==S

Questo utilizzo è **NON SICURO** perché fornisce all'attante S che può essere utilizzato come «oracolo».

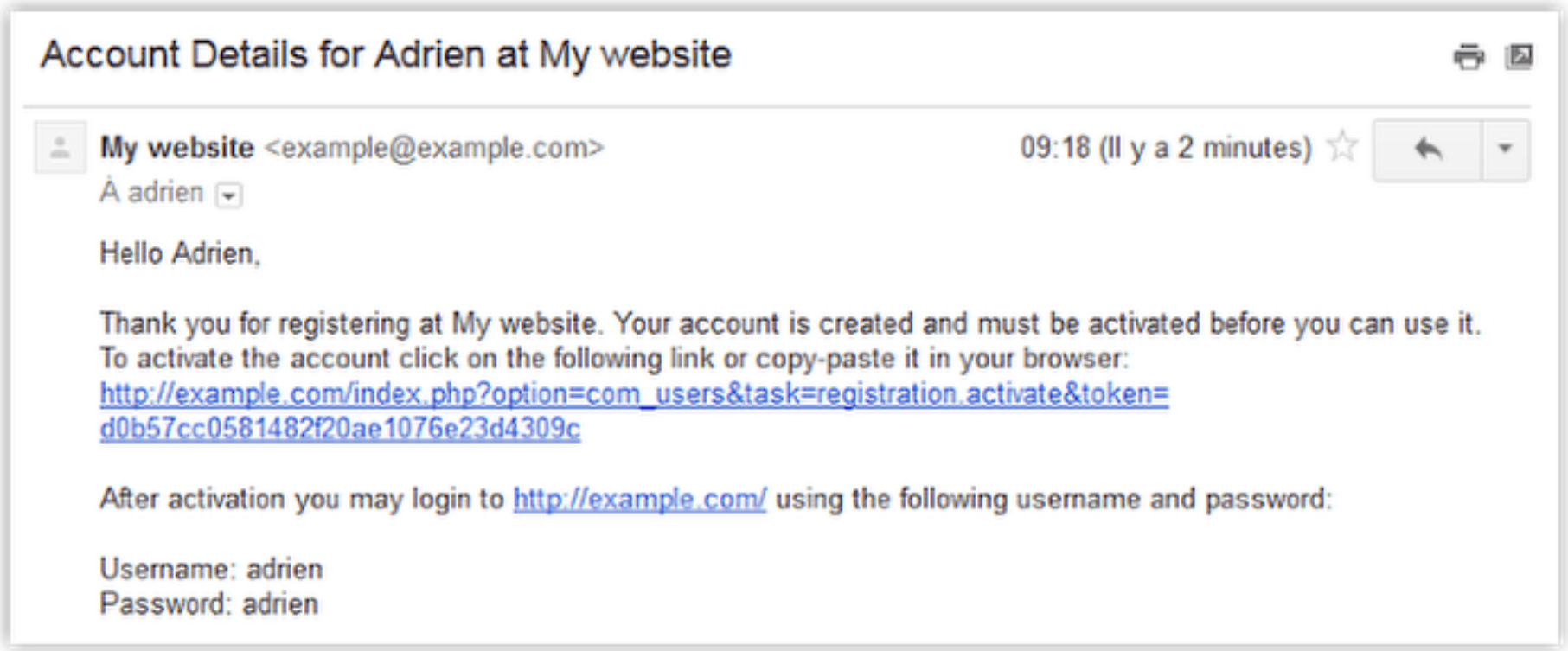
Utilizzi di hash

Un esempio di utilizzo per funzioni hash è l'autenticazione dell'utente.

Ogni applicazione web che richiede un login con username e password ha qualche implementazione di un pwdDB.

Nel pwdDB sono memorizzate i valori hash *(si spera)* delle password di ogni utente. *Si spera* perché non sono rari i casi di piccoli siti che memorizzano le password degli utenti in chiaro!

Utilizzi di hash



Account Details for Adrien at My website

My website <example@example.com> 09:18 (Il y a 2 minutes)

À adrien

Hello Adrien,

Thank you for registering at My website. Your account is created and must be activated before you can use it.
To activate the account click on the following link or copy-paste it in your browser:
http://example.com/index.php?option=com_users&task=registration.activate&token=
d0b57cc0581482f20ae1076e23d4309c

After activation you may login to http://example.com/ using the following username and password:

Username: adrien
Password: adrien

Utilizzi di hash

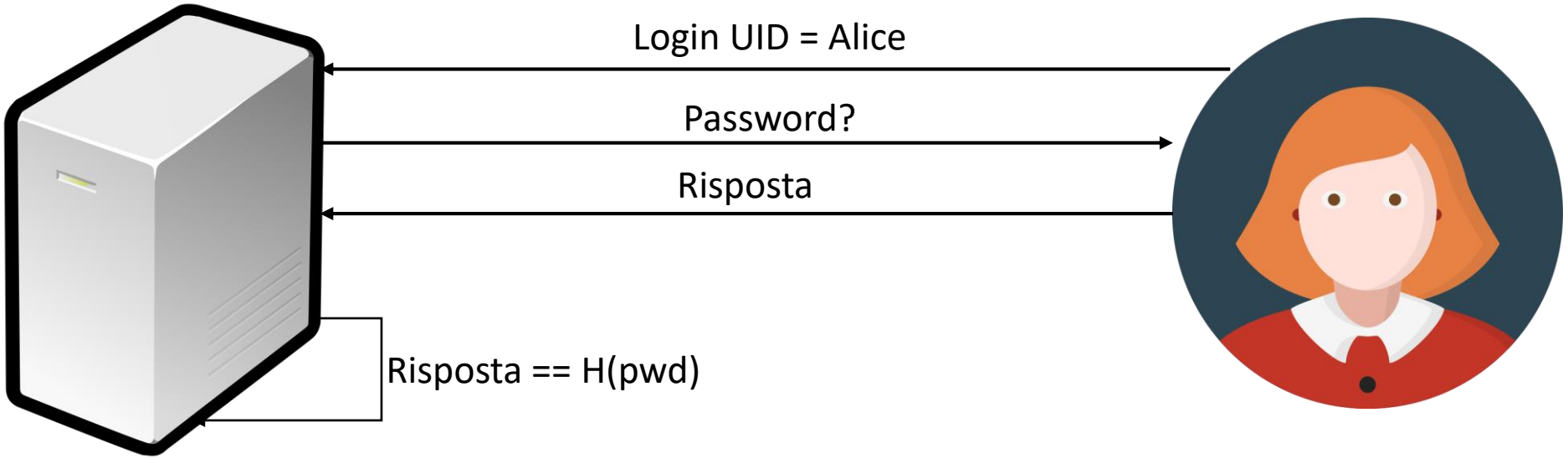


Tabella (UID, H(pwd))

MD5 e passwords

Molti utenti utilizzano password semplici da ricordare, come una data importante, od un semplice nome.

MD5 è un algoritmo debole ma ancora molto usato, pertanto molte delle password «più frequenti» sono state già scoperte e memorizzate in dizionari, per effettuare attacchi di lookup.

admin → 21232f297a57a5a743894a0e4a801fc3
1234321 → 6c30734811916b0f0f24a4630b08036f
password → 5f4dcc3b5aa765d61d8327deb882cf99

Altri possibili attacchi

Se i password database vengono bucati, un attaccante può recuperare le password nel caso siano in chiaro od oscurate con MD5 (tramite lookup o bruteforce).

- Entrare nell'applicazione web e spacciarsi per il legittimo utente.
- Provare a riutilizzare le combinazioni UID-password per entrare in altre applicazioni web con le stesse credenziali.
- Contattare la vittima ed estorcerle denaro.

Altri utilizzi di Hash Function: OTP

Le One Time Password sono una «raccolta» di password generate utilizzando il metodo delle «catene di hash».

L'utente sceglie una password P, e genera altre N password con la seguente procedura:

$P_0 = P$;

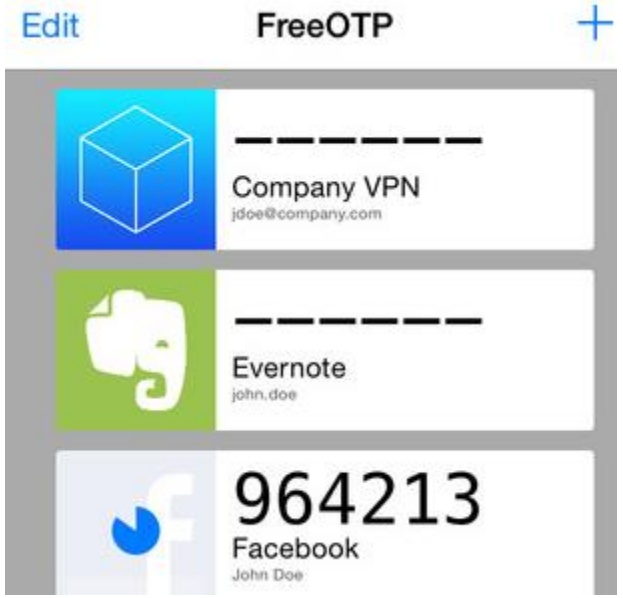
$P_1 = H(P_0)$;

$P_i = H(P_{i-1})$;

E comunica P_N al server. Per le successive N autenticazioni, l'utente manda P_i al server, che ne esegue l'hash e lo confronta con P_{i+1} che già possedeva.

Altri utilizzi di Hash Function: Time-Based OTP

Le TOTP sono una variante di OTP basata sull'orario in cui si cerca di effettuare l'autenticazione.
L'utente solitamente ha un dispositivo od un'applicazione che, sulla base dell'orario attuale e di un segreto inserito all'interno del dispositivo stesso, fornisce un codice di 6 cifre.



Per approfondire

Tool per fare look-up di hash MD5 online:
https://crackstation.net/

Per saperne di più (Bruteforcing su password oscurate con MD5): https://www.youtube.com/watch?v=7U-RbOKanYs

STEGANOGRAFIA

Steganografia

La parola steganografia proviene dal Greco e letteralmente significa “scrittura coperta”. Il significato è ovviamente veicolare un messaggio in modo che questo non venga intercettato da una controparte.

Al contrario della crittografia, però, la steganografia non cerca di rendere il messaggio incomprensibile, ma cerca di nasconderlo, di farlo passare inosservato!

Di solito, il messaggio nascosto è parte di qualcos'altro: immagini, video, brani di testo, e molto altro.

Steganografia - Esempi

«Infatti Istieo, volendo comunicare ad Aristagora l'ordine di insorgere, non aveva sistema sufficientemente sicuro per avvisarlo, dato che le strade erano tutte sotto controllo; allora, rasato il capo al più fidato dei suoi servi, vi tatuò dei segni, attese che ricrescessero i capelli e appena furono ricresciuti lo mandò a Mileto con il solo incarico, una volta giuntovi, di invitare Aristagora a radergli i capelli e a dargli una occhiata sulla testa. Il tatuaggio ordinava, come ho già detto, la ribellione.»

Erodoto, libro V delle storie

Steganografia - Esempi

«Demarato che si trovava a Susa e ne venne a conoscenza, volle informarne gli Spartani. Non aveva altri sistemi per avvisarli, giacché correva il rischio di essere scoperto, e quindi escogitò questo sotterfugio: prese una tavoletta doppia, ne raschiò via la cera e poi incise sul legno della tavoletta la decisione del re; dopodiché riversò della cera sullo scritto, affinché la tavoletta, non contenendo nulla, non procurasse noie a chi la portava da parte delle guardie delle strade. Quando essa giunse a Sparta, gli Spartani non riuscivano a raccapezzarsi finché un suggerimento non venne da Gorgo, figlia di Cleomene e moglie di Leonida, che ci era arrivata da sola: li esortò a raschiare via la cera e avrebbero trovato il messaggio inciso nel legno.»

Erodoto, libro VII delle storie

Steganografia - Esempi

«Un microdot è un testo o un'immagine che viene considerevolmente ridotta per nascondere l'informazione da destinatari inconsapevoli. I microdot normalmente sono circolari con un diametro di circa un millimetro ma possono essere di varie forme e dimensioni, oltre che a esser realizzati con vari materiali come poliestere e metallo. Il nome deriva dal fatto che i microdot spesso avevano la dimensione e la forma di un punto tipografico. Erano spesso nascosti sotto forma di punto di fine periodo o come punto sulle lettere i e j. I Microdot rappresentano un approccio steganografico alla protezione dei messaggi.»

Wikipedia

Steganografia

Steganografia iniettiva:

- Inserire il messaggio (segreto) all'interno di un altro messaggio (contenitore), in modo tale da non risultare visibile all'occhio umano e da essere praticamente indistinguibile dall'originale.

Steganografia generativa:

- Selezionare un messaggio (segreto) e costruirgli intorno un opportuno contenitore.

Steganografia iniettiva

Steganografia generativa

Oggi sono andato sul monte, e mi sono addentrato nei boschi… o mangiato una bistcca ala brace e bevuto un calice di vino novelo. Stasera vedrò la partita con Antnio in quel nuovo pub irlandese «Anyay» che vuoldire «ad ogni mdo» in inglese. Mi dispiace scocciati sempre con i dettagli del mio week-end. Ah, un'atra cosa: domani esco con Madalena, tua sorella, spero non ti dispiaccia!

Oggi sono andato sul monte, e mi sono addentrato nei boschi… **h**o mangiato una bist**e**cca a**l**la brace e bevuto un calice di vino novel**l**o. Stasera vedrò la partita con Ant**o**nio in quel nuovo pub irlandese «Any**w**ay» che vuol dire «ad ogni m**o**do» in inglese. Mi dispiace scocciar**r**ti sempre con i dettagli del mio week-end. Ah, un'a**l**tra cosa, domani esco con Mad**d**alena, tua sorella, spero non ti dispiaccia!

Hello world

STEGANOGRAFIA GENERATIVA

Tecniche Steganografiche: Steganografia di stampa

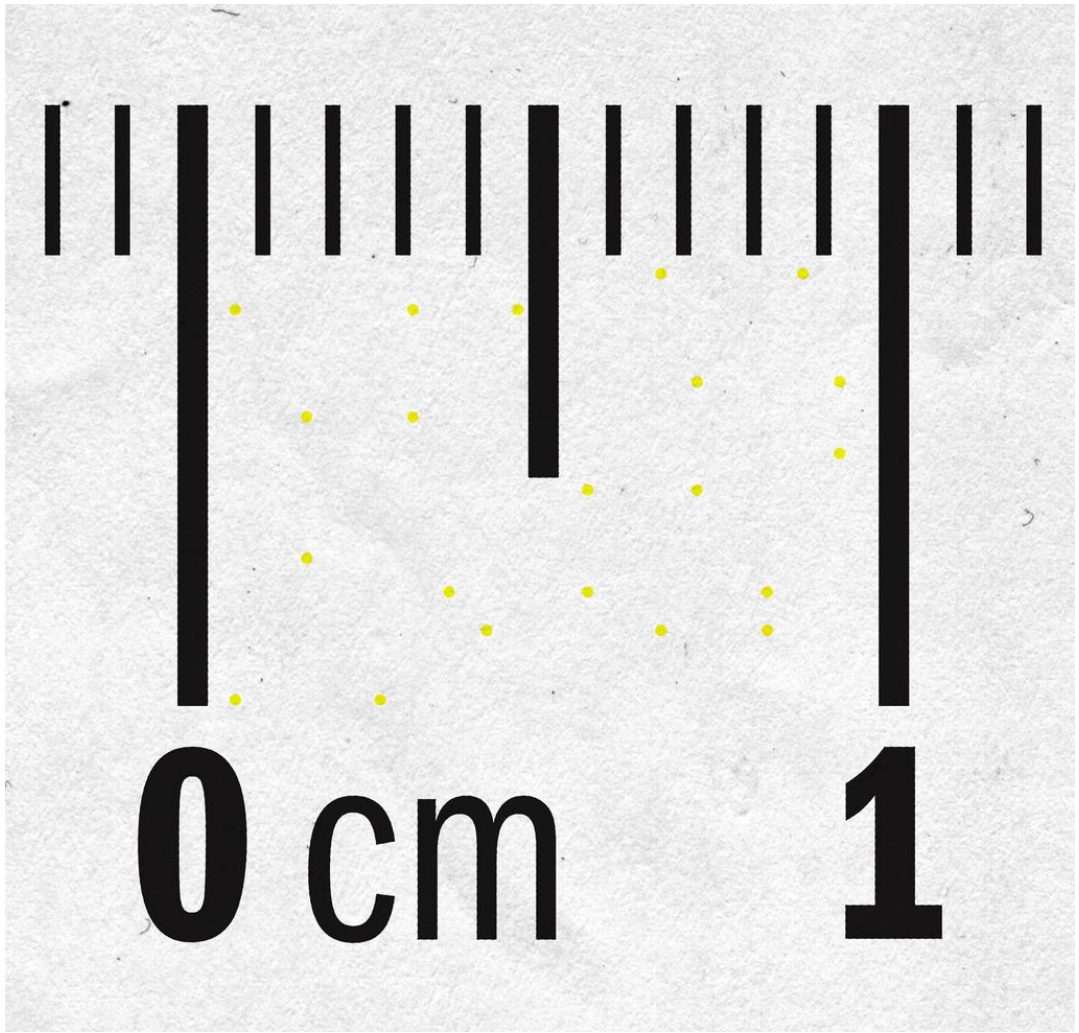
La steganografia di stampa è usata nelle stampanti laser a colori di diversi produttori. Dei minuscoli puntini gialli sono stampati su ogni pagina. I puntini sono appena visibili e contengono informazioni come il numero di serie della stampante, la data e l'ora della stampa.

Le stampanti a colori sono quelle in cui il sistema è maggiormente applicato. La misura è stata presa durante gli anni '90 da aziende come Xerox per convincere i governi sul fatto che le loro stampanti non possano essere usate per la contraffazione. L'identificazione avviene per mezzo di una filigrana, spesso formata da puntini gialli su sfondo bianco, situata su ogni foglio stampato, e può essere usata per identificare la macchina che ha stampato il documento in un'ampia gamma di stampanti. Può essere un testo oppure un motivo composto da puntini ripetuto in tutta la pagina, visibile con una luce blu o con una lente di ingrandimento abbastanza potente, ed è stato studiato per essere molto difficile da distinguere ad occhio nudo.

L'Electronic Frontier Foundation (EFF) ha craccato nel 2005 i codici delle stampanti DocuColor della Xerox ed ha pubblicato in rete una guida per individuarli.
L'EFF ha anche pubblicato un elenco delle stampanti che producono una steganografia di stampa attraverso un codice formato da puntini gialli.

https://it.wikipedia.org/wiki/Steganografia_di_stampa

Tecniche Steganografiche: Steganografia di stampa



Tecniche Steganografiche: Steganografia LSB

Si basa sulla teoria secondo la quale l'aspetto di un'immagine digitale ad alta definizione non cambia se i colori vengono modificati in modo impercettibile.
Ogni pixel è rappresentato da un colore differente, cambiando i bit meno significativi di ogni pixel, il singolo colore non risulterà variato in modo significativo e il contenuto dell'immagine sarà preservato nonostante questa manipolazione.

Nello standard bitmap a 24 bit di profondità ogni pixel è codificato con:

- 1 byte per codificare il rosso (valore 0 – 255)
- 1 byte per codificare il verde (valore 0 – 255)
- 1 byte per codificare il blu (valore 0 – 255)

Tecniche Steganografiche: Steganografia LSB

Il colore rosso, per esempio è codificato con la tripletta (255,0,0) mentre il colore lavanda è codificato con la tripletta (230, 230, 250).

Rosso

R	1	1	1	1	1	1	1	1
G	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0

Lavanda

R	1	1	1	0	0	1	1	0
G	1	1	1	0	0	1	1	0
B	1	1	1	1	1	0	1	0

ENC_LSB 2 = RES

R	1	1	1	1	1	1	1	1
G	0	0	0	0	0	0	1	1
B	0	0	0	0	0	0	1	1

DEC_LSB 2 = RES

R	1	1	0	0	0	0	0	0
G	1	1	0	0	0	0	0	0
B	1	1	0	0	0	0	0	0

Tecniche Steganografiche: Steganografia LSB

Il colore rosso, per esempio è codificato con la tripletta (255,0,0) mentre il colore lavanda è codificato con la tripletta (230, 230, 250).

Rosso

R	1	1	1	1	1	1	1	1
G	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0

Lavanda

R	1	1	1	0	0	1	1	0
G	1	1	1	0	0	1	1	0
B	1	1	1	1	1	0	1	0

ENC_LSB 4 = RES

R	1	1	1	1	1	1	1	0
G	0	0	0	0	1	1	1	0
B	0	0	0	0	1	1	1	1

DEC_LSB 4 = RES

R	1	1	1	0	0	0	0	0
G	1	1	1	0	0	0	0	0
B	1	1	1	1	0	0	0	0

Tecniche Steganografiche: Steganografia LSB

Link utili:

Steganografia LSB, esempio:
https://www.petitcolas.net/steganography/image_downgrading/

Steganografia LSB guida python:
https://towardsdatascience.com/steganography-hiding-an-image-inside-another-77ca66b2acb1