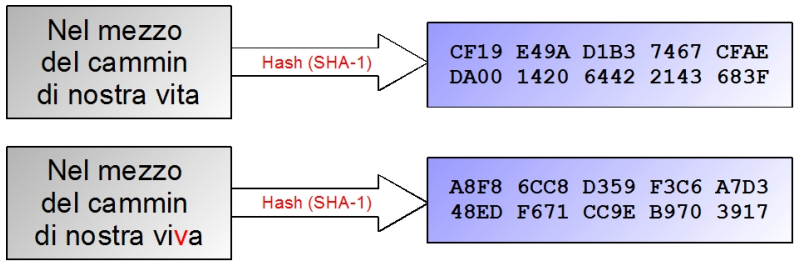
Metodologie e strumenti per la crittoanalisi della funzione di hash SHA-1 e sue implicazioni sulla sicurezza di rete

Luigi Esposito

Secondo Borsisti Day, Roma, 23.02.2011

Funzioni di hash crittografiche

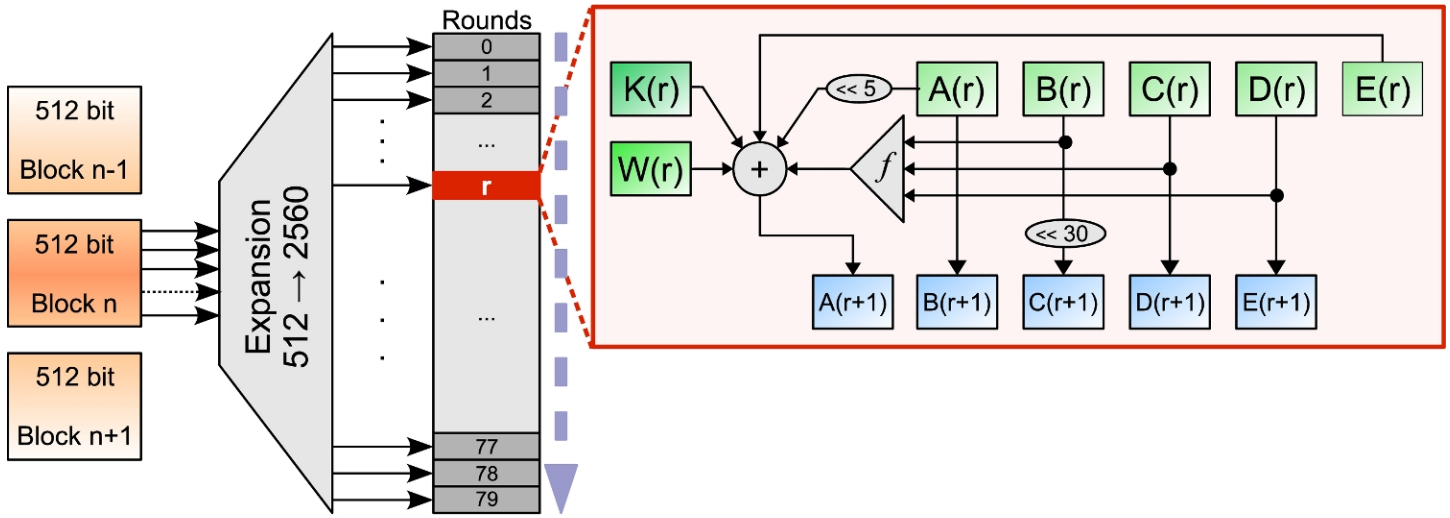
- Associano a messaggi di lunghezza arbitraria brevi stringhe di lunghezza fissa, dette valori di *hash*



- Tre proprietà rispetto alle funzioni di hash classiche:
  - Resistenza all' individuazione della prima preimmagine
  - Resistenza all' individuazione della seconda preimmagine
  - Resistenza alle collisioni
- La capacità di resistere alle collisioni è la prima proprietà ad essere attaccata
  - Funzione considerata rischio quando è scoperto un attacco migliore del *birthday attack*
  - Viene meno il ruolo principale della funzione: quello di essere una "firma" virtualmente univoca del messaggio

SHA-1

- SHA-1 è stata introdotta nel 1995 dal National Institute of Standards and Technologies, USA
- Basata sull' esecuzione di 80 iterazioni di una **funzione di compressione**
  - La funzione di compressione aggiorna ad ogni iterazione 5 registri di stato, usando anche il messaggio originario
  - I valori finali dei registri sono concatenati a formare l' hash



Applicazioni di SHA-1

- Verifica dell'integrità ed autenticità dei messaggi in vari protocolli di rete
  - A livello applicazione (PGP, S/MIME, SSH)
  - A livello trasporto (TLS/SSL)
  - A livello networking (IPsec)
- Identificazione di file, dati o software
  - Controllo versione software: Git, Mercurial, Monotone
  - Identificazione file (ad es. nei P2P e negli archivi)
- Derivazione di sequenze di chiavi e password

Attacchi – Crittanalisi differenziale

- Le vulnerabilità nelle funzioni di hash sono oggetto di un'intensa attività di ricerca:
  1. per allertare con sufficiente anticipo gli enti responsabili
  2. per comprendere a fondo le caratteristiche del processo di hashing e prevenire l'introduzione di analoghe debolezze nelle nuove funzioni promosse come standard

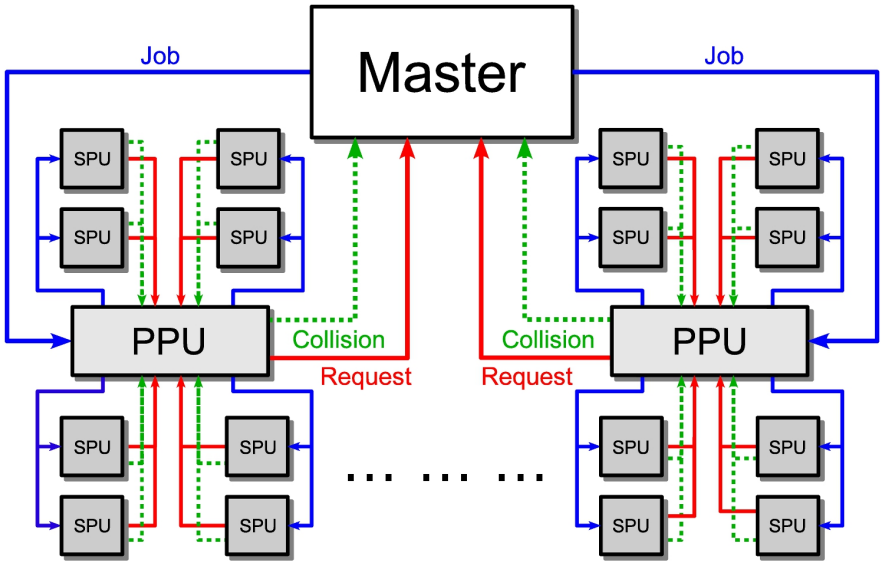
- La *crittanalisi differenziale* studia come differenze introdotte negli input si ripercuotono negli stati intermedi e negli output
- L'obiettivo è individuare collisioni, ovvero trovare coppie di messaggi diversi che generano lo stesso hash
- Una *caratteristica differenziale* è un insieme di vincoli che si impongono tra bit omologhi:
  - dei messaggi di input
  - degli stati intermedi dei registri durante tutte le iterazioni della funzione di hash

Tool per la ricerca di collisioni 1/2

- Sviluppo di un tool per la ricerca di collisioni, in due fasi:
  1. Strumento per la costruzione di caratteristiche differenziali
  2. Strumento per la ricerca di collisioni a partire da una caratteristica

- Porting dell'applicazione per il processore IBM Cell BE
  - Multicore: 1 PPU ed 8 SPU
  - SIMD (128 bit)
- Applicazione testata sul cluster prototipo MariCel del BSC
  - dotato di 144 PowerXCell 8i

- Porting (per ora virtuale) per array di FPGA in collaborazione con il tutor

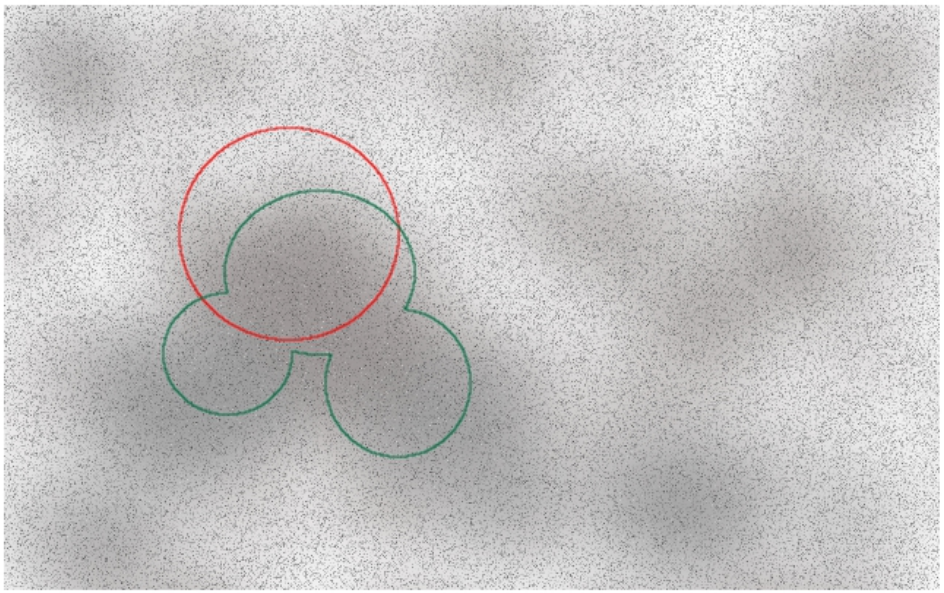


Tool per la ricerca di collisioni 2/2

- L'applicazione è ottimizzata in modo spinto, in quanto anche piccoli dettagli portano grosse differenze temporali nell'esecuzione.
  - L'ultima versione ha sorgenti meta programmati in funzione dell'input
- Ideate nuove tecniche complementari alle caratteristiche differenziali, progressivamente integrate nell'applicazione:
  1. Interbit constraints
  2. Constraint relaxation
  3. Splitting patterns

Spazio di ricerca delimitato dalla sola caratteristica

Spazio di ricerca delimitato da caratteristica + nuove tecniche



Interbit constraints

- La caratteristica differenziale è una struttura che definisce relazioni che intercorrono tra bit in posizioni omologhe di due diversi messaggi
- L'innovazione è stata quella di individuare particolari schemi di relazioni tra bit in posizioni *diverse* che consentissero un aumento della probabilità di collisione
- A fronte di un moderato aumento della complessità, si è ottenuto un fattore di incremento di velocità nella ricerca notevole (da 4-8x fino a 32x)

Constraint relaxation

- Non tutti i vincoli, generati per la ricerca dei messaggi con la costruzione tradizionale della caratteristica, si sono rivelati essere realmente necessari
- Alcuni di essi imponevano infatti condizioni eccessivamente stringenti
  - Di fatto, lo spazio di ricerca risultava ridotto arbitrariamente
  - Coppie di messaggi aventi una buona probabilità di collidere venivano subito scartate
- Con un costo computazionale irrisorio è stato possibile rimuovere tali vincoli

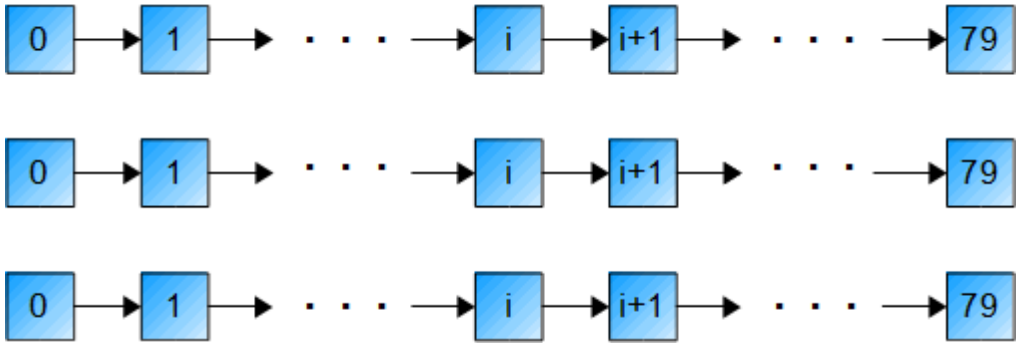
Splitting patterns 1/2

- Nella strategia tradizionale di ricerca, si analizza una coppia di messaggi per volta
  - Per ciascuna coppia si eseguono le iterazioni previste dall'algoritmo
  - La ricerca si arresta quando si riesce a stabilire che i messaggi non collideranno (o quando si trova la collisione)

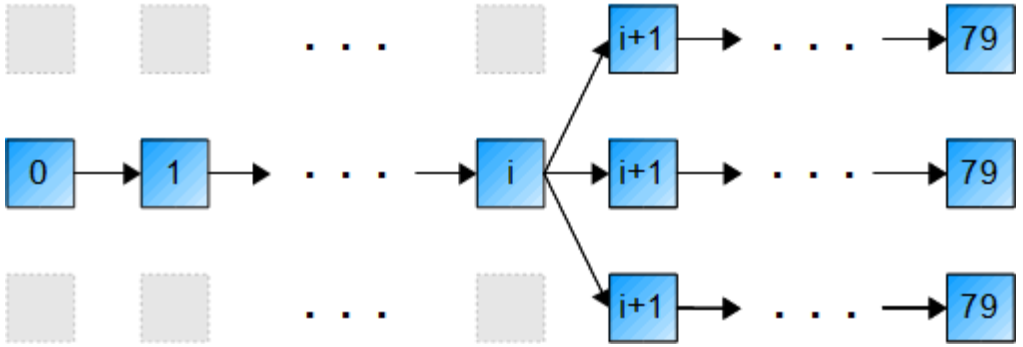
- L'innovativa tecnica ideata invece consente di analizzare in parallelo le prime iterazioni di molteplici coppie di messaggi

Splitting patterns 2/2

- Funzionamento tradizionale



- Funzionamento con applicazione Splitting patterns



Risultati – Collisioni

- All'avvio del progetto, erano note collisioni per versioni ridotte a 70 round di SHA-1 (T.Peyrin e C. Rechberger et al.)

- Partendo dal loro lavoro, grazie al tool sviluppato integrando le nuove tecniche
  - Maggio 2010: collisione a 71 round
  - Luglio/Agosto 2010: collisione a 72 round

- Abbiamo dunque rivolto la ricerca negli ultimi mesi verso ulteriori possibilità che consentano di raggiungere collisioni per SHA-1 a 73/74 round (tecnica degli *splitting patterns*)

Collisioni a SHA-1 71- e 72-round

71-round Collision							
Message 1				Message 2			
Word	Value (Hex)	Word	Value (Hex)	Word	Value (Hex)	Word	Value (Hex)
0	A031284A	16	C66928E5	0	10312819	16	766928B6
1	8E0B07E7	17	B8D273A2	1	BE0B07EF	17	88D273AA
2	259E60AA	18	136947A4	2	259E60E9	18	136947E7
3	26865A7F	19	4C7277A5	3	F6865A0D	19	9C7277D7
4	3F7A9945	20	87640DAB	4	8F7A9955	20	37640DBB
5	1F3B9AF1	21	1E439843	5	EF3B9A93	21	EE439821
6	B79EC755	22	8F78C12A	6	779EC717	22	4F78C168
7	41CB1152	23	7EA5FA75	7	41CB1162	23	7EA5FA45
8	311807D8	24	582F0B12	8	D118079A	24	B82F0B50
9	EA18241F	25	80286705	9	CA18247F	25	A0286765
10	C126D406	26	BA240B89	10	2126D447	26	5A240BC8
11	AAF12C3D	27	8A955AE7	11	8AF12C6D	27	AA955AB7
12	7F82700C	28	AB5CB1CA	12	BF82704D	28	6B5CB18B
13	464034BF	29	2BC7E7B5	13	A64034CD	29	CBC7E7C7
14	EF55783A	30	4132CCAA	14	4F557839	30	E132CCA9
15	B805685B	31	A57DD240	15	78056849	31	657DD252
Hash	89EE5B219C39AAB795FEED4483361F39D9B52E69						

72-round Collision							
Message 1				Message 2			
Word	Value (Hex)	Word	Value (Hex)	Word	Value (Hex)	Word	Value (Hex)
0	E03BE94A	16	17E06210	0	503BE919	16	A7E06243
1	55429082	17	4228938E	1	6542908A	17	72289386
2	51F58CAD	18	8F080AE2	2	51F58CEE	18	8F080AA1
3	165504EB	19	69723D6F	3	C6550499	19	B9723D1D
4	6FA80C12	20	28629C47	4	DFA80C02	20	98629C57
5	C0C0E90B	21	371A4D30	5	30C0E969	21	C71A4D52
6	A571346E	22	8CFECD5F	6	6571342C	22	4CFECD1D
7	F541EB71	23	6A92FE7F	7	F541EB41	23	6A92FE4F
8	741493FE	24	12C368B9	8	941493BC	24	F2C368FB
9	E46596B8	25	4C5C4030	9	C46596D8	25	6C5C4050
10	257FCC2C	26	15E3BB2D	10	C57FCC6D	26	F5E3BB6C
11	68A706C2	27	DEA02DF5	11	48A70692	27	FEA02DA5
12	7A282A59	28	3D54D825	12	BA282A18	28	FD54D864
13	467204D2	29	C5FE48AA	13	A67204A0	29	25FE48D8
14	A3B36574	30	1A280A1F	14	03B36577	30	BA280A1C
15	0E1D76B2	31	D2124E30	15	CE1D76A0	31	12124E22
Hash	BD28230B0A52F51F0FEA3B7EFCBE8120EE1C4036						

Disseminazione

- Un primo articolo è stato pubblicato per una conferenza del luglio scorso (collisione a 71 round)
  - A. Cilardo, L. Esposito, A. Veniero, A. Mazzeo, V. Beltran, E. Ayugadé, *A CellBE-based HPC application for the analysis of vulnerabilities in cryptographic hash functions*, submitted to ***HPCC 10***, 2010
- Un secondo articolo, in fase di preparazione, è stato posticipato per includere nuovi risultati
- Verrà in ogni caso redatto un survey generale sull'attività, che illustrerà in dettaglio le tecniche sviluppate ed i risultati raggiunti
- E' stato inoltre realizzato un sito web che illustra la problematica, gli avanzamenti ed i risultati raggiunti:
  - http://www.hashproject.eu

Altre direzioni di sviluppo

- Con il tutor ed altri importanti gruppi di ricerca attivi nel settore a livello europeo è stato formato un consorzio informale, nell'ottica di creare un europeo: **Exacrypt**
  - Tra i partner figurano: TU Eindhoven, Univ. Ruhr, TU Dortmund, Univ. Bristol, BSC, UC Louvain, KU Leuven
- La linea di ricerca avviata nell'ambito della borsa sarà parte delle tematiche affrontate dal progetto
- L'obiettivo globale è la valutazione dell'impatto delle tecnologie emergenti su crittosistemi di varia natura
- Si propone di fornire linee guida (dimensione chiavi, parametri, …) per la sicurezza dei crittosistemi nei prossimi anni (prossima decade)

Implicazioni per SHA-2 e SHA-3

- SHA-2 è poco studiata
  - Poco diffusa (incompatibilità con Windows prima di XP SP2)
  - Strutturalmente molto simile a SHA-1 (anche se più complesso).
    - Qualora si trovasse un attacco molto efficiente per SHA-1, SHA-2 ne risentirebbe

- SHA-3 è in fase di definizione, mediante concorso aperto
  - Di 64 proposte, 5 candidate rimasti in lizza dopo il terzo round
  - Le concorrenti hanno strutture e funzionalità piuttosto eterogenee, legate in diversa misura a costruzioni preesistenti e soluzioni innovative
    - In diversi casi la costruzione è ispirata ad AES, in altri basata sul nuovo framework HAIFA, o alle *sponge functions*
    - *Randomized hashing*, che introduce il *salt*: $H_s(M)$
    - *Multiple* (double in genere) pipe

Proposta di estensione dell' attività 1/2

- Momento storico fondamentale per le funzioni di hash: è in corso di definizione lo standard SHA-3 (FR Notice Vol. 72 No. 212)
  - Una volta standardizzata, SHA-3 sostituirà SHA-1 ed altre funzioni in tutte le applicazioni principali
  - Candidate finaliste: BLAKE, Grøstl, JH, Keccak, Skein
- La proposta di estensione si configura come un prolungamento dell' attività già svolta:
  - Valutazione di se e come le vulnerabilità individuate per SHA-1 possano influenzare la scelta dei candidati di SHA-3
    - Determinazione della possibilità di estendere o adattare le tecniche ideate alle nuove funzioni
  - Analisi crittanalitica dettagliata delle candidate, individuando punti deboli e punti forti di ciascuna

Proposta di estensione dell'attività 2/2

- Definizione di approcci e tecniche finalizzati ad abbassare la complessità della ricerca di collisioni per i candidati a SHA-3
- Presentazione di lavori scientifici e rapporti tecnici relativi alle caratteristiche di sicurezza dei diversi candidati
- Punto centrale dell'attività: preparazione di un rapporto tecnico, eventualmente patrocinato dal GARR, inerente le vulnerabilità individuate per SHA-1 ed estese ai candidati di SHA-3
  - Verrà sottoposto al gruppo del NIST che si occupa della selezione per SHA-3: il *Cryptographic Technology group*
  - Sarà redatto entro il terzo/ quarto trimestre 2011, in tempo utile per la selezione del vincitore (prevista 2012)