Attacchi ai FileSystem cifrati

Negli articoli precedenti abbiamo illustrato alcuni modi per cifrare dischi, partizioni eccetera, occupandoci di aspetti pratici per indicare le principali problematiche tecniche coinvolte nella implementazione di tali procedure. Non abbiamo però affrontato due punti comunque interessanti: il primo riguarda l'utilità/applicabilità di queste tecnologie ed il secondo le problematiche di sicurezza. E' abbastanza ovvio che la cifratura di dati in grandi quantità sia utile in molte situazioni, basta ad esempio citare la problematica della cifratura dei backup ed associarla ai noti incidenti occorsi recentemente che riguardavano nastri di backup contenenti informazioni riservate *persi* nei trasporti [1]. Ma in questa sede vorremmo occuparci delle problematiche di sicurezza, il che vuol dire che dobbiamo parlare di crittografia. Per cercare di essere pratici, seguendo il filone di questa rubrica, cerchiamo di illustrare di semplici attacchi ai filesystem cifrati che dovrebbero dare una indicazione sulla difficoltà del problema.

Cifrare a blocchi

Inevitabilmente dobbiamo fare un mini-riassunto di tecniche crittografiche moderne. Gli algoritmi che ci interessano sono gli algoritmi simmetrici, quali DES[1] e AES. Il loro funzionamento è basato sul dividere i dati da cifrare/decifrare in blocchi di lunghezza fissa, e cifrare ogni blocco con la stessa chiave. Per concretezza consideriamo il caso di cifratura di un intero disco fisico immaginando di utilizzare un device crittografico hardware. I parametri che ci interessano maggiormente sono:

1. la lunghezza di ogni *blocco* di dati da cifrare: di solito 64 o 128 bit (8 o 16 byte)
2. la lunghezza della *chiave* usata per cifrare: negli algoritmi più noti da 56 a 256 bit (da 7 a 32 byte)
3. la dimensione di un *settore* fisico di un disco: di norma 4096 bit (512 byte).

1 DES non è sicuro poiché con una chiave di soli 56 bit, è suscettibile ad attacchi di forza bruta, per questo in pratica si utilizza 3DES, ovvero la composizione di 3 operazioni DES con chiavi diverse; questo fatto però non è importante per quello che segue.

Ovviamente i primi due dipendono dall'algoritmo crittografico scelto, ad esempio DES ha un blocco di 64 bit a chiave di 56 bit mentre AES ha un blocco di 128 bit e chiave di 128, 192 o 256 bit.

Ricordiamo inoltre che la principale tecnica di crittoanalisi, è quella della ricerca di ripetizioni o strutture nel testo cifrato. Più il testo cifrato appare casuale, più è sicuro; al contrario più ripetizioni o strutture sono riconoscibili nel testo cifrato, più è facile che sia possibile risalire al testo in chiaro anche senza conoscere la chiave.

Supponiamo di cifrare ogni blocco di dati presente sul disco con la stessa chiave segreta ed in modo indipendente (detto Modo ECB, *Electronic Code Book*). Ovviamente avremmo subito dei problemi poiché è facile che vengano cifrati due blocchi di dati uguali (ad esempio intestazioni di messaggi od altro) e adottando questa procedura, la cifratura produrrebbe due testi cifrati identici. Per un buon crittoanalista sarebbe non troppo difficile estrarre molte informazioni dal disco così cifrato. Pertanto per cifrare un disco si adotta un'altra strategia. Prima di tutto i blocchi non vengono cifrati in modo indipendente ma concatenati con il Modo CBC (*Cipher Block Chaining*) o modi derivati da questo. Il funzionamento di CBC è molto semplice. All'inizio della catena si specifica un numero (meglio se pseudo-casuale) spesso non segreto detto *Vettore di Inizializzazione IV* lungo quanto un blocco, poi si prosegue come segue[2]:

- Cifratura: $C_0 = IV$; $C_n=E(P_n \times C_{n-1})$
- Decifrazione: $C_0 = IV$; $P_n= C_{n-1} \times D(C_n)$

ove E è l'operazione di cifratura, D quella di decifrazione, x l'operazione di XOR bit a bit, P_n il n-esimo blocco in chiaro e C_n il n-esimo blocco cifrato. Si noti come il n-esimo blocco cifrato dipende da tutti i blocchi precedenti e dall'IV, mentre l'operazione di decifrazione richiede la sola conoscenza del blocco cifrato precedente. E' chiaro che adottando questa procedura due blocchi in chiaro identici forniscono due blocchi cifrati differenti.

CBC e la cifratura di un disco

2 Nelle operazioni di cifratura e decifrazione per semplicità di notazione non abbiamo indicato la chiave segreta utilizzata.

Se CBC è molto più sicuro di ECB, la sua applicazione ad un disco fisico non è immediata poiché se cifrassimo tutto il disco in catena avremmo dei seri problemi di prestazioni: in pratica ogni qual volta si accedesse a dei dati sul disco saremmo costretti a ri-cifrare o decifrare tutti i dati. Pertanto si procede in modo diverso. Visto che le operazioni di accesso al disco fisico sono fatte di norma per settore fisico, si cifra ogni settore in modo indipendente. Il modo più semplice per fare ciò è di cambiare l'IV ad ogni settore ed il modo più veloce di farlo è quello di legare il valore dell'IV al numero del settore. Il modo più semplice di legare l'IV al numero del settore è quello di mettere l'IV uguale al numero del settore, in questo caso ovviamente l'IV è *pubblico*, nel senso che è noto a chiunque sia in grado di accedere al disco. In alternativa vi sono procedure per creare un IV diverso per ogni settore basate sia sul numero del settore che sulla chiave segreta[3], in questo modo solo chi è in possesso della chiave segreta può calcolarsi l'IV usato in ogni settore. Queste procedure ovviamente richiedono più risorse per effettuare il calcolo dell'IV per ogni settore, ma aiutano a prevenire alcuni tipi di attacchi quale ad esempio il Watermarking Attack.

Un attacco a CBC

Supponiamo di aver trovato due blocchi cifrati identici, bit a bit, sul disco: questa scoperta ci può essere utile se fossimo un crittoanalista? Chiamiamo i due blocchi identici C_n e C_m, ricordandoci di come i blocchi cifrati sono ottenuti dai blocchi in chiaro, e ricordando che la chiave di cifratura usata è sempre la stessa, otteniamo l'equazione $P_n \times C_{n-1} = P_m \times C_{m-1}$ che, utilizzando le proprietà dell'XOR può essere riscritta come $C_{m-1} \times C_{n-1} = P_m \times P_n$. Riassumendo:

se $C_n = C_m$ ==> allora $P_m \times P_n = C_{m-1} \times C_{n-1}$

In parole, se due blocchi cifrati sono identici, il crittoanalista ottiene l'XOR dei due corrispondenti blocchi in chiaro direttamente dall'XOR dei due blocchi cifrati precedenti, il che in pratica vuol dire che è possibile ottenere il testo in chiaro senza molto sforzo. Questa è ovviamente una debolezza di CBC che è in astratto indipendente dall'algoritmo crittografico utilizzato.

3 Quale ad esempio l'algoritmo ESSIV menzionato nei precedenti articoli.

La dimensione del *blocco*

Dobbiamo a questo punto chiederci quanto è probabile trovare due blocchi cifrati identici sullo stesso disco. Un buon algoritmo crittografico produce blocchi che si distribuiscono molto similmente ad una distribuzione statistica uniforme, il che vuol dire la probabilità di trovare due blocchi cifrati identici dipende principalmente dalla dimensione del blocco usato dall'algoritmo crittografico. Senza entrare nei dettagli dei calcoli, è ovvio che il numero di possibili diversi blocchi è 2 elevato alla lunghezza del blocco. In realtà l'attacco più efficace è basato su di un teorema matematico detto il *paradosso dei compleanni* e possiamo riassumere come segue le sue conseguenze.

Consideriamo un algoritmo crittografico con blocco di 64 bit, allora c'è una probabilità di circa il 50% di trovare due blocchi identici se ci cifrano con la stessa chiave 32 Gbyte di dati. DES e 3DES hanno un blocco di 64 bit, pertanto possono essere considerati sicuri solo se usati per cifrare al più qualche centinaio di Mbyte di dati con la stessa chiave. Si noti che questo vale anche per la cifratura delle comunicazioni, pertanto se ad esempio si usa 3DES su di una linea ad 1Gbps è necessario cambiare la chiave di cifratura ogni 5 minuti circa.

Se invece il blocco è di 128 bit, come ad esempio per AES, la probabilità di circa il 50% di trovare due blocchi identici avviene se ci cifrano con la stessa chiave 256 Exabyte (256 miliardi di Gbyte) di dati. Questa è una quantità di dati al momento non raggiungibile con alcuna tecnologia.

Conclusioni

Non è ovviamente questa la sede per addentrarci nei dettagli tecnici di questa problematica. Il nostro intento era solo di segnalare al lettore la complessità dell'argomento presentando i più semplici esempi. Bisogna comunque sottolineare che questo è sicuramente un campo che vedrà nel prossimo futuro dei vigorosi sviluppi, a partire ad esempio dai lavori dell'IEEE [1] o dalla implementazione del modo LWR presentato nel 2002 da Liskov, Rivest e Wagner [2].

Andrea Pasquinucci

pasquinucci@ucci.it

Riferimenti Bibliografici

[1] Si veda ad esempio: IEEE Security in Storage Working Group, *Standard Architecture for Encrypted Shared Storage Media*, draft IEEE P1619 e P1619.1, http://www.siswg.org/

[2] M. Liskov, R. Rivest e D. Wagner, *Tweakable block ciphers,* CRYPTO '02 (LNCS, volume 2442), 2002

[3] per una introduzione alle problematiche della cifratura di dischi si veda http://clemens.endorphin.org/LinuxHDEncSettings ed anche http://en.wikipedia.org/wiki/Disk_encryption