

Utilizzo di Advanced Forensic Format nell'Informatica forense

Stefano Fratepietro e Cesare Maioli
Cirsfid e Università di Bologna

Informatica forense e open EnCase

L'informatica forense è la disciplina che concerne le attività di individuazione, conservazione, protezione, estrazione, documentazione e ogni altra forma di trattamento e interpretazione del dato memorizzato su supporto informatico, al fine di essere valutato come prova nei processi giudiziari [1].

L'informatica forense ha strette connessioni con la sicurezza in quanto si ha interesse a raccogliere reperti digitali una volta che c'è l'ipotesi di un reato o di un comportamento irregolare che verosimilmente ha superato i meccanismi di controllo predisposti dagli amministratori di sistema per evitare intrusioni e alterazioni di una componente informatica.

Il prodotto leader del mercato per l'analisi forense è attualmente EnCase della Guidance Software [2] che consente di reperire, analizzare e presentare dati nell'uso professionale e investigativo da parte di numerose agenzie e forze dell'ordine in tutto il mondo e che è considerato in linea con gli standard internazionali per le analisi delle tracce informatiche. Esso utilizza un format proprietario per le immagine di dati digitali basato su ASR Data's Expert Witness Compression Format.

Il format del file Evidence File [3] contiene un bitstream fisico del disco acquisito - prefisso da un header che contiene meta informazioni sul caso in esame - intrecciato con i CRC per ciascun blocco di 64 settori (32 Kb), e seguito da un footer che contiene lo hash MD5 per l'intero bitstream. L'header contiene data e ora dell'acquisizione, il nome dell'operatore, note sulle acquisizione, una password opzionale e il proprio CRC. Il formato è comprimibile e su di esso si possono eseguire operazioni di search. La compressione si basa sui blocchi; tabelle di salto e puntatori sono mantenuti tra i blocchi e nell'header per migliorare le prestazioni. Le immagini di disco possono essere suddivise in file multipli (per esempio per memorizzare CD e DVD). I file non possono superare i due Gigabytes.

EnCase dunque memorizza l'immagine di un disco come una serie di pagine compresse univocamente individuabili e gestibili: ogni pagina può venire reperita in modo random e decompressa secondo le esigenze investigative.

EnCase consente inoltre di inserire meta informazioni sulle varie parti dei documenti sotto esame.

Sul primo punto va notato che compressori diffusi come gzip o bzip2 non consentono l'accesso random all'interno di un file compresso; sul secondo che la prassi corrente e inevitabile di inserire meta informazioni in un data base separato dal file sotto esame rende possibili smarrimenti, sovrapposizioni e disordine su informazioni spesso di interesse nei procedimenti giudiziari.

Il software EnCase è proprietario; in questi ultimi anni sono stati progettati con continuità e a forte ritmo [4] prodotti open source che presentano capacità analoghe a quelle di EnCase per la memorizzazione di copie di dati grezzi prelevati da hard disk che consentano di evitare la copia di enormi quantità di dati anche se il file in esame è di dimensioni contenute ovvero di poter accedere selettivamente a parti di file compressi, oltre a gestire in modo efficiente meta informazioni come numeri identificativi dei drive in esame, le date, l'identificativo dell'operatore coinvolto in quella indagine e simili.

L'iniziativa Common Digital Evidence Storage Format

Il rischio che reperti e basi di prove per i procedimenti giudiziari vadano persi o divengano inammissibili in giudizio è causato dalla presenza di format diversi per le immagini digitali, tipi diversi di reperti (si va dai log di reti a memorie di dispositivi mobili), caratteristiche e comportamenti diversi degli strumenti di analisi forense ed è accresciuto dalla assenza di standard condivisi e tecnicamente robusti che consentano il congelamento della situazione rilevata, garantiscano la catena di custodia dei reperti, e siano analizzati con strumenti disponibili a tutte le

parti coinvolte in un procedimento giudiziario; quest'ultimo punto suggerisce l'opportunità di soluzioni open source.

Le perdite di informazioni che si hanno convertendo dati grezzi rappresentati in formati diversi, la dimensione dei file sequestrati di cui vengono eseguite copie settore per settore, la generale mancanza di meta dati sono fattori ulteriori che complicano la situazione.

Oltre al citato EnCase sono significativi i seguenti formati di file: ProDiscover, PyFlag, RAID, SDi32, SMART della famiglia open source e ILook, SafeBack proprietari.

L'iniziativa Common Digital Evidence Storage Format [5] nasce per definire un formato open che risolva tali problemi basandosi sui formati attuali, sulle esigenze dell'utenza e sugli standard giudiziari. In più la cura della catena di custodia, la cui best practice attuale sembra essere la trascrizione manuale in quadernetti o verbali delle forze investigative degli hash MD5o SHA-1 delle immagini acquisite dai supporti sotto esame, e la flessibilità per tener conto di più forme di reperto digitale (traffico in rete, dump di memorie, struttura logica dei file) suggeriscono la necessità di una *evidence bag* [6] e associata targhetta digitale in cui raccogliere tutti i reperti e le informazioni che li riguardano in maniera compatta e standardizzata come si suole in scene criminis più tradizionali.

L'adozione di un formato standard incoraggia lo sviluppo e la commercializzazione di prodotti più maturi per le analisi forensi e facilita la cooperazione tra forze investigative nazionali e internazionali.

Il format Advanced Forensic Format

Advanced Forensic Format (AFF) è una recente implementazione [7] open source ed estensibile distribuita sotto licenza BSD modificata [8] di un formato che analogamente a quello di EnCase memorizza l'immagine in maniera compressa e indirizzabile e, a differenza di quello di EnCase, consente di memorizzare le meta informazioni sia all'interno del file che in un file esterno collegato a quello di riferimento.

AFF è articolato in due layer per tener conto della compatibilità in avanti e in indietro in riferimento a un periodo temporale: il data storage layer descrive come una serie di coppie nome e valore sono memorizzate in uno o più file di disco, in maniera indipendente sia dal sistema operativo che dell'ordine dei byte; il disk representation layer definisce una serie di coppie nome e valore che vengono utilizzate per memorizzare le immagini del disco e le meta informazioni associate.

E' interessante osservare che i progettisti hanno rinunciato alla idea originale di implementazione del data storage layer tramite una distribuzione open source di b-tree ritenendo che l'articolazione delle informazioni contenute in un b-tree fosse troppo complessa da spiegare, laddove se ne presentasse la necessità, in un dibattito giudiziario; pertanto è stato adottato un approccio più semplice basato su una struttura detta AFF segment [9], ripetibile e di lunghezza variabile. Ogni AFF segment consiste di un header, un nome del segmento, un flag di 32 bit, una area dati di lunghezza variabile, un footer. La lunghezza è memorizzata sia nello header che nel footer. Un file AFF inizia con un file header e termina con una directory che contiene la lista di tutti i segmenti del file e il loro offset in byte dall'inizio del file.

Il disk representation layer definisce nomi specifici di segmento per rappresentare informazioni sui dischi e meta informazioni. Queste possono essere memorizzate nello stesso AFF file dell'immagine oppure in un file separato; lo schema può essere memorizzato anche in un file XML. I segmenti di dati hanno tutti la stessa ampiezza che viene determinata al momento della creazione del file immagine. I segmenti possono essere compressi con lo strumento open source zlib o lasciati non compressi secondo scelte da compiere al momento dell'esecuzione.

Per rendere più usabile il sistema e sollevare i programmatori dalla comprensione di molti dettagli implementativi è stata costruita la libreria AFFLIB che fornisce un'astrazione semplice dei file

immagine AFF che appaiono come l'insieme di un data base nomi-valori e di un file standard che può essere aperto, letto, e acceduto in ricerca con chiamate di libreria.

Il codice AFF è fornito [10] assieme a un insieme di strumenti come un programma per eseguire l'imaging del disco, un programma di conversione da AFF a XML, un programma per convertire, nei due versi, file grezzi in file AFF.

Utilizzo di Advanced Forensic Format

Lo scopo che perseguiamo è di partecipare alla progettazione e implementazione di un "open EnCase" in linea con le iniziative dei paragrafi precedenti e riteniamo AFF uno strumento molto valido come formato di riferimento.

Per quanto riguarda AFF abbiamo eseguito alcune prove utilizzando:

- FCCU Gnu/Linux Forensic Boot CD 10.0 [11];
- immagini grezze di reperti relativi ad alcuni casi giudiziari per la conversione da dati grezzi a AFF;
- storage USB di vario tipo per la creazione di immagini in formato AFF.

FCCU è una distribuzione live cd Linux per architetture X86 basata su Knoppix che comprende un insieme di strumenti open source per l'analisi forense ed è la prima distribuzione che integra nativamente le funzioni AFF.

Le prove hanno avuto lo scopo di confrontare AFF con altri strumenti utilizzati in attività di consulenza in casi penali; si è rilevato che rispetto:

- a prodotti simili e a EnCase, AFF mentre esegue l'acquisizione del reperto calcola anche lo hash SHA1 e MD5 dell'immagine grezza consentendo un sostanzioso risparmio di tempo rispetto alle procedure a più passi;
- a prodotti simili, AFF consente di creare l'immagine dei dati grezzi compressa permettendo l'apertura e la lettura del file senza dover decomprimere in un secondo momento l'immagine;
- a prodotti simili, AFF consente una visualizzazione a elevata usabilità di informazioni dettagliate riguardanti i segment, meta informazioni, dati sullo hash dei singoli file;
- a prodotti simili e a EnCase, AFF produce file compressi più piccoli;
- a EnCase, AFF consente di superare il limite di creazione di immagine di due Gigabytes.

Gli strumenti di AFF che abbiamo sperimentato riguardano:

- aimage per la creazione di nuove immagini in formato AFF;
- aconvert per la conversione di immagini grezze in immagini AFF;
- acompare per confrontare un'immagine grezza con una in formato AFF;
- ainfo per visualizzare a video le informazioni dettagliate riguardanti l'immagine AFF;
- acat per creare un immagine grezza da un immagine AFF.

Attualmente le funzioni di lettura ed apertura delle immagini in formato AFF, la citata libreria AFFLIB, sono state implementate nella nuova release di Sleuthkit [12], che non è stata ancora rilasciata pubblicamente.

Intendiamo inserire AFF nel prodotto Autopsy [13] che da tempo utilizziamo in pratiche forensi; più precisamente l'integrazione può avvenire integrando il format AFF per costruire strumenti come un motore di ricerca tramite parole chiave più efficiente di quelli ora disponibile utilizzando le potenzialità di ricerca di Google desktop all'interno dell'applicazione. Pur non essendo questo un software open source, si possono scrivere interfacce da integrare all'interno di Autopsy per poter sfruttare le sue potenzialità per la ricerca di file e contenuti nei vari dispositivi che risparmino spazio e tempo.

Referenze

- [1] Maioli C., *Dar voce alle prove: elementi di Informatica forense*, in P. Pozzi (a cura), *La sicurezza preventiva dell'informazione e della comunicazione*, FrancoAngeli, 2004
- [2] <http://www.guidancesoftware.com>
- [3] http://www.forensicswiki.org/index.php?title=Forensic_file_formats
- [4] Panda B. e J. Giordano, D. Kalil, *Next-generation cyber forensics: introduction*, CACM 49, 2, February 2006
- [5] CDESF, *Standardizing digital evidence storage*, CACM 49, 2, February 2006
- [6] Turner P., *Unification of digital evidence from disparate sources*, 5th Annual Digital Forensic Workshop, New Orleans, 2005
- [7] Garfinkel S.L. et alii, *Advanced Forensic Format: an open extensible format for disk imaging*, IFIP WG 11.9 International Conference on Digital Forensics, Orlando, January 2006
- [8] http://en.wikipedia.org/wiki/BSD_License
- [9] Garfinkel S.L., *AFF: a new format for storing hard drive images*, CACM 49, 2, February 2006
- [10] <http://www.afflib.org>
- [11] <http://www.lnx4n6.be/>
- [12] <http://www.sleuthkit.org/>
- [13] www.sleuthkit.org/autopsy

Cesare Maioli (cesare.maioli@unibo.it) è professore ordinario di Informatica giuridica e Informatica forense all'Università di Bologna e membro del Cirsfid; Stefano Fratepietro (fratepietro@cirsfid.unibo.it) è sistemista di rete al Cirsfid dell'Università di Bologna.