



Utilizzo di Advanced Forensic Format nell'Informatica Forense

Stefano D. Fratepietro, Cesare Maioli

CIRSFID e Università di Bologna

Roma, 21 giugno 2006

Contenuti

- **Cenni di Informatica forense**
- Informatica forense e strumenti di analisi open source
- Advanced Forensic Format
- Caso pratico

La criminalità usa la tecnologia informatica che non ha confini



- Terrorismo
- Cracking
- Accesso abusivo
- Danneggiamento informatico
- Pedopornografia
- Discriminazione razziale
- Ingiuria e diffamazione
- Spamming
- Bilanci falsi
- Riciclaggio
- Phishing
- Truffe on line
- Estorsioni
- Violazione della privacy
- Violazioni al diritto d'Autore
- Frode informatica
- “Furto” di dati

Crescita della domanda di analisi



All'aumento del trattamento di dati con sistemi informatici consegue l'incremento della domanda di analisi dei dati digitali a fini di investigazione e di giustizia per

- reati informatici e telematici (es. L. 547/93)
- reati non informatici ma commessi con sistemi informatici
- reati di cui si rinvengono tracce o indizi nei sistemi informatici

Comune denominatore:
il dato digitalizzato come oggetto di indagine

Emergenza di una nuova disciplina

Accanto all'Investigazione scientifica classica su:

- oggetti vari (vestiti, auto, armi, ecc.)
- materiale biomedico
- impronte digitali
- sostanze di vario tipo (legno, metalli, liquidi, esplosivi, ecc.)
- fotografie
- suoni
- documenti vari
- documenti cartacei

anche in Italia si è iniziato a parlare sempre più diffusamente di **Informatica Forense**

Definizione e obiettivi dell'Informatica forense

Informatica forense è la disciplina che concerne le attività di individuazione, conservazione, protezione, estrazione, documentazione ed ogni altra forma di trattamento ed interpretazione del dato memorizzato su supporto informatico, al fine di essere valutato come prova nel processo

Informatica forense studia a fini probatori i processi, le tecniche e gli strumenti per l'esame metodologico dei sistemi informatici (memorie, hard disk, dischetti, nastri, cartaceo, altri), nonché l'analisi forense di ogni sistema informatico e telematico (computer, rete di computer, e ogni altro dispositivo per il trattamento di dati in formato digitale), l'esibizione della prova elettronica, l'esibizione del dato digitale, il recupero di dati e la loro esibizione, l'analisi ed esame del sistema informatico e telematico.



- data seizure (**sequestro di dati**)
- data duplicaton (**duplicazione di dati**)
- data preservation (**conservazione di dati**)
- data recovery (**recupero dati**)
- document searches (**ricerca di documenti**)
- media conversion (**conversione di formati**)
- expert witness services (**servizi di testimoni periti**)



Elemento comune: il dato digitalizzato come oggetto di indagine

Nel corso di numerose indagini sono emersi:

Limiti culturali

- scarsa cultura informatica tecnica di base
- scarsa cultura di diritto dell'informatica
- no dottrina
- no giurisprudenza

Limiti organizzativi ed operativi

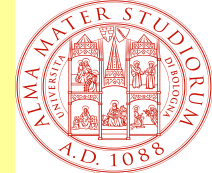
- know how investigativo
- know how difensivo
- poca formazione specifica
- pochissimi laboratori pubblici e privati di analisi dei dati a fini investigativi

Tendenze in atto - I

Attualmente nei processi è molto frequente che il Pubblico Ministero chieda che vengano considerate "prove":

- log di server inviati via fax dal provider
- stampe di home page
- stampe di e-mail (**nemmeno firmate digitalmente**)
- stampe di sessioni di chat
- stampe di log
- stampe di tabulati telefonici
- contenuti di supporti di memorizzazione utilizzati dagli accertatori prima di apporre i sigilli
- contenuti di supporti sequestrati ma non sigillati
- "perizie" d'ufficio predisposte da "consulenti" **privi di formazione specifica** nel settore della *digital evidence* e, in qualche caso, nemmeno laureati o laureati in materie non tecniche
- relazioni di servizio sui contenuti di un sito remoto predisposte da agenti e ufficiali di polizia giudiziaria privi di competenze specifiche
- identificazione di un soggetto solo tramite user id e intestazione della eventuale utenza telefonica impiegata per il collegamento in rete

Tendenze in atto - II



Va detto che, stranamente, questo atteggiamento di inquirenti e giudicanti si è manifestato in modo particolare per quanto riguarda l'informatica

Mentre, infatti, l'ammissione in un processo di altri tipi di prova (il test del DNA, per esempio) è stata soggetta a un lungo periodo di stretta "osservazione" e di analisi critica da parte di illustri scienziati, questo non è accaduto con la *digital evidence*

Che entra nei processi dalla porta principale come se fosse già elemento consolidato e non, come in effetti è, tema ignoto sul quale tutto è ancora da scrivere

Prova scientifica *nuova*

- strumenti tecnico scientifici a elevata specializzazione non oggetto di una condivisa e consolidata esperienza nell'uso giudiziario
- due aspetti che devono trovare un punto di sutura:
 - epistemologia scientifica e giudiziaria
 - congegni procedurali a trasferirla nella ricostruzione processuale di un fatto
- in quanto
 - « *le tradizionali categorie concettuali sedimentatesi con gli studi e gli enunciati giurisprudenziali sugli istituti della perizia e della consulenza tecnica sono in grado di dare apporti di assai scarso rilievo* »

Principi inderogabili nelle perizie a tema informatico

2. Le memorie di massa e ogni dispositivo di memorizzazione devono essere “congelati” al più presto possibile; cioè i reperti vanno raccolti nel tempo più prossimo all’accadere di un evento di interesse e senza che i contenuti dei dispositivi di memoria vengano alterati
4. Deve essere garantita la continuità della prova; cioè deve esistere una “catena di custodia” senza discontinuità per quel che riguarda la gestione del reperto, dal momento in cui viene sequestrato al momento in cui compare in giudizio
5. Tutte le procedure utilizzate durante l’esame dei reperti devono essere **controllabili e ripetibili**; cioè un esperto indipendente deve essere in grado, leggendo i documenti di ripetere tutte le operazioni che sono state eseguite durante le indagini

Informatica forense e reperto informatico

Tenendo presente la volatilità del dato informatico si esamina il reperto informatico

Le fasi del trattamento sono:

- Individuazione
- Acquisizione
- Analisi
- Valutazione

Bit stream image

- Eseguire un copia integrale, bit per bit, del disco su un altro dispositivo
- Calcolare l'hash del disco sorgente e del disco copia e confrontarli, meglio se con la firma digitale
- Creare almeno tre copie

Contenuti

- Cenni di Informatica forense
- **Informatica forense e strumenti di analisi open source**
- Advanced Forensic Format
- Caso pratico

Analisi di Informatica forense con EnCase

- Il prodotto leader è EnCase della Guidance Software
 - 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
 - è considerato in linea con gli standard internazionali per le analisi delle tracce informatiche
 - 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
 - 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)
 - seguito da un footer che contiene lo hash MD5 per l'intero bitstream

Analisi di Informatica forense con EnCase

- 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

Memorizzazione con EnCase

- EnCase 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
 - compressori diffusi come gzip o bzip2 non consentono l'accesso random all'interno di un file compresso
- EnCase consente di inserire meta informazioni sulle varie parti dei documenti sotto esame
 - la prassi corrente 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

Progetti di open EnCase

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

Motivazioni per un formato comune di storage - I

- 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
 - 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

Motivazioni per un formato comune di storage - II

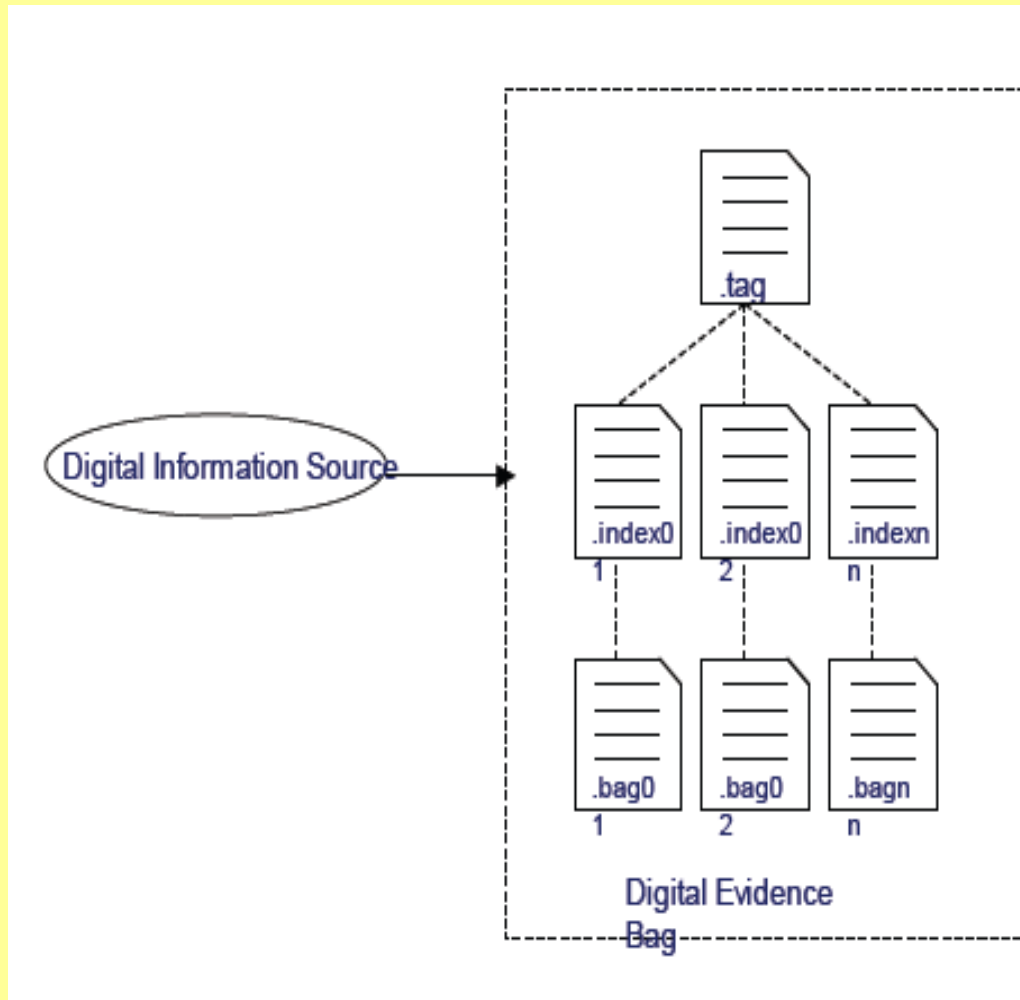
- Fattori ulteriori sono
 - le perdite di informazioni che si hanno convertendo dati grezzi rappresentati in formati diversi,
 - la dimensione dei file sequestrati di cui vengono eseguite copie settore per settore
 - la generale mancanza di meta dati
- Principali formati di file: ProDiscover, PyFlag, RAID, SDi32, SMART della famiglia open source e ILook, SafeBack proprietari

Common Digital Evidence Storage Format initiative



- L'iniziativa Common Digital Evidence Storage Format nasce per definire un formato open che risolva tali problemi basandosi sui format attuali, sulle esigenze dell'utenza e sugli standard giudiziari
- Necessità di una *evidence bag*
 - Garantire 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)
 - Consentire flessibilità per tener conto di più forme di reperto digitale (traffico in rete, dump di memorie, struttura logica dei file)
- con 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

Evidence Bag



Contenuti

- Cenni di Informatica forense
- Informatica forense e strumenti di analisi open source
- **Advanced Forensic Format**
- Caso pratico

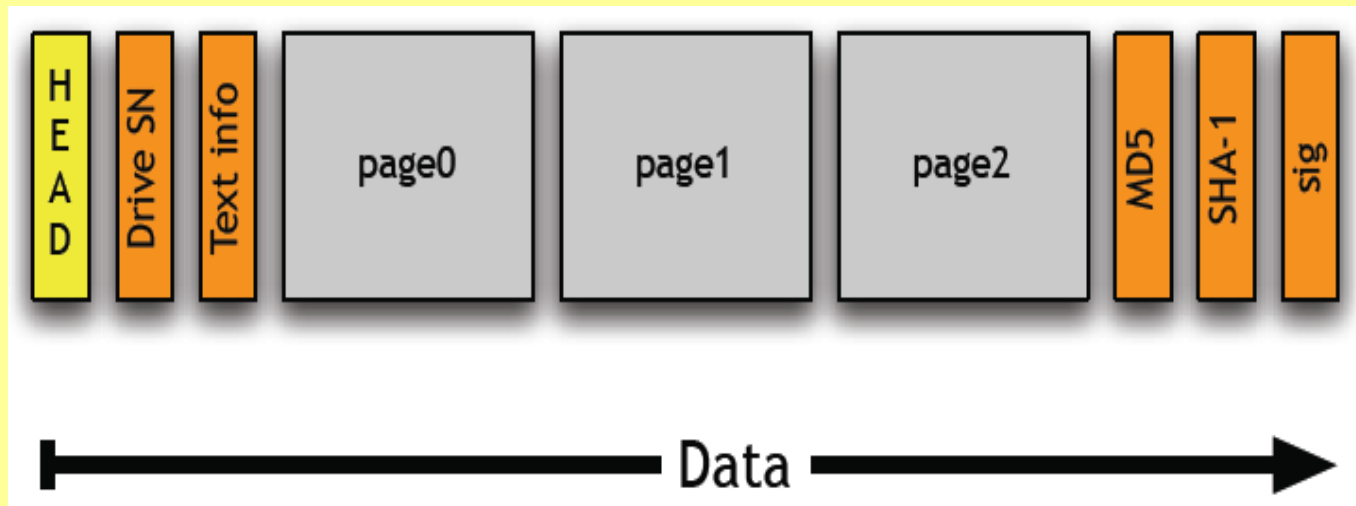
Architettura di Advanced Forensic Format

- AFF è un' implementazione open source ed estensibile distribuita sotto licenza BSD modificata di un formato che
 - analogamente a quello di EnCase memorizza l'immagine in maniera compressa e indirizzabile
 - 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 forward e backward 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 b-tree ritenendo che l' articolazione delle informazioni contenute in un b-tree fosse troppo complessa da spiegare, laddove se ne presentasse la necessità, nella fase dibattimentale di un processo; pertanto è stato adottato un approccio più semplice basato su una struttura detta AFF segment, ripetibile e di lunghezza variabile

AFF e AFFLIB

- 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

Architettura di Advanced Forensic Format



AFF verso digital Evidence Bag



- AFF è un formato che si avvicina al concetto di *evidence bag* permettendo di memorizzare al suo interno (in un contenitore “virtuale”) informazioni riguardanti i dati acquisiti proponendosi come una adeguata alternativa alle attuali best practice
- Siamo ancora lontani dalle complete funzionalità di *evidence bag* che deve garantire politiche di accesso ai file contenuti nel contenitore (già sviluppato in EnCase)

Contenuti

- Cenni di Informatica forense
- Informatica forense e strumenti di analisi open source
- Advanced Forensic Format
- **Caso pratico**

Utilizzo di AFF

Prove eseguite utilizzando:

- AFFlib compilato su Debian Gnu/Linux 3.1
- 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

Gli strumenti di AFF 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

Aimage: Caso di acquisizione - I

Abbiamo fatto alcune prove utilizzando reperti di varie dimensioni per poter confrontare le prestazioni di *aimage* e di *dd rescue*

L'acquisizione di un'immagine acquisita con *dd rescue* è terminata dopo 19 secondi, mentre l'immagine creata utilizzando *aimage* ha richiesto 31 secondi

Aimage: Caso di acquisizione - II

L' output generato da *dd rescue* è una normale immagine bit stream

L'output generato da *aimage* è un immagine compressa al 51% (66MB sui 128 MB), cifrata in SSL e con lo hash SHA1 e MD5 già calcolati e scritti all'interno delle meta informazioni del file leggibili con il tool *ainfo*



Aimage: Caso di acquisizione - III

```
Elapsed Time: 00:00:08          IMAGING          Wed May 31 16:24:16 2006
Source device: /dev/sda1        AFF Output: nuovoaff.aff
Model #:
                                Disk Size: 129 MB (1024 byte sectors)
                                Total sectors: 126,448

.<.MSWIN4.1.....!. .....)....NO NAME   FAT16   3.....{...x..v..V
U."..~..N.....|.E..8N$} ...~...:f..|f;..W.u.....V....s.3....}.F...f..F..V
.F....v.`.F..V.. ....^...H...F..N.a....(.r>8-t.`.....}.at=Nt... ;.r.....}{...}
.....@t.Ht.....)}.....^..f.....}.}.E..N...F..V.....r....p.RP.Sj.j.
[=====] ]

Currently reading sector:      32,768 (32768 sector chunks) (25.91% done)

      blank sectors:          0
      Done in:                00:00:23 (this drive)

      Bytes read:             33,554,432
      Bytes written:          18,464,189

Overall compression ratio:     44.97% (0% is none; 100% is perfect)

Free space on capture drive: 34,515 MB
                                WRITING  ==>
```

Aimage: Caso di acquisizione - IV

```
/home/steve/Documents/digital_evidence/nuovachiavetta.aff
[skipping data segments]

Segment          arg          data
=====          =====
badflag           0           512  BAD SECTOR..9...$. ..}.m.....
badsectors        2           8      = 0 (64-bit value)
afflib_version    0           6      1.6.26
aff_file_type     0           3      AFF
acquisition_commandline  0          31     ./aimage /dev/sda1 nuovoaff.aff
acquisition_device 0           9      /dev/sda1
sectorsize        1024         0
pagesize          16777216     0
devicesectors     2           8      = 126448 (64-bit value)
acquisition_macaddr 0           18     00:01:02:9B:CB:6F.
acquisition_dmesg 0          22128  [4294667.296000] Linux version 2
image_gid         0           16     .....D.JN...
acquisition_date  0           19     2006-05-31 16:24:08
md5               0           16     AC8F DF33 A82A 2208 D30E 4B2C 161D 2BBE
sha1              0           20     7820 374C 4658 F160 0784 EA34 5491 538A
                  DEC1 A12A
blanksectors      2           8      = 0 (64-bit value)
acquisition_seconds 32          0      = 00:00:32 (hh:mm:ss)
imagesize         2           8      = 129482752 (64-bit value)

Page segments:    8
Empty segments:   0
Total segments:   26
```

Aimage: Caso di acquisizione - IV

```

/home/steve/Documents/digital_evidence/nuovachiavetta.aff
[skipping data segments]

Segment          arg          data
=====          =====
badflag           0           512  BAD SECTOR..9...$. ..}.m.....
badsectors        2           8      = 0 (64-bit value)
afflib_version    0           6      1.6.26
aff_file_type     0           3      AFF
acquisition_commandline 0          31     ./aimage /dev/sda1 nuovoaff.aff
acquisition_device 0           9      /dev/sda1
sectorsize        1024        0
pagesize         16777216    0
devicesectors    2           8      = 126448 (64-bit value)
acquisition_macaddr 0          18     00:01:02:9B:CB:6F.
acquisition_dmesg 0          22128  [4294667.296000] Linux version 2
image_gid        0           16     .....D.JN...
acquisition_date  0           19     2006-05-31 16:24:08
md5              0           16     AC8F DF33 A82A 2208 D30E 4B2C 161D 2BBE
sha1             0           20     7820 374C 4658 F160 0784 EA34 5491 538A
                DEC1 A12A
blanksectors     2           8      = 0 (64-bit value)
acquisition_seconds 32          0      = 00:00:32 (hh:mm:ss)
imagesize        2           8      = 129482752 (64-bit value)

Page segments:   8
Empty segments:  0
Total segments: 26

```



Funzioni di lettura, apertura e ricerca in AFF

Le funzioni di lettura, apertura, ricerca e scrittura sono state implementate nell'ultima release dello Sleuthkit

Autopsy, usato per la lettura delle immagini in formato AFF, ha le stesse prestazioni di lettura e ricerca su normali immagini bit stream

Vantaggi di AFF - I



- Compressione dell'immagine creata con possibilità di lettura dei dati contenuti all'interno dell'immagine
- Calcolo hash SHA1 e MD5 durante l'acquisizione
- Cifratura dell'immagine in SSL
- Nessun limite di grandezza delle immagini
- Supportato dallo Sleuthkit
- Netto risparmio di tempo

- 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

Corso di Informatica Forense

Prof. Cesare Maioli - Avv. Antonio Gammarota



Obiettivi

- Il corso esamina gli aspetti giuridici e tecnologici attinenti alla prova digitale.
- Muovendo dalla computer forensics internazionale si analizzano le modalità di indagine informatica alla luce dell'ordinamento giuridico italiano: tecniche di indagine scientifica, indagine informatica, investigazione difensiva nel campo dei crimini informatici e dei crimini comuni **la cui prova sia costituita da dati digitali o veicolati da sistemi informatici**. Vengono altresì esaminate le indagini su sistemi di telefonia fissa e mobile, per **reati attinenti ai sistemi di pagamento elettronici e su strumenti che si basano sull'elaborazione di dati in forma digitale**.
- Allo scopo di confrontare le esperienze maturate in altri Paesi col contesto giuridico italiano si fornisce un quadro dei problemi tecnici tipicamente informatici in connessione con le problematiche giuridiche che sottendono a tali tipi di indagini, soprattutto per quanto riguarda la corretta applicazione del diritto penale e del diritto processuale penale.
- L'attenzione si sofferma **sull'analisi delle norme rilevanti per le tecniche di acquisizione, conservazione, analisi e produzione dei dati digitali rinvenuti nei computer e dei flussi telematici, per la loro utilizzabilità nell'ambito dei vari tipi di processi** (civile, penale, tributario, amministrativo, contabile) nonché in altri tipi di istruttoria e procedimento amministrativo sia della Pubblica Amministrazione che delle autorità indipendenti (Banca d'Italia, Consob, Privacy, Antitrust, Telecomunicazioni).

(sito del corso - www.stevelab.net/didattica)

Futuri sviluppi



- Progetto e sviluppo di nuovi tool da integrare in AFF per la collaborazione allo sviluppo di Open EnCase
- Collaborazione allo sviluppo di AFFlib
- Utilizzo di AFF per attività didattiche
- Utilizzo di AFF in attività peritali

Conclusioni



- Ruolo crescente dell'Informatica forense
- Necessità di metodi e strumenti condivisi
- Importanza dell'open source
- Qualità di AFF

Computer forensics is one of the largest growth profession of the 21st century

(Michael Erbschloe, in Foreword, Vacca, J. R., *Computer Forensics – Computer Crime Scene Investigation*, Hingham, 2002)