TECNICHE AVANZATE DI RECUPERO DEI DATI CANCELLATI: IL *CARVING* DELLE STRUTTURE DATI

di Paolo Reale e Maurizio La Porta

Il recupero dei dati cancellati dai supporti di memorizzazione di massa (dischi fissi, chiavette USB, memorie, etc.), frutto di un'azione volontaria per finalità di occultamento o per l'insorgere di problemi di altra natura (p. es. difettosità dei supporti), è una delle attività più richieste nell'ambito della *digital forensics*, e quasi sempre di grande interesse investigativo.

Senza entrare nel merito delle modalità tecniche di memorizzazione dei dati sui supporti di memoria, e dei principali metodi per il recupero dei dati cancellati[(1)], è utile comunque riepilogare che esistono diversi approcci possibili, con livelli di complessità crescenti:

- → le tecniche di "*undelete*", che si basano sui *metadati* ancora presenti nel *file system*;
- → **le tecniche di "*file carving*"' che consistono nel ricercare i *files* cancellati basandosi sul loro stesso contenuto, piuttosto che sui *metadati***. Ciò consente di superare molti limiti, tra i quali la conoscenza della stessa struttura del *file system* oppure se questa è stata proprio cancellata.

Va detto che la maggior parte dei *tools* d'informatica forense consente sia l'*undelete* che il *carving* dei *files*, ovvero:

- ➲ il recupero di *files* cancellati: possibile quando la *entry* del *file* (ad esempio nella tabella dei *files* del *file system* NTFS) viene invalidata, cioè resa disponibile, ma non ancora riutilizzata, ed inoltre i *clusters* del *file* non sono stati ancora riallocati;
- ➲ il riconoscimento di *files* all'interno di *clusters* non allocati: (parti di) *files* possono essere recuperati da porzioni di *cluster* contigue quando vengono riconosciuti determinati *pattern* caratteristici di specifici tipi di *files*.

Risulta chiaro, dal tipo di approccio, come gli strumenti indicati nel secondo punto possano essere molto più efficaci dei primi, in quanto consentono il recupero di dati cancellati che altrimenti non sarebbero più individuabili. Ma anche questa metodologia ha dei limiti: la maggior parte di questi *tools* effettua il tentativo di recupero ricercando *files* non frammentati, non compressi, e in cui l'inizio del *file* non è stato sovrascritto. Questo approccio viene anche definito come "*carving*" semplice (o *basic*).

Dove questi strumenti non riescono ad essere efficaci, ovvero quando la situazione è molto più complessa, si deve ricorrere a tecniche di *carving* "avanzato": si definisce così quando il tentativo di recupero avviene in presenza anche di *files* frammentati, in cui i frammenti possono essere persi, o non essere sequenziali, basandosi sulla struttura interna dei *files*.

Riassumendo: se non è possibile ricostruire i *files* cancellati perché le *entry* del *file system* sono state sovrascritte, e il *file* stesso è stato in parte sovrascritto per cui falliscono anche le tecniche di *carving* semplice, una possibilità che può essere perseguita è il tentativo di recuperare i soli *metadati* contenuti nelle strutture dati memorizzate all'interno dei *files* stessi. Alcune di queste strutture dati, contenenti *metadati* di interesse, hanno infatti un'"impronta" riconoscibile, ad esempio una struttura variabile, ma che risponde a determinate regole che danno luogo a *pattern* identificabili. Ricercando in tutti i *clusters* del disco tali *pattern* è possibile identificare i settori del disco candidati a contenere tali strutture dati e, quindi, tentare di leggere le informazioni ivi contenute.

Volendo utilizzare una metafora, per comprendere il tipo di approccio utilizzato nel *carving* delle strutture dati, si può dire che tale lavoro sia simile a quello di un archeologo che trova un frammento di papiro con alcuni caratteri scritti sopra: per poterli interpretare, viene ipotizzato che facciano parte di un determinato tipo di testo in una particolare lingua e, sotto questa ipotesi, si prova a leggerli per vedere se risultino intellegibili. Se quanto letto rispetta le regole grammaticali, sintattiche e di contesto, allora l'ipotesi fatta è corretta ed il frammento può essere effettivamente letto. Ripetendo questo procedimento *byte per byte* su tutto il supporto di memorizzazione di massa si possono riconoscere, ricostruire e quindi interpretare le strutture dati ivi contenute.

In questo contesto sarà analizzato un approccio di *carving* "avanzato", basato sul tentativo di ricostruire non più il *file* originale nella sua forma integrale, perché irrimediabilmente sovrascritto in molte delle sue porzioni, ma parte di esso, tramite le strutture dati che contengono i *metadati* del *file* stesso, in quanto ancora utili a caratterizzare il *file* ormai "perduto".

Il procedimento, da un punto di vista concettuale, è il seguente (a meno di ottimizzazioni implementative):

1. si posiziona il puntatore in lettura al primo *byte* del primo *cluster* del disco;
2. se una struttura con un certo *pattern* cominciasse dove è stato posizionato il puntatore, il valore che assume i(l) byte letto(i) potrebbe avere uno dei seguenti significati:
   - un *marker* di identificazione della struttura (se la struttura inizia con un *marker*);
   - un *offset* al quale trovare il primo *byte* stringa significativo;
   - la dimensione della struttura;
   - altri parametri che determinano la semantica dei *bytes* successivi;
3. interpretando i(l) *byte(s)* in esame come se fosse(ro) parte della struttura ipotizzata, si traggono conclusioni sulla struttura stessa e si continua a leggere i *bytes* successivi;
4. se procedendo in questo modo si trovano risultati consistenti quali:
   - le informazioni di tipo stringa della mia ipotetica struttura contengono solo caratteri ASCII o UNICODE che costituiscono stringhe "sensate" (tale valutazione non può che essere euristica);

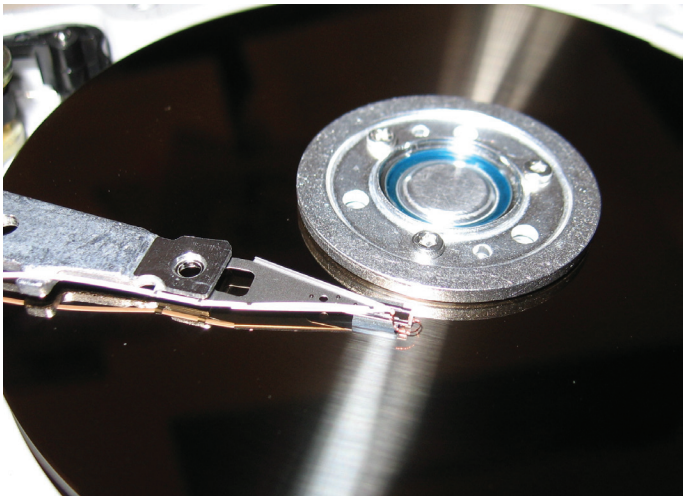
- i valori degli *offset* sono tra loro compatibili e compatibili con le dimensioni della struttura;
- si trovano tutti i campi attesi con valori che si collocano nei *range* previsti per la struttura in esame;
  allora si può concludere di avere potenzialmente trovato una struttura dati del tipo in esame, estrarne le informazioni contenute e spostare il puntatore di lettura alla fine della struttura appena letta. In alternativa, si sposta il puntatore di lettura al *byte* successivo a quello appena esaminato;

4. si torna al punto 2, sino a quando non è terminato il disco.

Il procedimento sopra descritto, per quanto valido (in astratto) per qualunque tipologia di struttura dati, è stato oggetto di una specifica implementazione in un caso giudiziario di grande rilevanza, per il recupero delle informazioni presenti nei metadati di Microsoft Office.

I *files* di Microsoft Office contengono una struttura, denominata "*SummaryInfo*" che contiene metadati di particolare interesse nell'ambito dell'analisi forense. Si tratta proprio della struttura dati che contiene tutte quelle informazioni visualizzabili nella finestra "Proprietà del documento" (in Word 2007 si accede a tale finestra attraverso il menù "Office -> Prepara -> Proprietà -> Proprietà Documento -> Proprietà avanzate"" mentre in precedenti versioni di Word era più facilmente accessibile…) quali: la data e ora di creazione del *file*, la data e ora dell'ultima modifica, l'autore dell'ultimo salvataggio, il numero di revisione (un valore progressivo che viene incrementato ad ogni salvataggio), il numero di caratteri contenuti nel documento, il tempo totale di modifica, e altri ancora facilmente rilevabili nella finestra citata.
La "*SummaryInfo*" è memorizzata all'interno dei *files* di Office, in alcuni casi anche in modo ridondato. Le caratteristiche di tale struttura dati, e dei *files* di Office, sono documentate in modo sufficientemente adeguato nella letteratura tecnica resa disponibile da Microsoft.

Unendo l'approccio algoritmico descritto precedentemente, e le informazioni puntuali sull'organizzazione dei dati della "*SummaryInfo*" è possibile automatizzare un *tool* per il *carving* avanzato di questa struttura dati: è possibile implementare una procedura *software* in grado di effettuare con successo l'analisi descritta sopra, producendo in *output* un *report* automatico con l'evidenza dei dettagli di tutte le strutture dati di questa tipologia individuate all'interno del supporto di memoria analizzato, sia esso un supporto fisico o una copia forense su *file*. Questo strumento di automazione del *carving*, applicato a Microsoft Word, consente il recupero di una serie di informazioni essenziali, irrecuperabili in un approccio di *carving* tradizionale: oltre alle informazioni sulla data ed autore di ciascuna versione del *file* (ormai in gran parte sovrascritto), si rivelano estremamente interessanti anche le informazioni sul numero di revisione, sul numero complessivo di caratteri del documento Word, e sul tempo di attività totale, che consentono di ricostruire la progressione delle attività di scrittura nel periodo in esame.



Oltre all'utilizzo per il recupero delle informazioni di Microsoft Word, la stessa procedura è utilizzabile anche con *files* di diversa tipologia ottenendo risultati altrettanto interessanti. In particolare, sono analizzabili le strutture dati "*ThumbCatalogItem*" dei *files* di tipo "thumbs.db". Si ricorda che questo particolare *file* viene prodotto dai sistemi operativi Microsoft, e rappresenta una memoria *cache* di miniature (anteprime) per i *files* di tipo immagine (ma anche di tipo AVI, PDF, etc.). L'utilità di questo *file*, che viene creato automaticamente dal sistema in modo "nascosto" (salvo configurazione utente diversa per disabilitare la funzione), è quella di evitare la necessità del "ricalcolo" delle immagini a dimensione ridotta, da visualizzare come "anteprime" del *file*, in modo da velocizzare le operazioni, soprattutto nei casi in cui sono presenti molte immagini all'interno della stessa cartella.
La struttura dati "*ThumbCatalogItem*" contiene il nome del *file* di cui viene creata l'anteprima e la corrispondente data di ultima modifica. Queste due informazioni possono sembrare estremamente limitate, ma in realtà consentono di recuperare delle indicazioni fondamentali, in particolare è possibile ricostruire la presenza nell'*hard disk* dell'indagato di determinati *files* video ormai cancellati e sovrascritti tempo addietro, associando anche i titoli degli stessi con le relative anteprime (ove possibile), ciò quando evidentemente i relativi *files* "thumbs.db" sono irrecuperabili nella loro interezza.

Risulta peraltro evidente che le informazioni recuperate con questa tecnica siano meno ricche di quelle che possono essere recuperate dalla ricostruzione integrale del *file* contenente tali strutture, in quanto si tratta di informazioni puntuali che richiedono una successiva fase di interpretazione. Per questo il *carving* delle strutture dati è da utilizzare quando si è nell'impossibilità di ricostruire i dati cancellati nella loro interezza. In situazioni di questo tipo, il metodo di *carving* avanzato consente di recuperare informazioni parziali, ma probabilmente uniche: informazioni che possono essere effettivamente determinanti nella comprensione degli eventi oggetto di indagine. ©

NOTE

1. "Cancellazione dei *files* e recupero in ambito forense" di Simone Tacconi, Sicurezza e Giustizia, n. 2, Anno 2012, pagg. 5-7. ◊