GESTIONE DELL'INCERTEZZA NELLA *DIGITAL FORENSICS*

di Marco Mattiucci

❶ *Premessa*

La *Digital Forensics* (DF) è una scienza forense nata a seguito della necessità di rispondere a consulenze tecniche di polizia giudiziaria inerenti sistemi digitali. Lungi quindi dall'essere partita da basi teoriche assestate, ha sfruttato la cultura esistente (prettamente nel settore informatico ed elettronico) per determinare metodi il più possibile certi ed inattaccabili, atti ad ottenere fonti di prova ammissibili in dibattimento. La stessa informatica teorica, la logica matematica che sottende l'informatica, la statistica alla base del funzionamento dei sistemi elettronici e la teoria delle probabilità, non sono mai stati coinvolti nell'iniziale creazione del DF e questo oggi inizia ad evidenziare delle possibili debolezze teoriche e strutturali della materia che potrebbero essere negativamente impiegate in dibattimento a scopo di invalidazione dei risultati.
In questo articolo andremo a vedere proprio una di tali debolezze strutturali e precisamente quella correlata alla teoria delle probabilità, ossia alla possibilità di legare ai risultati delle analisi un valore numerico, generalmente percentuale, di affidabilità.

❷ *Incertezza e Digital Forensics*

Come appena evidenziato il DF viene dal lavoro pratico di informatici "applicativi" e non da quello di informatici "teorici" per cui l'interpretazione dei risultati di un'analisi forense viene riportato sempre ad un qualcosa tipo "trovato"/"non trovato". Si tratta dell'approccio più semplificato ed intuitivo all'informatica, ossia quello che enfatizza al massimo la discretizzazione del digitale basato appunto sui valori binari 0 ed 1. Un minimo di approfondimento nel settore, soprattutto facendo riferimento all'elettronica che sottende il digitale, permette di comprendere senza problemi che il mondo digitale è solo apparentemente binario in quanto fatto di una miriade di situazioni "intermedie" che gioco forza vengono riportate ad uno dei due valori 0 ed 1. Quel giocoforza vuole dire una cosa semplice: incertezza!

Si può fare un esempio abbastanza comprensibile di cosa ciò comporti: un sistema A in condizioni limite (non normali di funzionamento), a partire dalle stesse condizioni iniziali (es. contenuto della memoria) e dagli stessi stimoli esterni (es. stessi *input* dall'utente) può avere reazioni ed *output* differenti.

Questa idea attacca la fondamentale certezza (falsa ma sicuramente rassicurante) che il *computer* (sistema digitale per eccellenza) sia una macchina deterministica. È sicuramente più corretto dire che un sistema digitale è generalmente deterministico, o meglio, è generalmente deterministico in condizioni normali di lavoro, mentre diviene non (facilmente) prevedibile in molti altri casi.
A questa idea di imprevedibilità bisogna sommare un'ulteriore fonte di incertezza che è quella relativa all'utente ed alle procedure di lavoro. L'esperienza dello scrivente in ambito di direzione di attività forense dei consulenti tecnici di DF, lo porta a considerare che raramente due operatori hanno le stesse procedure per svolgere lo stesso lavoro di analisi. Qualora poi venissero forzati a seguire le stesse procedure (formalmente definite), il modo di applicarle risulterebbe sostanzialmente diverso. Questo per un motivo fondamentale: le attività forensi del DF lasciano spazio alla creatività del tecnico e sostanzialmente ne hanno anche bisogno (molte soluzioni innovative sono spesso trovate grazie alle creatività e raramente nel rispetto delle procedure).

In definitiva le incertezze sul funzionamento interno dei sistemi analizzati, sulle procedure e sul materiale umano che le applica costituiscono una fonte di debolezza per i risultati del DF che in diversi articoli scientifici molti studiosi e specialisti nel settore hanno già iniziato a studiare[1,2,3,4].

❸ *Esempi di incertezza nel Digital Forensics*

Il primo esempio di incertezza intrinseca nel DF è quello da individuare nella *Live Digital Forensics*, in particolare di quella DF che si svolge direttamente sulla scena del crimine su sistemi attivi e spesso non disattivabili o sequestrabili/trasportabili.

Come si procede all'analisi di un sistema digitale "live" sul campo? L'operatore deve ovviamente interagire direttamente con esso. Questo nel migliore dei casi avviene attraverso dei *software* forensi specifici ma, dato che sulla scena possono presentarsi situazioni non preventivabili, la bontà dell'interazione è spesso nelle mani del singolo operatore scientifico forense. Fattori come la preparazione specifica nel singolo fatto che si presenta, la stanchezza, la capacità di resistenza allo stress, ecc. divengono elementi determinanti ed il risultato è che il determinismo dei risultati non è assicurabile affatto. A questo va aggiunto che tali attività molto raramente sono ripetibili sotto il profilo procedurale e quindi non possono essere discusse se non a livello meramente di studio documentale (su referti redatti dallo stesso che ha operato - per maggiori garanzie si consiglia la ripresa video in tali circostanze).
Come si manifesta quindi l'incertezza nella live DF sulla scena? Attraverso la possibile perdita di informazioni utili e/o la creazione di tracce inesistenti che vengono trasformate in fonti di prova o indizi.

Sempre nella *live* DF si possono poi rinvenire elementi di incertezza intrinseca dal punto di vista meramente tecnico[3], quando essa va infatti a toccare elementi volatili nei contenuti, come le memorie RAM, si generano dei curiosi paradossi in quanto il metodo di osservazione del contenuto di una memoria elettronica è a sua volta un *software* e questi deve sussistere nella memoria stessa per consentire all'osservatore di discernere conoscenza dalla macchina osservata. In pratica l'osservazione del contenuto ne implica la sua alterazione, ovviamente in misura diversa a seconda dell'operazione tecnica posta in atto e dell'abilità del tecnico, ma non si può prescindere da tale alterazione. Quello che è veramente interessante a questo punto è comprendere che l'alterazione di cui si parlava differisce da macchina a macchina, da tempo a tempo, da *tool* a *tool*. Questo significa ad esempio che anche impiegando lo stesso *tool* di osservazione forense sulla stessa macchina in istanti di tempo leggermente diversi si potrebbero ottenere risultati fondamentalmente differenti. In questo caso il livello di aleatorietà del fenomeno "osservazione" è veramente alto.

Un ulteriore ed ultimo esempio di incertezza nel DF lo si potrà facilmente comprendere in quello che si chiama "parallelizzazione forzata dell'impiego dei *tool*". La velocità nell'esecuzione delle attività scientifico forensi è un elemento ormai imprescindibile (spesso causa i processi mediatici cui i casi investigativi sono sottoposti). Il consulente quindi tende ad usare un solo strumento per svolgere le analisi e normalmente quello che gli dá più affidabilità o che per esperienza conosce e sa usare meglio e più speditamente. È purtroppo esperienza comune che *tools* forensi diversi diano risultati diversi nelle analisi, questo magari non completamente, ma i dettagli che differiscono possono essere tanti ed in dibattimento sono spesso i dettagli a fare da padroni. Ottenere quindi un risultato dall'impiego di un solo *tool* ed ottenere lo stesso risultato dall'impiego di due o più *tools* e magari più operatori (non interagenti tra loro = ripetizione indipendente dell'analisi) rende al risultato stesso un grado d'affidabilità enormemente più elevato.

❹ *Affidabilità e Probabilità nel Digital Forensics*

Come si può misurare l'affidabilità dei risultati nel DF? Nell'ambito delle scienze forensi in generale (es. genetica/biologia forense) questo problema è stato già affrontato mediante valori di probabilità. La situazione per il DF è particolarmente complessa data la "flessibilità" e l'ampiezza delle operazioni di analisi possibili, ma lo scrivente, assieme ad alcuni altri studiosi di DF a livello internazionale ha già iniziato la formalizzazione di un sistema descrittivo che permette la determinazione di valori percentuali di riferimento[6].

La necessità di avere un valore di probabilità (almeno di correttezza) delle analisi di DF è molto sentita nell'ambito delle certificazioni di qualità dei laboratori e degli specialisti forensi[5]. Per i certificatori è prassi comune chiedere all'analista di laboratorio: "qual è il grado di affidabilità dei suoi strumenti e dei suoi risultati?" e quando tale domanda viene posta ad uno specialista informatico l'imbarazzo è generalmente notevole.
Strumenti, procedure, situazione, reperti, personale, stato del laboratorio, ecc. sono alcuni tra gli elementi che concorrono alla formazione della probabilità di correttezza dell'analisi (caso semplificato e generale).
La probabilità in un singolo fenomeno aleatorio si può estrinsecare in due modi fondamentali:

1. soggettiva/teorica: in base a valutazioni di ditte/istituti esterni o relative a teorie assestate del mondo scientifico;
2. frequentistica: in base all'esperienza reale già svolta o a simulazioni numeriche su calcolatore.

Ciò posto, il primo passo è determinare la probabilità per ognuno dei fenomeni di base che determinano l'analisi nel DF (es. il processo di copia di una memoria di massa con uno specifico *tool* potrebbe essere visto come un passo di base, il processo di *hashing* anch'esso, e così via…) e poi combinarle mediante un calcolo numerico che tenga conto delle procedure impiegate per l'analisi (es. la parallelizzazione riduce le probabilità di errore, la serializzazione le aumenta)[6]. Quello che è certo è che per ottenere tali valori di probabilità è necessario formalizzare gran parte della attività di analisi forense e ciò ha un grande valore sia tecnico che legale, aumenta le garanzie della bontà dei risultati in esso stesso (migliore documentazione = meno errori) ed incrementa l'esaustività descrittiva dei referti tecnici.

❺ *Conclusioni*

I sistemi digitali, come visto, non sono sempre deterministici in quanto spesso i loro *outputs* sono dovuti a delle "fluttuazioni" inerenti lo stato del sistema, quello degli operatori umani, le possibilità di errore dei *tools* forensi impiegati, la farraginosità delle procedure, ecc.
Ad avviso dello scrivente è assolutamente necessario introdurre almeno una teoria assestata che consenta la valutazione della probabilità di correttezza delle attività di DF in maniera tale che **durante il dibattimento i legali possano valutare un semplice valore di percentuale che dia loro un'idea di affidabilità dei risultati**[5,6]. In caso contrario il legale è portato a pensare che i risultati del DF siano certi al 100% ma questo è mediamente falso, fatto testimoniato dalle diatribe tecniche che sempre più spesso si presentano in aula e che portano solo confusione a discapito spesso della verità. ©

NOTE

1. **E**. Casey (2002) – "*Error, Uncertainty, and Loss in Digital Evidence*" - *International Journal of Digital Evidence* - Summer 2002, Volume 1, Issue 2
2. **M**.Y.K. Kwan, K.P. Chow, F.Y.W. Law, P. K.Y. Lai (2008) – "*Reasoning about evidence using bayesian networks*" – IFIP Proceedings Vol. 285 "*Advances in Digital Forensics IV*" (pp. 275-289)
3. **A**. Savoldi, P. Gubian, I. Echizen (2010) – "*Uncertainty in Live Forensics*" - IFIP AICT 337 "*Advances in Digital Forensics VI*" (pp. 171-184)
4. **R**. Harriman (2011) – "*A case of forensic uncertainty*" – Digital Forensics Magazine Issue 9 December 2011 – http://www.digitalforensicsmagazine.com
5. **P**er lo studio dello scrivente (attualmente in corso) sulla formalizzazione dei sistemi di lavoro per il DF si può leggere un'antipazione nel "*Computer Forensics* & Indagini Digitali", Manuale tecnico-giuridico e casi pratici (Editore: Experta srl - http://www.experta.it - 2011) - S. Aterno, F. Cajani, G. Costabile, M. Mattiucci, G. Mazzaraco
6. **P**er lo studio dello scrivente (attualmente in corso) sull'incertezza nel *Digital Forensics* si può avere un'anticipazione in uno dei video didattici (liberamente disponibili) su http://www.marcomattiucci.it/myvideodf.php.◊