Osservatorio Italiano di Informatica Forense
Seminario formativo ONIF 2022
"Back To Amelia"
11 novembre 2022

Una “pistola fumante” nel processo civile telematico

Marco “*Darth Adobe*” Calamari
marco.calamari@ordineingegneripisa.it
Osservatorio Nazionale Informatica Forense - Ordine degli Ingegneri della provincia di Pisa

Il vostro anfitrione

https://www.linkedin.com/in/marcocalamari/

- **Marco Calamari, classe 1955, ingegnere nucleare, nell'ICT ha seguito un lungo cammino da umile sviluppatore ad architetto di applicazioni, ed è specializzato in gestione di software legacy. Opera come consulente in ambito informatico e di Computer Forensics dal 1990, ed ha maturato 15 anni di esperienza nella formazione in Olivetti ed Elea.**
- **Affiliazioni: ONIF, AIP, Opsi, PWS**
- **Appassionato di privacy e crittografia, ha contribuito ai progetti FOSS Freenet, Mixmaster, Mixminion, Tor e Globaleaks.**
- **Fondatore del *Progetto Winston Smith* e del convegno e-privacy, che quest'anno è giunto alla trentunesima edizione.**
- **Dal 2003 scrive su Punto Informatico, ZeusNews.it, Medium, Giano.news, Galileo ed altre testate la rubrica "Cassandra Crossing", che ha superato le 500 puntate (www.cassandracrossing.org).**
- **Membro della Commissione Informazione dell'Ordine di Pisa, ha tenuto numerosi corsi per il CNI e gli ordini provinciali della Toscana.**

Di cosa parleremo

- **Il Processo Civile Telematico ed i suoi documenti**
- **PDL – i linguaggi di descrizione della pagina**
- **Struttura di un documento PDF**
- **Dati, metadati, meta-metadati, meta-meta-metadati …**
- **Il “caso” reale**

Il Processo Civile Telematico ed i suoi documenti

I documenti del PCT

I documenti del fascicolo elettronico, scaricabile (con le opportune autorizzazioni) dal Processo Civile Telematico o dal sito Giustizia, sono a tutti gli effetti degli "originali", sia nel senso legale che nel senso informatico della parola.

I loro contenuti sono pienamente utilizzabili nel PCT dal punto di vista legale, e possono costituire prova o fonte di prova.

Ma quali sono i loro "contenuti"? Non sempre un "contenuto" è evidente o pienamente comprensibile, e non sempre l'autenticità del documento è scontata.

Ad esempio, un testamento olografo firmato e poi depositato nel PCT dopo essere stato sottoposto a scansione, può essere messo in discussione dal punto di vista dell'autenticità.

In questo caso si "esce" dal fascicolo elettronico, e si richiede di esibire l'originale, sottoposto poi a CTU grafologica.

I documenti del PCT

Come dicevamo, sempre più di frequente i documenti vengono prodotti direttamente in formato elettronico, iniziando dall'atto principale di un deposito, che deve obbligatoriamente essere depositato in pdf con testo selezionabile.

Ma se i documenti sono stati prodotti per via elettronica, e non sono stati successivamente "manipolati", il documento che viene depositato e poi scaricato dal PCT è un secondo originale.

Questo "originale elettronico" contiene sia i "normali" metadati, ma anche una "struttura interna", che a sua volta può contenere informazioni non visibili nel documento stampato.

Per proseguire, dobbiamo prima richiamare alcune nozioni su come funziona il processo di stampa da computer a PDF.

I linguaggi PDL

(Page Description Languages - Linguaggi di Descrizione della Pagina)

I linguaggi PDL

Le moderne stampanti non funzionano più con i martelletti o le palline ad impatto delle macchine da scrivere elettromeccaniche, e neppure come le stampanti ad aghi.

Ambedue queste tipologie erano in grado di produrre solo documenti testuali o poco più.

Le stampanti (o meglio, il loro driver di stampa) colloquiano col computer (o meglio con l'applicazione che sta stampando) con un vero e proprio linguaggio di programmazione, linguaggio che permette di descrivere e stampare testo e grafica in maniera più o meno sofisticata.

Dopo diversi linguaggi di stampa relativamente primitivi, nel 1982 Adobe System realizza il linguaggio PostScript e le prime famiglie di font vettoriali.

I linguaggi PDL

La tecnologià PostScript è talmente superiore alle preesistenti che si impone rapidamente come standard per le unità di fotocomposizione professionali, la più famosa delle quali fu la Linotype Linotronic 300.

Adobe® PostScript® 3™

I linguaggi PDL

Il PostScript si diffonde rapidamente dovunque.

Apple inserisce il PostScript nella sua innovativa stampante laser Apple LaserWriter; successivamente il linguaggio viene licenziato da Adobe sulle stampanti di fascia alta di quasi tutti I produttori. Jobs ne fa un fulcro della potenza dei Macintosh, e successivamente degli affascinanti e sfortunati NeXT.

Tuttavia la lentezza della produzione di software della Adobe, unita ai prezzi altissimi delle licenze, scatena una vera e propria "corsa" a clonare il Postscript, riscrivendolo da zero.

Della dozzina di aziende che partecipano alla gara, tre arrivano al traguardo. Viene anche realizzato un interprete con licenza libera, il Ghostscript.

Struttura di un documento PDF

Struttura di un documento PDF

Oggi il PostScript è dappertutto, perché ha figliato due parenti strettissimi, EPS (Encapsulated PostScript) e l'arcinoto PDF.

Il PDF (Portable Document Format) è un formato aperto, standard ISO 32000-2. E' al 99% PostScript, perché ne eredita tutti i comandi di stampa, la struttura del linguaggio e la gestione dei font.

Un file PDF contiene dati e metadati, ed ha un meccanismo per includere file binari come ad esempio le immagini TIFF o JPG, codificandole ROT64, ZIP o con altri metodi.

Un file PDF può consistere anche interamente di caratteri stampabili.
Normalmente non è così per occupare meno spazio.
L'inizio di ogni file è sempre in ASCII, come nella slide seguente.

L’inizio di un file PDF

%PDF-1.4

%\E2\E9\CD\D3

333 0 obj

<</Linearized 1/L 47721/O 783/E 5747/N 17/T 37053/H [576 728]>>

endobj

xref

333 15

0000000033 00000 n

0000001133 00000 n

0000000533 00000 n

trailer

<</Size 396/Prev 37041/XRefStm 924/Root 382 0 R/Info 43 0 R/ID[<0C215302E743484600B5E9E00D365D><B7881071E4004687F2A90A2DF56D26>]>>

startxref

0

%%EOF

355 0 obj

<</Length 237/C 311/Filter/FlateDecode/I 373/L 275/S 752 >> stream

PDF forensics:

Dati,
Metadati,
Meta-metadati,
Meta-meta-metadati ...

Ma cosa interessa ai forenser?

Il testo o le immagini che un PDF permette di visualizzare e stampare non sono solo bitmap, ma anche vettoriali.

Un file pdf può essere costituito anche solo da un'unica bitmap, come i file PDF generati da uno scanner.

Se generati da una normale applicazione, invece, i file PDF sono costituiti da singole entità grafiche, sovrapposte e posizionate opportunamente. Ogni singola stringa, talvolta ogni singolo carattere, è un'entità perfettamente riconoscibile anche nel file di stampa PDF.

In generale, ogni programma produce i file PDF in maniera tipica, con una struttura interna caratteristica come un'impronta digitale. Ad esempio, tutti i programmi che manipolano e sovrappongono bitmap, rettangoli od entità grafiche, le inseriscono singolarmente, come "oggetti" separati, nel file PDF.

Quindi, ad esempio, se una pagina viene "censurata" sovrapponendovi un rettangolo bianco, l'analisi del file PDF permette di ricostruirla..

PDF forensics

Ora abbiamo il necessario per raccontare un “caso”.

Riassumendo: il PDF scaricato dal PCT (ricordiamo che è un originale) può contenere più livelli di metadati. Oltre a quelli classici legati al filesystem ed all’inode (date, etc) esistono quelli propri del PostScript, che può contenere non solo commenti e metadati (i quali normalmente iniziano con uno o due caratteri “%”) ma anche oggetti di testo non stampabili, o commenti inseriti dall’applicazione generatrice del PDF.

Già in questi “internals” PDF possono essere contenute informazioni che il sistema operativo non fa vedere tra le proprietà del file.

Ma è meglio di così; ogni singola entità grafica inserita nel pdf, ad esempio una bitmap TIFF o JPG, mantiene nel suo header i propri metadati, completamente invisibili al filesystem ed all’intero processo di stampa, ma recuperabili in maniera forensically sound, decodificando la struttura del file PDF, ed identificando le entità di cui è composto.

Quindi è possibile che i dati della fotocamera di una foto in un documento PDF possano essere recuperati e, nel caso di entità gerarchiche, iterare il processo, facendo emergere veri tesori nascosti.

La “pistola fumante”

La pistola fumante: un “caso” reale

(Il procedimento non è concluso, quindi la descrizione sarà forzatamente generica.)

I fatti

Tra gli allegati di un procedimento a PCT viene prodotto un file pdf, contenente grafici prodotti da un applicativo per la gestione di veicoli.

Una pagina del documento prodotto, contenente un unico grafico, localizza un veicolo ed il suo conducente in certi tempi e luoghi.

Gli indizi

Il grafico appariva perfetto, ma ingrandendo l’immagine a colori, gli artefatti di compressione jpeg, quando sottoposti all’esame visuale di un esperto di grafica computerizzata, apparivano “asimmetrici”.

Una pagina conteneva inoltre la targa del veicolo, che il programma che aveva prodotto la stampa non era in grado di inserirvi.

La pistola fumante: un caso reale

L’analisi

Valendosi di una combinazione di comandi unix e di librerie di analisi in Python, si procedeva ad estrarre dal file pdf tutti i commenti, i metadati e gli oggetti testuali.

Tra i metadati si rinvenivano i dati del driver utilizzato, che permettevano di identificare il driver della stampante usata, e la data ed ora di stampa del pdf stesso.

Tali dati venivano confermati dalla identifi-cazione dello scanner impiegato (meglio descritta nella prossima slide), e permettevano di ricostruire quale marca e modello di fotocopiatrice multifunzione era stata usata per la manipolazione del documento.

La pistola fumante: un caso reale

La struttura interna del file PDF

Dopo aver verificato che l'applicazione generava il grafico come singola bitmap, si procedeva ad estrarre tutte le bitmap contenute nel file pdf.

La struttura del file era formata da un'unico grande file jpeg di base, con sovrapposte numerose piccole bitmap a sfondo trasparente, contenenti dati numerici rilevanti, ed anche il nome del conducente.

Il file jpeg di base era la scansione della stampa cartacea di un grafico del programma di gestione dei veicoli, file che conteneva nell'header JFIF i dati del driver di scanner utilizzato, che confermavano il tipo di fotocopiatrice multifunzione usata anche come scanner.

Il file jpeg di base presentava evidenti cancellature di parti significative del grafico, eseguite in digitale utilizzando un programma di fotoritocco ad oggetti tipo Photoshop.

In particolare, una delle immagini, posizionata sopra la pagina del grafico, conteneva l'informazione "impossibile", che era stata inserita per far risultare "a forza" nel grafico la targa del veicolo.

Conclusioni

Riassumendo, l'analisi informatica di un documento pdf proveniente direttamente dal PCT, della sua esatta struttura interna e degli oggetti in esso contenuti, ha permesso di provare la manipolazione di un atto rilevante, anzi decisivo, per la ricostruzione dei fatti oggetto di causa.

Aldilà del caso tecnico di analisi di un particolare tipo di documento PDF e delle tecniche utilizzate, il punto centrale di questa presentazione è evidenziare che, sia dal punto di vista legale che tecnico, qualunque contenuto del fascicolo del PCT è suscettibile di esame forense.

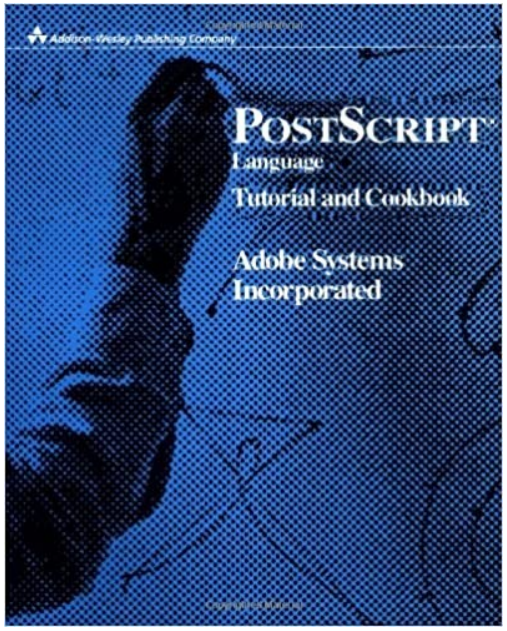
Dai soli documenti presenti nel fascicolo del PCT, il tecnico forense, sempre alla ricerca di elementi utili alla ricostruzione della verità, può accertare fatti rilevanti, quali la manipolazione di un documento, incongruenze tra documenti di uno stesso fascicolo ed altro.

Per approfondire il PostScript

Corso pratico di Postscript (1985, opera dell'autore, vero esempio di archeologia informatica!)
https://www.marcoc.it/corsops/indice.htm

PostScript ® LANGUAGE REFERENCE - third edition
https://www.adobe.com/jp/print/postscript/pdfs/PLRM.pdf

PDF Association resources index
https://www.pdfa.org/resource/pdf-specification-index/



Grazie per l'attenzione.
Ci sono domande?

+ Marco A. Calamari marco.calamari@ordineingegneripisa.it --+
| |
| DSS/DH: 8F3E 5BAE 906F B416 9242 1C10 8661 24A9 BFCE 822B |
| Cell: (+39) 347 8530279   Tel: (+39) 050 576031 |
| Skype-Twitter: calamarim |
| |
+ P.E.C.: marcoanselmoluca.calamari@ingpec.eu -------------+