SECUREBIZ 2024
Dal 10 all'11 Maggio a FIRENZE

Ottenere risposte immediate

DALL'ANALISI DEI FILE DI LOG CON SQLITE

ANDREA LAZZAROTTO

ANDREALAZZAROTTO.COM

A proposito di me

- Consulente informatico forense e sviluppatore
- Gli interessi includono la WhatsApp forensics e anti-forensics (manipolazione delle chat) e l'acquisizione e analisi dei metadati dei profili Instagram
- Autore di alcuni strumenti open source, inclusi **RecuperaBit** per la ricostruzione di NTFS e **Carbon14** per datare le pagine web (entrambi si trovano in CAINE)
- **A breve rilascerò Fuji,** un software per l'acquisizione forense dei computer con macOS

Agenda

I FILE DI LOG

Contenuto e approcci per analizzarli

IL CASO DEGLI SCREENSHOT COMPULSIVI

Un amministratore di sistema con vocazione da paparazzo

IL CASO DELLA POSTA VIOLATA

Un metodo peculiare di controllo del lavoratore

I file di log

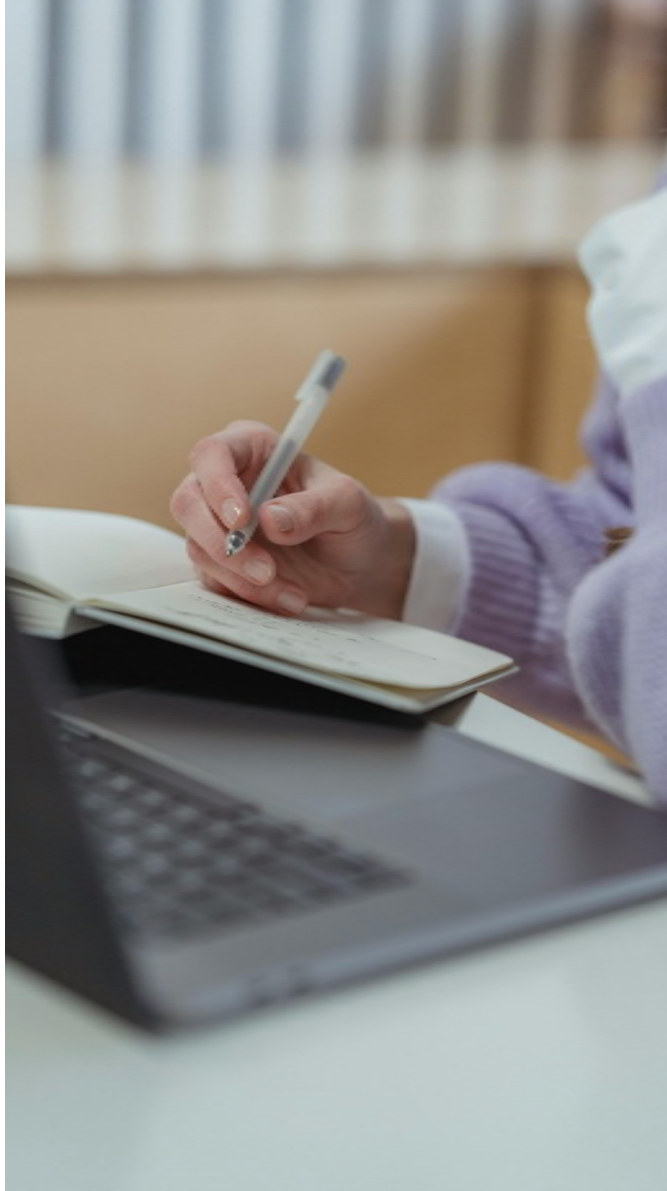
CONTENUTO E APPROCCI PER ANALIZZARLI

Struttura dei file di log

Concettualmente questi file si possono vedere come registri che memorizzano una sequenza di eventi accaduti nel sistema.

In genere sono presenti:

- Data e ora (*timestamp*)
- Nome utente e/o indirizzo IP
- Evento
- Altri dettagli (in funzione del tipo di evento)



Un file di log è un CSV
che non ci ha creduto abbastanza.

IN ESTREMA SINTESI

Trasformazione del contenuto

Si può partire da un file di log testuale:

```
2020-01-01 12:34,login,"device=PC01;user=root"
```

I fogli di calcolo hanno la funzione **“da testo a colonne”** che permette di suddividere i dati in base a un separatore.

Altri caratteri divisori possono essere gestiti con la ricerca e la sostituzione prima della suddivisione.

Infine si possono riordinare o cancellare le colonne.

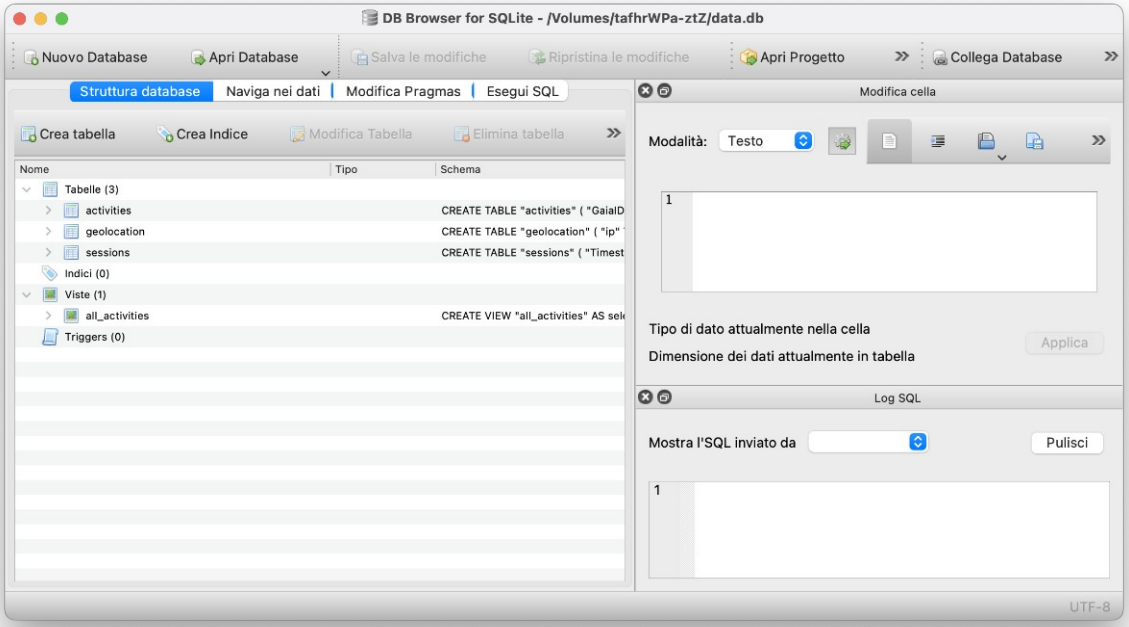
OPERAZIONI COMUNEMENTE APPLICATE

- Scegliere le colonne importanti
- Correlarle con altri dati
- Filtrare le righe
- Contare gli eventi
- Farlo in base a un dato comune
- Ordinare i risultati

OPERAZIONI COMUNEMENTE APPLICATE

- Scegliere le colonne importanti `SELECT`
- Correlarle con altri dati `JOIN`
- Filtrare le righe `WHERE`
- Contare gli eventi `COUNT`
- Farlo in base a un dato comune `GROUP BY`
- Ordinare i risultati `ORDER BY`

SQLITE



https://sqlitebrowser.org

SQLite è un motore di database, con dati memorizzati in un unico file.

L'ideale è usarlo con **DB Browser for SQLite,** molto comodo per importare file CSV e analizzarli in modo veloce e intuitivo.

Il caso degli screenshot compulsivi

UN AMMINISTRATORE DI SISTEMA CON VOCAZIONE DA PAPARAZZO

Scenario

Nell’arco di diversi mesi l’amministratore di sistema di un’azienda ha effettuato *screenshot* sui PC di diversi dirigenti e dipendenti, usando la piattaforma *Panda Systems Management*.

I log **non erano modificabili,** perciò chi è subentrato al ruolo ha potuto notare l’attività sospetta.

L’azienda ha richiesto l’analisi forense volta a determinare quanto estesa fosse stata l’attività e **chi fossero i lavoratori più colpiti.**

IMPORTAZIONE DEI DATI

Il file di log era già in formato CSV, tuttavia il formato dei dati non era adeguato:

```
2000-01-01 11:58:54 UTC,tizio,1.2.3.4,device : screenshot,
    "[deviceId:123, deviceUid:aa-bb-cc, deviceName:NB-TIZIO]"
```

Usando il foglio di calcolo, le righe relative agli screenshot sono state spezzate sul simbolo ":" e arricchite includendo colonne per **deviceId, deviceUid e deviceName:**

```
2000-01-01 11:58:54 UTC,tizio,1.2.3.4,device : screenshot,
    "[deviceId:123, deviceUid:aa-bb-cc, deviceName:NB-TIZIO]",
    123,aa-bb-cc,NB-TIZIO
```

Informazioni sugli indirizzi IP

Gli indirizzi IP coinvolti si possono ricavare con:

```
SELECT DISTINCT IndirizzoIP FROM UserActivity
```

Usando uno script esterno e apposite API, si può generare un **secondo file CSV** che contiene:

- Indirizzo IP
- Geolocalizzazione stimata
- ASN e nome del fornitore
- Contatti del provider, tipo di linea, ecc

HTTPS://IPINFO.IO

TIPO DI LINEA

Domestica

Spesso utilizza indirizzi IP che variano nel tempo. A volte è riconoscibile dall’indicazione del termine *Consumer* nei contatti specificati dal fornitore del servizio.

Aziendale

Generalmente utilizza un indirizzo IP statico, quindi invariabile per mesi o anni. In alcuni casi si riconosce dal termine *Business* o dal nome dell’azienda indicata come titolare.

CONTEGGIO PER TIPO DI OPERAZIONE

Tipologia	Conteggio	Prima operazione	Ultima operazione
account : login	349	2011-05-16 14:39:23 UTC	2013-05-28 12:57:46 UTC
account : logout	4	2012-01-11 06:54:19 UTC	2013-05-28 13:00:28 UTC
csv : download	2	2011-05-28 08:54:49 UTC	2011-06-18 09:58:16 UTC
device : screenshot	1305	2011-05-16 14:41:08 UTC	2013-05-25 11:58:54 UTC
user : create	1	2013-05-28 13:00:20 UTC	2013-05-28 13:00:20 UTC

Con una sola query possiamo generare una tabella di riepilogo:

```
SELECT Dettagli,
COUNT(*) as Conteggio,
MIN("Data/Ora") as "Prima",
MAX("Data/Ora") as "Ultima"
FROM UserActivity
GROUP BY Dettagli
```

CONTEGGIO PER CATEGORIA DI LINEA

Possiamo popolare una colonna “category” nella tabella degli IP:

```
UPDATE IPdata SET category = 'Aziendale'
WHERE abuse_address LIKE '%Business%'
```

```
SELECT category as "Tipologia",
COUNT(*) as "Conteggio",
MIN("Data/Ora") as "Prima operazione",
MAX("Data/Ora") as "Ultima operazione"
FROM UserActivity JOIN Ipdata
ON UserActivity.IndirizzoIP = Ipdata.ip
WHERE deviceId IS NOT NULL
GROUP BY category
ORDER BY Conteggio DESC
```

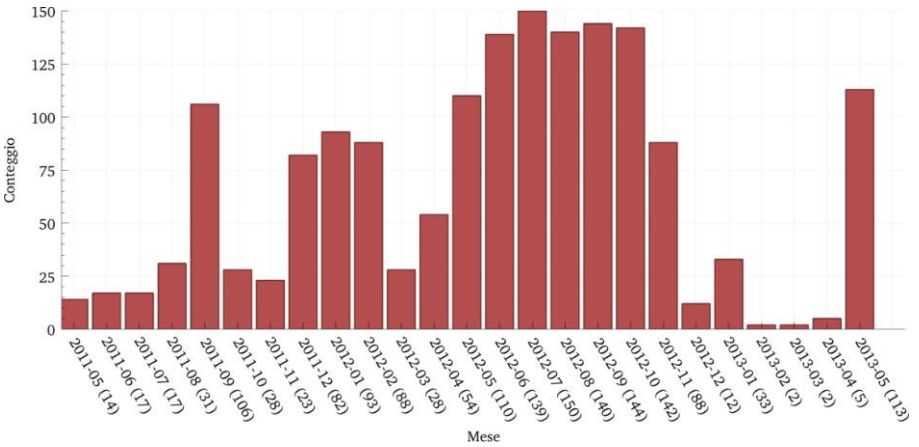
Tipologia	Conteggio	Prima operazione	Ultima operazione
Aziendale	1061	2011-05-16 14:41:08 UTC	2013-05-25 11:58:54 UTC
Privato	243	2011-05-18 10:01:12 UTC	2013-01-16 08:38:40 UTC
???	1	2012-02-11 15:47:46 UTC	2012-02-11 15:47:46 UTC

ANDAMENTO TEMPORALE

Possiamo contare le attività intercorse per singolo mese:

```
SELECT SUBSTR("Data/Ora", 1, 7) as "Mese1",
SUBSTR("Data/Ora", 1, 7) || ' (' || COUNT(*) || ')'
as "Mese",
COUNT(*) as "Conteggio"
FROM UserActivity
GROUP BY Mese1
ORDER BY Mese1 ASC
```

DB Browser for SQLite permette di generare un grafico, che si può stampare in PDF e/o modificare con un programma esterno.



Dispositivi coinvolti e conclusioni

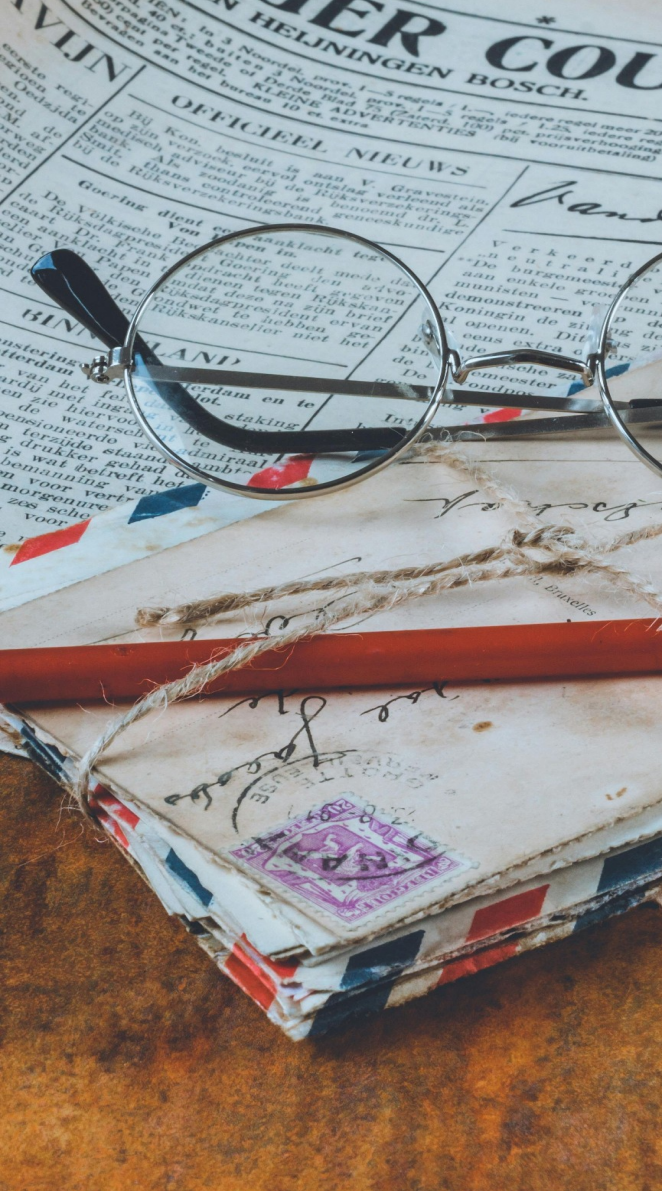
Con una query simile si può generare una tabella di tutti i PC coinvolti, individuando le vittime più colpite:

```
SELECT deviceId, deviceName, COUNT(*) as "Conteggio", MIN("Data/Ora")
as "Prima operazione", MAX("Data/Ora") as "Ultima operazione"
FROM UserActivity WHERE deviceId IS NOT NULL
GROUP BY deviceId, deviceName
ORDER BY Conteggio DESC
```

Il PC “preferito” è quello dell’amministratore delegato.

Il caso della posta violata

UN METODO PECULIARE DI CONTROLLO DEL LAVORATORE



Scenario

Una persona ha subito il licenziamento dal datore di lavoro, il quale ha contestato che il soggetto abbia inviato email in violazione del dovere di fedeltà all'azienda.

Il datore di lavoro ha mostrato al lavoratore una stampa della casella Gmail **personale** del dipendente.

È stato chiesto di **accertare eventuali accessi abusivi** all'account Google privato, specialmente se provenienti dalla rete aziendale.

Importazione

Tramite la funzione di Google Takeout sono stati richiesti i dati relativi a *Access Log Activity, Account Google* e *Le mie attività*.

L'archivio conteneva due tabelle:

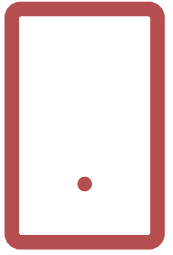
- **File CSV con le attività di utilizzo** dei servizi Google, molto dettagliato ma relativo solo agli ultimi 30 giorni
- **File HTML "SubscriberInfo"** con le sole indicazioni di *login* e *logout,* avente copertura temporale più estesa

I dati degli indirizzi IP sono stati arricchiti come già visto, predisponendo una tabella a parte.

TIPO DI DISPOSITIVO

PC

Identificato da stringhe quali *WINDOWS_OS, Windows NT* oppure *PC*.



Smartphone

Identificato da stringhe quali *IOS_OS, iPhone* oppure *MOBILE*.

Sconosciuto

Identificato dalla stringa *UNKNOWN* o comunque non coperto da altri casi.

Tipo di linea

Questo caso presenta una situazione poco agevole per quanto riguarda gli indirizzi IP coinvolti.

Infatti:

- Lo smartphone ha una SIM di Fastweb
- La linea di casa del lavoratore usa Fastweb
- La linea dell’azienda è fornita da Fastweb

Linee di casa e linee aziendali non si distinguono. Si rivela necessario cambiare approccio: **bisogna cercare anomalie comportamentali nelle attività sull’account Google.**

Giorni e fasce orarie

Con l’ausilio di uno script Python, si possono aggiungere le seguenti colonne a entrambe le tabelle con gli eventi:

- **TimestampITA:** valore nel fuso orario italiano
- **Data** e **Ora:** le due componenti staccate
- **Giorno:** nome del giorno in inglese
- **Ambito:** colonna con valore *Lavorativo* nei giorni e fasce orarie di lavoro, altrimenti *Personale*

Si può creare una vista per unire le colonne comuni tra le tabelle.

SFODERIAMO L'ARTIGLIERIA PESANTE

```
SELECT IPAddress as IP,
    MIN(TimestampITA) as 'Primo',
    MAX(TimeStampITA) as 'Ultimo',
    SUM(CASE WHEN Ambito = 'Lavorativo' THEN 1 ELSE 0 END) as Lavorativo,
    SUM(CASE WHEN Ambito = 'Personale' THEN 1 ELSE 0 END) as Personale,
    COUNT(*) as Totale,
    SUM(CASE WHEN Dispositivo = 'Smartphone' THEN 1 ELSE 0 END) as Smartphone,
    SUM(CASE WHEN Dispositivo = 'PC' THEN 1 ELSE 0 END) as PC,
    geolocation.company_name
FROM all_activities LEFT JOIN geolocation ON all_activities.IPAddress = geolocation.ip
GROUP BY IPAddress
HAVING Primo < '2014-04-01' and PC > 0
ORDER BY Ultimo desc
```

Indirizzi individuati e conclusioni

IP	Primo	Ultimo	Lav.	Pers.	Totale	Tel.	PC	*Company n.*
1.2.3.126	2013-07-15 10:56:14	2014-04-02 09:11:41	2960	576	3536	102	3363	Fastweb SpA
1.2.3.34	2014-03-30 14:29:18	2014-03-30 15:09:37	5	0	5	4	1	Fastweb SpA
aaaa:bbbb:cccc:dddd:eeee:ffff:gggg:6679	2014-03-30 14:29:18	2014-03-30 14:51:14	92	0	92	0	90	Fastweb SpA
aaaa:bbbb:cccc:dddd:eeee:ffff:gggg:94ab	2014-03-29 08:30:46	2014-03-29 09:10:51	403	0	403	0	403	Fastweb SpA
1.2.3.130	2014-03-29 08:44:58	2014-03-29 09:04:51	31	0	31	0	31	Fastweb SpA
aaaa:bbbb:cccc:dddd:eeee:ffff:gggg:b8b6	2014-02-29 15:01:37	2014-02-29 15:13:33	2	0	2	0	2	Fastweb SpA

Tra oltre 250 indirizzi IP presenti, **se ne identifica solo uno** utilizzato nell'arco di mesi e molto spesso in orario lavorativo. Dev'essere quello dell'azienda.

Conteggiando le attività di quell'IP suddivise per giorni, **si nota un picco di attività serali** prima che il dipendente fosse licenziato. Alcuni accessi sono occorsi anche nei giorni seguenti.

Altri casi interessanti

COMMESSA SOFTWARE

Tramite l’analisi dei commit di Git è stato stimato il numero di sviluppatori coinvolti e le giornate impiegate nel progetto.

ATTIVITÀ SU GOOGLE WORKSPACE

Le query con SQLite sono state usate per filtrare gli *audit log* prodotti da Google Workspace e controllare quanto svolto da due amministratori di sistema.

CONTATTI

Web
andrealazzarotto.com

GitHub
Lazza

Twitter
@thelazza

Mastodon
@lazza@mastodon.social