I log dei sistemi ICT:

un miraggio?

Andrea Pasquinucci

Indice:

- Log ed Audit
- Informazioni e log, qualche esempio
- L'eredità di syslog
- Formato, dimensioni, contenuti dei log
- Qualche spunto su cosa dovremmo fare

Log

- I sistemi informatici registrano informazioni sul proprio funzionamento in archivi chiamati **Log** (file, tabelle di database ecc.)
- Di norma le informazioni registrate sono per il **Debugging** del sistema (spesso interpretabili solo da chi lo ha sviluppato) e **non per il controllo delle attività** (**Audit**)
- Ogni sistema registra informazioni di tipo diverso in formato diverso

Audit

Per un efficace controllo dei sistemi, è necessario

- **Tracciare** le attività più rischiose
- **Correlare** attività svolte su sistemi diversi ma logicamente connesse
- **Archiviare** le informazioni in modo *sicuro* e tale da poter eseguire ricerche
- Creare **report** statistici/di soglia/puntuali delle attività

Soluzione ... ?

Archiviare i log ed estrarne le informazioni richieste

MA...

- I log NON ci sono!
- E quando ci sono, le informazioni NON ci sono!
- E quando ci sono, i log sono troppi!
- E chi li legge?

Log e Informazioni

Ci interessa sapere:

- **Chi** si è connesso ai sistemi (**come**, **dove** e **quando**)
- **Cosa** sta facendo, ovvero quali
  - Applicazioni
  - Dati
  - Risorse

  sta utilizzando, leggendo, creando, modificando ecc.

Sorgenti di Log

- Queste informazioni dovrebbero essere presenti nei **log** di
  - **Workstation**
  - **Apparecchi di comunicazione (rete)**
  - **Server**
- E in ognuno di questi potremmo avere log di
  - **Sistemi Operativi**
  - **Base dati**
  - **Applicazioni**

Esempio

Tracciare le attività su di una applicazione con front-end web:

- Utente si autentica al front-end web
- I dati passano ad un application server
- Il quale accede con una utenza applicativa al DB

Esempio - 2

- **Problema**: data una transazione sul DB, risalire all'utente che l'ha eseguita senza utilizzare i dati della transazione stessa

Esempio - 3

- **Problema**: data una transazione sul DB, risalire all'utente che l'ha eseguita senza utilizzare i dati della transazione stessa
- **Soluzione**: correlare i log dei sistemi coinvolti nella transazione

Esempio - 4

- **Problema**: data una transazione sul DB, risalire all'utente che l'ha eseguita senza utilizzare i dati della transazione stessa
- **Soluzione**: correlare i log dei sistemi coinvolti nella transazione
- **Problemi**:
  - I log non ci sono
  - I log non riportano le informazioni necessarie
  - I log non sono correlabili

Esempio - 5

- *Attenzione*: se manca anche solo un log/informazione nella catena, può risultare impossibile correlare gli eventi, o comunque farlo in modo inoppugnabile (ad esempio per usi forensi)

L'eredità di Syslog

- Per raccogliere i log di solito sono usati protocolli/formati (syslog, snmp ecc.) che **non garantiscono**:
  - L'origine
  - L'integrità
  - La protezione
  - la consegna dei messaggi.

L'eredità di Syslog - 2

- Storicamente i log sono:
  - In formato human-readable/free-format
  - Con contenuto non strutturato
  - Con contenuto di informazioni lasciato alla decisione/volontà/interesse del programmatore
- Pochi applicativi riportano nei log i *Chi/Quando/Dove/Come/Cosa* necessari per l'Audit (e sono applicativi specifici per l'Audit ovviamente)
- Di solito i log riportano solo messaggi di errore

Gestione dei log

Assumiamo comunque che i log siano generati, dobbiamo:

- **Raccogliere** i log in maniera efficiente e sicura
- **Archiviare** i log in modo efficace e tale da facilitare le ricerche
- **Correlare** i log di sorgenti diverse per l'estrazione delle informazioni di interesse.

Raccolta Log

- **Come** raccogliamo i log?
- **Quale** protocollo usiamo per garantire
  - Autenticità
  - Integrità
  - Completezza
  - Riservatezza

  dei log?
- **Quale** architettura usiamo per raccogliere così tanti dati?

Archiviazione Log

- **Dove** archiviamo da centinaia di GB a decine di TB al giorno di dati?
- **Come** possiamo fare ricerche su questa quantità enorme di dati?
- É necessario che i log siano strutturati in modo particolare in modo da poter permettere:
  - Ricerche veloci su quantità enormi di dati
  - Correlazioni tra log di sorgenti diverse

Correlazione Log

- Per estrarre e correlare le informazioni, è necessario che in ogni riga di log siano presenti come minimo
  - Chi
  - Quando
  - Dove
  - Come
  - Cosa

  Per ogni evento e nello stesso formato!

Correlazione Log - 2

- Oggi questo richiede spesso la **modifica** del codice degli applicativi per garantire che nei log siano riportate tutte le informazioni necessarie
- Manca uno **Standard** che indichi *come (cosa* e in che *formato)* il software debba produrre log utili all'Audit e non solo al Debugging

In pratica oggi ...

- Per generare i log possiamo attivare il Debugging di OS, DB, Applicazioni
- Facendo attenzione a cosa si traccia
- Facendo attenzione a quanti dati si generano
- Trasferendo i log con procedure ad-hoc e protocolli sicuri
- Formattando ed integrando i log prima dell'archiviazione
- Archiviandoli in DB specializzati.

Il problema log

- Sta diventando attuale per esigenze di
  - Sicurezza
  - Normative (i.e. Privacy)

  oltre ovviamente all'Audit, e pertanto
- Stanno nascendo le prime soluzioni

Riferimenti

A. Pasquinucci, *The Difficult Art of Managing Logs*, in stampa in *Computer Fraud & Security*

Copyright e Licenza

Grazie

Andrea Pasquinucci

pasquinucci-At-ucci.it www.ucci.it
a.pasquinucci-At-integra-group.it www.integra-group.it