FILE DI LOG: IMPORTANZA ED ANALISI

Yvette (vodka) Agostini vodka@s0ftpj.org

CNR - MILANO 4 novembre 2003

Copyright

Premessa

- Nel corso base abbiamo visto una gestione leggera dei logfiles, con un approccio orientato alla gestione della singola macchina
- in questa seconda parte amplieremo la prospettiva verso un ambiente più complesso, in cui non ha senso considerare solo i log della singola macchina, ma occorre un approccio globale

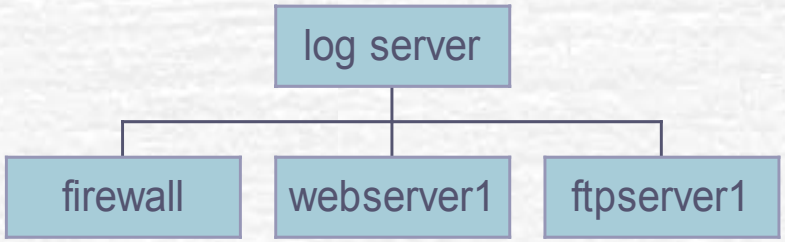
Di cosa parleremo

- RFC 3164: struttura dei messaggi syslog
- syslogd in dettaglio
- comunicare con syslogd (perl, C, logger)
- syslog-ng: un'ottima alternativa a syslogd
- centralizzazione dei log
- Strumenti di analisi e gestione dei log
- sicurezza: tips & tricks

Architetture per il logging centralizzato: RFC3164

- E' utile creare strutture centralizzate dove raccogliere i log

architettura logica struttura di log centralizzato



- il protocollo usato per la trasmissione dei messaggi syslog è UDP, sulla porta 514

- http://www.ietf.org/rfc/rfc3164.txt

RFC 3164: Struttura dei messaggi syslog (1)

PRI	HEADER		MSG	
<pri>	timestamp	hostname	Tag	message

un messaggio syslog e' costituito da 3 parti:

- PRI: è la priority, ottenuta da facility e severity
- HEADER: contiene il timestamp e l'hostname o indirizzo IP dell'host che genera il messaggio
- MSG: contiente il nome del processo che ha generato il messaggio e il testo del messaggio

RFC 3164: Struttura dei messaggi syslog (2)

PRI	HEADER		MSG	
<pri>	timestamp	hostname	Tag	message

PRIORITY = FACILITY * 8 + SEVERITY

- Poichè ci sono 24 facilites e 8 severities, il valore di PRI è un numero compreso tra 0 e 191

- la facility va da 0=kernel a 15=clock, ovvero at/cron; le rimanenti facilities sono di uso locale e possono essere assegnate in modo personalizzato)

- la severity va da 0=emergency a 7=debug-level

RFC 3164: Struttura dei messaggi syslog (3)

PRI	HEADER		MSG	
<pri>	timestamp	hostname	Tag	message

L'header è così composto:

Il **TIMESTAMP** contiene la data locale nel formato Mmm dd hh:mm:ss

L'**HOSTNAME** contiene il nome dell'host che ha generato il messaggio (senza il dominio), oppure il suo ip

RFC 3164: Struttura dei messaggi syslog (4)

PRI	HEADER		MSG	
<pri>	timestamp	hostname	Tag	message

MSG è così composto:

Il **TAG** contiene il nome del programma o processo che ha generato il messaggio

Il campo **CONTENT** contiene l'effettivo messaggio

FILE DI LOG: configurare syslogd (1)

Le principali opzioni di syslogd sono:

-a	Specifica sockets addizionali per ambienti chrooted
-d	Debug mode
-f	Specifica un differente file di configurazione
-h	Attiva l'host forwarding
-m	Intervallo per il --MARK-- in minuti (0 disattiva il mark)
-p	Socket da usare al posto di /dev/log
-r	Abilita la ricezione sulla porta 514 UDP

FILE DI LOG: configurare syslogd (2)

Il file di configurazione è syslog.conf, normalmente in /etc, la cui sintassi è:

FACILITY.SEVERITY<spazio/tab>AZIONE

=

Campo selettore <spazio/tab>Campo Azione

=

Cosa loggare<spazio/tab>Dove/come

FILE DI LOG: configurare syslogd (3 esercitazione)

Operatori del campo selettore e loro significato

*	tutte le facilities o tutte le severities
none	nessuna severity
,	multiple facilities e severities
;	multiple statement con stessa azione
=	esattamente una severity
!	negazione severity
\	separazione multiline

FILE DI LOG: configurare syslogd (4 esercitazione)

Le diverse azioni possibili per il campo azione sono:

Invia i log su normali file:	**/**
Invia i log su named Pipe – FIFO:	**|**
Invia i log su terminali virtuali e console:	**/dev/tty**
Invia i log su macchine remote:	**@**
Invia i log a una lista di utenti:	**,**
Invia i log a tutti gli utenti:	*****

Comunicare con syslogd (1): logger

E' possibile far comunicare i nostri shell scripts con syslogd, tramite:

- **logger**, un programma che consente di inviare a syslogd messaggi con priorità e tag definiti da noi secondo questa sintassi:

logger -p facility.severity -t tag message

- la priorità può essere indicata numericamente o come facility.priority

- di default logger utilizza la priorità user.notice

Comunicare con syslogd (2): Perl

E' possibile far comunicare i nostri programmi con syslogd, tramite:

- **Sys::Syslog**, un modulo Perl che consente di interagire con syslogd

- la sintassi e le funzioni del modulo si trovano qui:
http://www.perldoc.com/perl5.6/lib/Sys/Syslog.html

- una alternativa è il modulo **Unix::Syslog**, che differisce da Sys::Syslog in quanto logga solamente sul syslog locale, cioè non può essere usato per inviare log in rete direttamente. (consigliato per macchine stand-alone)

Comunicare con syslogd (3): C

E' possibile far comunicare i nostri programmi con syslogd, tramite:

- la libreria **syslog.h** ed utilizzando le funzioni che essa rende disponibili:
  - void openlog(char *ident, int option, int facility)
  - void syslog(int priority, char *format)
  - void closelog(void)

SYSLOG-NG: l'alternativa a syslogd

E' un progetto OpenSource, ormai molto maturo e sempre più utilizzato grazie alle sue caratteristiche di sicurezza e flessibilità.

L'ultima release è uscita pochi giorni fa (25 aprile 2003) ed è la:

syslog-ng 1.6.0rc3

Homepage: http://www.balabit.com/products/syslog_ng/

Freshmeat : http://freshmeat.net/projects/syslog-ng/

FAQ: http://www.campin.net/syslog-ng/faq.html

SYSLOG-NG: caratteristiche salienti

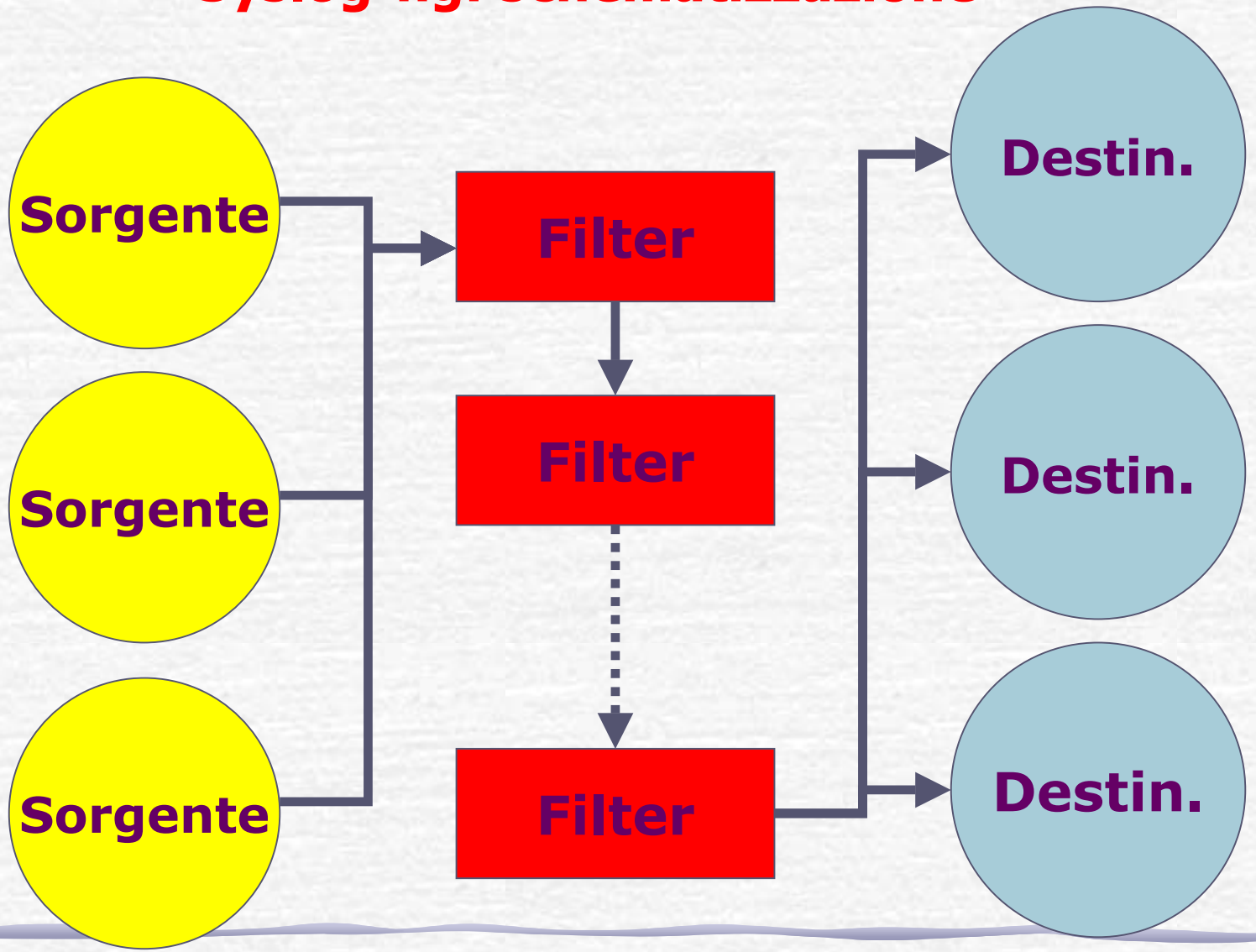
- Compatibilità con syslogd
- Possibilità di filtrare sul contenuto del messaggio
- File di configurazione chiaro e potente
- Forwarding log via UDP/TCP, con forwarding chain
- Creazione file di log basata su MACRO
- Formato dei log customizzabile
- Possibilità di risalire all'host generatore di messaggi

SYSLOG-NG: un approccio differente

- controllo più preciso sul filtering dei messaggi
- messaggi filtrati in base al **message path**, composto da:
  - una o più sorgenti di log
  - una o più regole di filtering
  - una o più destinazioni dei log
- ideato avendo in mente ambienti complessi, con segmenti di rete firewallati, e log centralizzato

quindi ci sono delle regole che legano una o più sorgenti, con uno o più filtri, in una o più destinazioni

Syslog-ng: schematizzazione



Courtesy of Valerio [Hypo] Verde

Syslog-ng: sorgenti, filtri, destinazioni (1)

Le **sorgenti** dichiarabili sono:

internal	per messaggi generati internamente da syslog-ng
unix-stream	apre un socket in modo SOCK_STREAM
unix-dgram	apre un socket in modo SOCK_DGRAM
file	apre un file (tipo /proc/kmsg)
pipe, fifo	per aprire una named pipe
udp	riceve messaggi via UDP
tcp	riceve messaggi via TCP
sun-stream	apre lo STREAM device su Solaris

Syslog-ng: sorgenti, filtri, destinazioni (2)

Le **destinazioni** dichiarabili sono:

file	scrive i messaggi su file
fifo, pipe	scrive i messaggi su una named pipe
unix-stream	invia i messaggi su un socket SOCK_STREAM
unix-dgram	invia i messaggi su un socket SOCK_DGRAM
udp	invia i messaggi via UDP
tcp	invia i messaggi via TCP
usertty	invia i messaggi ad un utente se collegato
program	forka, lancia un programma e manda il messaggio allo stdin

Syslog-ng: sorgenti, filtri, destinazioni (3)

I **filtri** (ogni filtro ha un identificatore univoco) usano questi operatori:

operatori logici	and, or, not
facility	controllo sulle facilities specificate
level	controllo sui levels specificati
program	controllo sul tag via regexp
host	controllo sull'hostname via regexp
match	controllo sul messaggio via regexp
filter	controllo su una diversa filter rule

Centralizzazione dei log

quando?

• In ambienti ad elevata complessita' ed elevate performances

Perchè?

• Per aumentare la sicurezza dei file di log, che vengono custoditi in un solo luogo
• Per semplificare l'analisi di attacchi (forensics) o di anomalie gravi

Centralizzazione dei log (2)

- In ambienti complessi e/o a rischio e' necessario non solo analizzare, ma anche **correlare** i log di differenti macchine e servizi.
  - Raccolta centralizzata (consolidamento)
  - Uniformazione di log con formati differenti
    - firewall
    - differenti sistemi operativi
    - apparati di rete
    - Intrusion Detection System
    - antivirus, ecc

Necessita' di strumenti avanzati e complessi

Aspetti interessanti della centralizzazione

- fonti diversificate (sistemi operativi differenti, tipologie di file di log differenti: apparati di rete, IDS, antivirus, ecc)

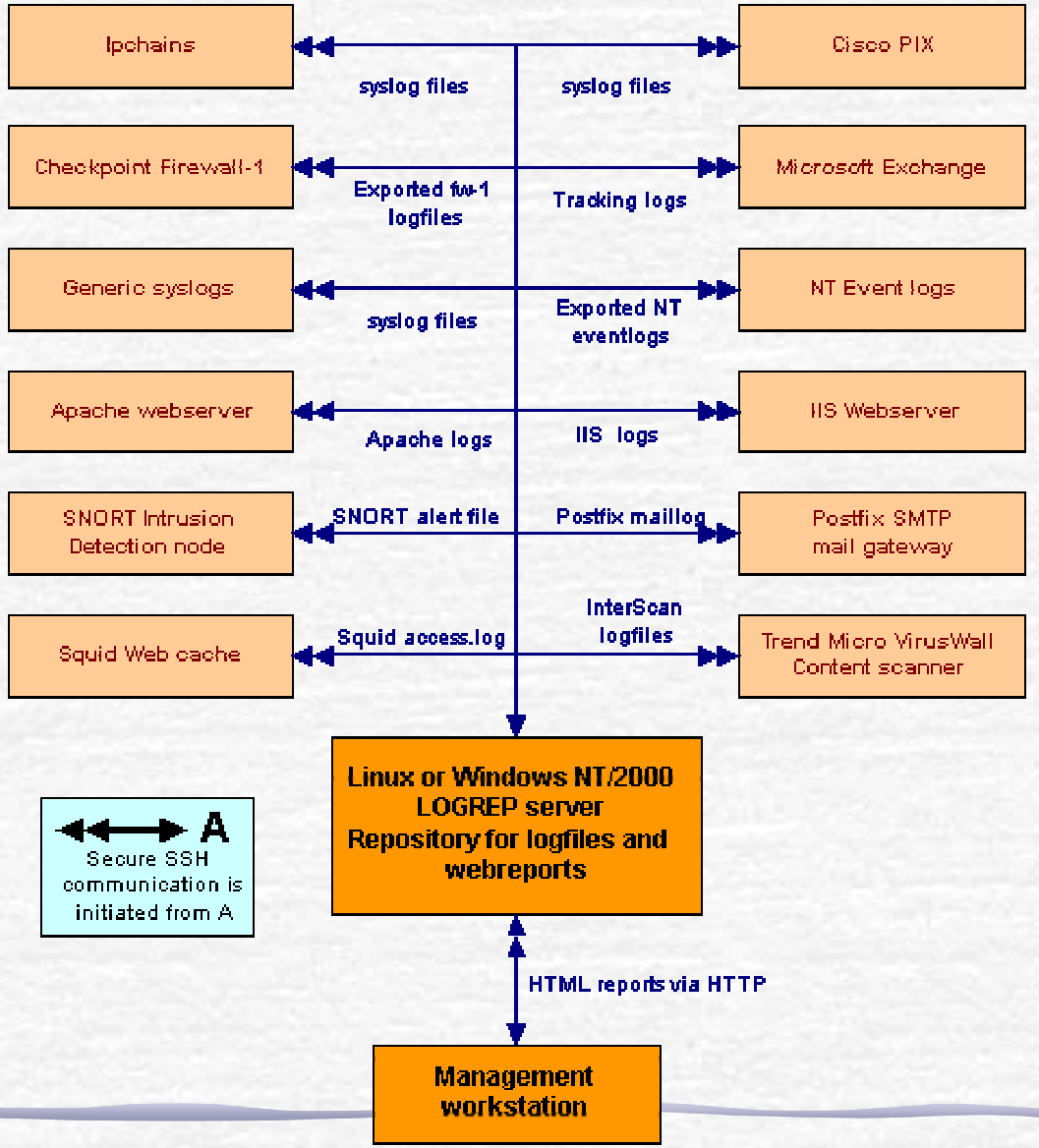
- segmenti di rete firewallati che devono essere attraversati dai flussi dei logfiles

- disomogeneità dei dati raccolti

- grandi quantità di informazioni da scremare per valorizzarle

Si parla perciò di consolidamento, gestione e analisi dei logfiles

Strumenti di consolidamento: LOGREP



Un interessante progetto LOGREP

http://logrep.sourceforge.net

- Multipiattaforma
- Trasferimento sicuro dei log
- Data mining e reporting

Strumenti di gestione: LOGROTATION

LOG ROTATION:
Ovvero, conservare in loco solo uno "slot" di file di log per ogni servizio che monitoriamo.

Come?

• a mano, scrivendo opportuni shell scripts e temporizzandone l'esecuzione tramite crontab

• utilizzando logrotate (ben noto a chi utilizza redhat) che da la possiblita' di comprimere, rimuovere, spedire per mail i file.

Strumenti di analisi periodica: LOGWATCH

LOGWATCH (http://www.logwatch.org):

- E' sviluppato in Perl
- Controlla a intervalli regolari, impostabili da file di configurazione oppure da linea di comando, i file di log ricercando stringhe particolari.
- E' abbastanza semplice da configurare e modificare per venire incontro alle proprie esigenze.

Warning: evitare le versioni<2.5, affette da un bug grave (race condition -> root in locale)

Una alternativa è Logcheck

Strumenti di analisi in realtime: SWATCH

Swatch (http://www.oit.ucsb.edu/~eta/swatch/)

- E' sviluppato in perl.
- Controlla in tempo reale i log alla ricerca dei "trigger" da noi scelti (esempio: parsa /var/log/maillog alla ricerca della stringa EXPN)
- Quando incontra un trigger swatch esegue un'azione
- E' abbastanza semplice da configurare e modificare.

Tutorial: http://www.enteract.com/~lspitz/swatch.html

Download: ftp://ftp.stanford.edu/general/security-tools/swatch/swatch-3.0.4.tar.gz

Una alternativa è Logsurfer

Sicurezza: tips & tricks (1)

- lasciare un file /etc/syslog.conf credibile

+

- utilizzarne invece uno "nascosto" che:
  - logga tutto il loggabile
  - lo invia a una macchina remota
  - ha un nome "innocente"

=

Eventuale attacker convinto di essere al sicuro

Utilizzare l'opzione -f di syslogd!!!!

Sicurezza: tips & tricks (2)

• Attenzione a trasferire in chiaro i file di log al logserver! Una delle cose che tutti gli attaccanti lasciano in ascolto è l'immancabile sniffer!!

Utilizzare connessioni cifrate!!!

• Attenzione ai Denial of Service sui file di log!! Sono possibili sia con syslogd che con syslog-ng, sia in locale che in remoto!!

Se non è un logserver allora non è necessario che il demone di logging sia in ascolto (syslogd sulla 514/udp e syslog-ng anche sulla tcp)

Sicurezza: tips & tricks (3a)

Attenzione ai Denial of Service!!

Sono possibili sia con syslogd che con syslog-ng, sia in locale che da remoto!!

Soluzioni:

Se non è un logserver allora non è necessario che il demone di logging sia in ascolto (syslogd sulla 514/udp e syslog-ng anche sulla tcp)

Ruotate i file, lasciate meno files possibili nella directory dove loggate, tenete sotto controllo la dimensione dei file di log (scriptino che vi avverte quando una certa soglia è superata)

Sicurezza: tips & tricks (3b)

Alle volte sono gli utenti legittimi del sistema che creano DoS via logging e senza volerlo (provate ad amministrare macchine di sviluppo e ve ne accorgerete ☺)

Per venire a capo della cosa basta creare un gruppo, e assegnare a questo gruppo /dev/log, tunando poi i permessi di lettura/scrittura opportunamente:

```
# groupadd nolog

# chgrp nolog /dev/log

# chmod g-rw,o+rw /dev/log

# ls -l /dev/log srw----rw- 1 root nolog 0 Apr 20 15:56 /dev/log
```

I log...

- Sono essenziali per condurre indagini relative a reati informatici (e non solo quelli)
- Contengono informazioni che possono violare la privacy degli utenti
- Contengono informazioni sui processi produttivi dell'azienda
- Sono facilmente corruttibili. Quale garanzia sulla loro veridicita'?
- Possono rivelarsi di intralcio invece che utili qualora non accuratamente raccolti e conservati

...la CIA...

C = Confidentiality = riservatezza → cifratura

I = Integrity = integrita' → hashing del file

A = Availability = disponibilita' → storage e backup

I file di log devono ottemperare a questi requisiti per poter essere utilizzati in ambito legale

..le leggi?!?

*"1. I responsabili dei **motori di telecomunicazione**, i portali Web, i provider, i gestori dei server e tutti gli operatori di telecomunicazione sono obbligati a conservare i **file di accesso al logo** per almeno dieci anni.*
2. In caso di mancata osservanza delle disposizioni di cui al comma 1, i soggetti di cui al medesimo comma incorrono nei reati di favoreggiamento e di concorso nella pedofilia e di sfruttamento dei minori. Salvo che il fatto non costituisca più grave reato, il responsabile è punito con la reclusione da uno a tre anni."

Disegno di legge S.57: Modifiche alla legge 3 agosto 1998, n. 269, e altre misure contro la pedofilia

I log e le leggi

- la direttiva 97/66/CE del 15 dicembre 1997, art. 4 - recepita dal decreto legislativo 13 maggio 1998, n. 171 – facendo riferimento ai fornitori di telecomunicazioni richiede che i dati relativi al traffico vengano resi anonimi o cancellati al termine della chiamata, fatte salve le finalità di fatturazione

Come dire che i log non vanno conservati

[i provider aderenti all'AIIP garantiscono la conservazione del log files per almeno 5 anni, anche se la legge non prevede questo obbligo]

- la legge 675/96 richiede, nell'ambito delle misure minime di sicurezza, che sia garantita la possibilità di identificare chi ha avuto accesso a dati sensibili o riservati.

Come dire che i log devono essere conservati

Testo Unico sulla Privacy

- Decreto legge del 27 giugno 2003 (pubblicato sulla Gazzetta Ufficiale del 29 luglio 2003)
- entrera' in vigore il 1 gennaio 2004
- prevede **l'obbligo di conservazione dei log** a fini investigativi, per favorire l'attivita' di indagine in caso di fatti rilevanti penalmente
- **non vengono indicate le modalità operative** di "gestione" delle informazioni, **ne' quali dati dovranno essere conservati** (la questione viene rimandata al Ministro della Giustizia, con il parere del Ministro dell'Interno, del Ministro delle comunicazione e parere del Garante della Privacy)

Risorse per approfondire (1)

Man page di: Syslogd, syslog.conf, logger

BSD Syslog Protocol (RFC 3164):
http://www.ietf.org/rfc/rfc3164.txt

Tools di vario genere (anche per altri OS):

http://online.securityfocus.com/cgi-bin/sfonline/tools.pl?platid=-1&cat=2&offset=0

Active Security Monitoring and Containment with Cross Technology Correlation: The Next Generation in Computer Security Technology:
http://online.securityfocus.com/guest/10414

Log analysis resource by Tina Byrd:

http://www.counterpane.com/log-analysis.html

Log analysis mailing list:

http://lists.shmoo.com/mailman/listinfo/loganalysis

Risorse per approfondire (2)

http://www.netjus.org

http://www.altalex.com

http://www.interlex.it

http://www.ictlaw.net

Ringraziamenti

Parte del materiale di queste slides è opera di Valerio [Hypo] Verde

Fabio (naif) Pietrosanti

http://fabio.pietrosanti.it

S0ftpj.org

http://www.s0ftpj.org

Lucrezia (luz) Gatti, cucciola smanettona

FILE DI LOG: IMPORTANZA ED ANALISI (advanced)

Contatti:

yvette.agostini@ieo-research.it vodka@s0ftpj.org