

Técnicas de Sumarização e Priorização de Logs

Klaus Steding-Jessen
jessen@cert.br

Centro de Estudos, Resposta e Tratamento
de Incidentes de Segurança no Brasil – CERT.br

<http://www.cert.br/>

Comitê Gestor da Internet no Brasil

<http://www.cgi.br/>

- Motivação e descrição do problema
- Algumas técnicas de análise de log
 - busca por padrões
 - sumarização
 - AI (“*Artificial Ignorance*”)
- descrição de algumas ferramentas
- protótipo de uma ferramenta de priorização

Motivação

Motivação

- administradores em geral não tem tempo para análise de logs
 - número crescente de sistemas
 - varreduras, worms/bots e ataques gerando ruído
- tendência a examinar apenas um subconjunto (ou ignorar...)
- Necessidade de mecanismos mais automatizados para priorizar os logs mais importantes

Motivação (cont)

- geralmente apenas as fontes mais comuns são analisadas:
 - `/var/log/messages`, logs de firewall, etc
- e várias outras desconsideradas:
 - `authlog`, `daemon`, `lpd-errs`, `failedlogin`, `xferlog`, `sudo`, `cron/log`
 - `maillog`, `spamd`
 - `access_log`, `error_log`, `ssl_request_log`
 - `named.log`
 - `error`, `mischief`, `vette`
 - ...

Algumas Técnicas de Análise de Log

Busca por Padrões

- padrões conhecidos
 - procura
 - descarte
- muito utilizado para alertas em “tempo real”
 - perdendo um pouco a utilidade
- risco: padrões nunca vistos
 - considerar o caso de “*else*”
- exemplos: `swatch`, `logsurfer`, etc

Sumarização

- produzido por um programa especializado para um determinado tipo de log
 - web (analog)
 - mail (pflogsumm)
 - etc
- pode mostrar tentativas maliciosas
- se for muito verboso, tende a ser ignorado

AI (Artificial Ignorance)

- discutido por Marcus Ranum, em 1997^{*}
- redução de variação não desejada (datas, PIDs, etc)
- remoção de linhas não interessantes
- linhas iguais são colapsadas, e a frequência é exibida

http://www.ranum.com/security/computer_security/papers/ai/

AI (cont)

Exemplo:

```
cd /var/log ; cat * | \  
sed -e 's/^.*demo//' -e 's/\[[0-9]*\]//' | \  
sort | uniq -c | \  
sort -r -n > /tmp/xx
```

```
297 cron: (root) CMD (/usr/bin/at)
```

```
167 sendmail: alias database /etc/aliases.db out of da
```

```
120 ftpd: PORT
```

```
61 lpd: restarted
```

```
48 kernel: wdpi0: transfer size=2048 intr cmd DRQ
```

```
[...]
```

Ferramentas

syslog-summary

- escrito em Python
- útil e simples
- específico para logs de syslog
- conceito de estado
 - sem suporte para rotação

syslog-summary (cont)

```
$ syslog-summary -s state -i ignore /var/log/daemon
```

```
Summarizing /var/log/daemon
```

```
0 Lines skipped (already processed)
```

```
1 Patterns to ignore
```

```
428 Ignored lines
```

```
2 hostname newsyslog: logfile turned over
```

```
1 hostname ntpd: peer 192.168.2.3 now invalid
```

```
2 hostname ntpd: peer 192.168.2.3 now valid
```

```
1 hostname ntpd: ntp engine ready
```

```
1 hostname savecore: no core dump
```

```
$ cat ignore
```

```
ntpd: adjusting local clock by
```

swatch

- monitoração em “tempo real”
- regex de interesse e de descarte
 - risco de desprezar logs, dependendo das regras
- ações: escrever a linha, beep, execução de comando (argumentos vindos log!), mail
- uma instância por arquivo

- conjunto de arquivos padrão que são monitorados
- **regras padrão:** `logcheck.hacking`,
`logcheck.violations`,
`logcheck.violations.ignore`, ...
- **risco de desprezar logs**

logsurfer e logsurfer+

- monitoração em “tempo real”
- flexível
- conceito de contextos
 - possibilidade de DoS
- difícil de configurar

Protótipo de uma Ferramenta

Requerimento de uma Ferramenta

- AI e sumarização
- uma única instância possa lidar com múltiplos arquivos
- configurável para lidar com vários formatos de log
- comportamento default: mostrar entradas desconhecidas
- mantenha estado e tenha suporte rotação de arquivos

Exemplo - arq. de configuração

```
##
## logsumm.conf -- logsum configuration file.
##

# input directives
input named_files = /var/named/log/named.log

# reduction directives /from/, /to/
reduction named_reduction =
/^d\d-\w\w\w-\d\d\d\d\s+\d\d:\d\d:\d\d\.\d+\s+/, //, \
/#\d+: view/, /#SRC: view/

# ignore directives
ignore named_ignore =    /^client 127\.0\.0\.1/, \
                        /^createfetch/, \
                        /(?i)^client .*: query: .*domain.name/
```

Exemplo - arq. de configuração (cont)

```
# option directives
options null =

# action directives
action null =

# rule directives
rule named = named_files, named_reduction, named_ignore, null, null

### logsumm.conf ends here.
```

Exemplo



```
34 client yyy.yyy.236.124#SRC: query: www.bb.com.br IN A +
14 client xxx.xxx.163.182#SRC: query: www.cartounibanco.com.br IN A +
14 client xxx.xxx.163.182#SRC: query: ibpf.unibanco.com.br IN A +
14 client xxx.xxx.163.182#SRC: query: www.banco1.net IN A +
14 client xxx.xxx.163.182#SRC: query: empresarial.unibanco.com.br IN A +
14 client xxx.xxx.163.182#SRC: query: www.unibanco.com.br IN A +
14 client xxx.xxx.163.182#SRC: query: unibanco.com IN NS +
14 client xxx.xxx.163.182#SRC: query: www.investshop.com.br IN A +
14 client xxx.xxx.163.182#SRC: query: www.unibanco.com IN A +
14 client xxx.xxx.163.182#SRC: query: unibanco.com.br IN NS +
 4 client zzz.zzz.19.155#SRC: query: images.americanas.com.br IN A +
 3 client zzz.zzz.19.152#SRC: query: www1.la.dell.com IN A +
 2 client zzz.zzz.19.155#SRC: query: energyzard.battleon.com IN A +
 2 client zzz.zzz.19.155#SRC: query: media.fastclick.net IN A +
 2 client zzz.zzz.19.152#SRC: query: www.fiesp.com.br IN A +
```

Conclusões

Conclusões

- alguns administradores se limitam a poucos logs (`/var/log/messages`)
- monitoração “em tempo real” não escala
 - a tendência é ignorar logs
- ao casar logs apenas com “match”, fica-se limitado a entradas conhecidas
- é fundamental usar ferramentas para automatizar parte do processo e incorporar a análise ao dia-a-dia
- lembrar dos possíveis riscos introduzidos

Links Relacionados

- **syslog-summary**

<http://packages.debian.org/stable/admin/syslog-summary.html>

- **swatch**

<http://swatch.sourceforge.net/>

- **logsurfer**

<http://www.cert.dfn.de/eng/logsurf/>

- **logsurfer+**

<http://www.crypt.gen.nz/logsurfer/>

- **logsentry**

<http://sourceforge.net/projects/sentrytools>

Links Relacionados (cont)

- Security Related Tools

<http://www.cert.br/tools/>

- Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil – CERT.br

<http://www.cert.br/>