Faculdade Senac Goiás

Curso de Gestão de Tecnologia da Informação

Laboratório de redes

Tratamento de arquivo log

Professor Orientador: Fernando

Acadêmicos: Henrique Sousa e Silva

Tarcisio Lopes

Tulio Barros

O que são arquivos de log?

Em computação, *log* de dados é uma expressão utilizada para descrever o processo de registro de eventos relevantes num sistema computacional. Esse registro pode ser utilizado para restabelecer o estado original de um sistema ou para que um administrador conheça o seu comportamento no passado. Um arquivo de log pode ser utilizado para auditoria e diagnóstico de problemas em sistemas computacionais.

Ademais, os *logs* possuem grande importância para o Direito da Tecnologia da Informação. A possibilidade de identificar a autoria de ações no ambiente virtual, permitindo a responsabilização dos autores, só é possível através da análise de *logs*. Os logs também podem ser entendidos como provas digitais.

O formato de log do HTTP

127.0.0.1 é o endereço IP do cliente (host remoto) que fez a solicitação ao servidor.

user-identifier é a identidade do RFC 1413 do cliente.

O frank é o userid da pessoa que solicita o documento.

[10 / Oct / 2000: 13: 55: 36 -0700] é a data, hora e fuso horário que a solicitação foi recebida, por padrão no formato strftime % d /% b /% Y:% H:% M: % S% z .

"GET /apache_pb.gif HTTP / 1.0" é a linha de solicitação do cliente. O método GET , /apache_pb.gif o recurso solicitado e HTTP / 1.0 o protocolo HTTP .

200 é o código de status HTTP retornado ao cliente. 2xx é uma resposta bem-sucedida, 3xx um redirecionamento, 4xx um erro do cliente e 5xx um erro do servidor.

2326 é o tamanho do objeto retornado ao cliente, medido em bytes .

Arquivo de log escolhido para o teste

http://pathalizer.sourceforge.net/wiki-access.log

Tratamento dos dados

Primeiramente é necessário baixar o arquivo log, nesse caso usando o comando : `wget` http://pathalizer.sourceforge.net/wiki-access.log

Feito o download pode-se começar a tratar o código primeiramente foi feito o seguinte comando:

```
cat wiki-access.log | awk {'print $1,$4,$12'} | sed 's/"//g' | sed 's/\[//g' | sed 's/ /-/g' | sed 's/:/-/g' | sed 's/\//-/g'   >> log1.txt
```

Este comando selecionou a primeira a quarta e a decima segunda coluna, substituindo aspas por espaço em branco, removendo colchete, substituindo espaço em branco por traço, dois pontos por traço e barra por traço.

Além de salvar as alterações no arquivo log1.txt que ficou da seguinte forma:

222.64.146.118-19-Jun-2005-06-44-17-Mozilla-4.0

218.84.191.50-19-Jun-2005-06-46-05-Mozilla-4.0

202.201.245.20-19-Jun-2005-06-47-37-Mozilla-4.0

138.243.201.10-19-Jun-2005-06-48-40-Mozilla-5.0

68.251.52.253-19-Jun-2005-06-50-49-Mozilla-5.0

```
cat log1.txt | sed 's/-/ /g' >> log2.txt
```

Após salvar as alterações o arquivo log1.txt foi alterado substituindo os traços por espaços e salvando as alterações no arquivo log2.txt.

222.64.146.118 19 Jun 2005 06 44 17 Mozilla 4.0

218.84.191.50 19 Jun 2005 06 46 05 Mozilla 4.0

202.201.245.20 19 Jun 2005 06 47 37 Mozilla 4.0

138.243.201.10 19 Jun 2005 06 48 40 Mozilla 5.0

```
cat log2.txt | awk {'print $1,$5,$8'} >> log3.txt
```

Depois do arquivo log2.txt foi pego a coluna 1 , 5 e 8 e essas colunas foram salvas no arquivo log3.txt, obtendo assim somente ip, hora e navegador.

222.64.146.118 06 Mozilla

218.84.191.50 06 Mozilla

202.201.245.20 06 Mozilla

138.243.201.10 06 Mozilla

Conclusão

Foi utilizado comandos do cent os 7 para transformar um arquivo log completo em um log.txt com as 3 informações que o software utiliza.