Sistema Operacionais: Trabalho Extra SysCall Tracer
Alan Portela Bandeira
15/06/2017

A nova versão do projeto apresentada inclui uma simples inteface em linha de comando para auxílio do usuário. A aplicação compreende as seguintes funções:

1. Import Syscall Log: Importa o arquivo de logs gerados pelo strace
2. Inport ".strace" File: Importa os arquivos .strace fornecidos no link para teste. Estes arquivos contem algumas informações extras em relação aos gerados pelo strace no arquivo .log
3. Consult DataBase: Consulta o banco de dados (árvore ternária) pelo gram fornecido por linha de comando.
4. Output N-Gram Data: Gera um arquivo com todos os grams de tamanho N, sendo N fornecido por linha de comando.
5. Calculate N-Gram Score: Calcula o score de um arquivo que contem todos os grams no arquivo .gram com os grams presentes np banco de dados.
6. Clear Data: Limpa o banco de dados.
7. Exit: Encerra o programa.

Para os testes realiza-se o seguinte procedimento:
1 – Importar o arquivo de teste (possivelmente infectado).
2 – Grar o arquivo .gram do arquivo de teste.
3 – Limpar o banco de dados.
4 – Importar o arquivo base (saudável)
5 – Calcular o grau de similaridade entre o arquivo .gram e os dados no baco.

Os testes fora realizados utlizando os arquivos do link: http://www.kayacik.ca/datasets/syscalls/
Foram utilizadas 5 instancias sadias e 2 infectadas, para um gram de tamanho 10 os resultados são apresentados na tabela a seguir:

Scores(absoluto, percentual)		
	case6_attack	case7_attack
case1	(0.8996655518394648, 89%)	(0.8734693877551021, 87%)
case2	(0.8996655518394648, 89%)	(0.8734693877551021, 87%)
case3	(0.8996655518394648, 89%)	(0.8734693877551021, 87%)
case4	(0.9565217391304348, 95%)	(0.9510204081632653, 95%)
case5	(0.9264214046822743, 92%)	(0.9061224489795918, 90%)

O método foi capaz de diferenciar arquivos infectados de arquivos saudáveis com um grau de diferença entre eles de no mínimo 87% para grams de tamanho 10 onde, a principal característica de infecção é a presença de chamadas de sistema que existem na versão infectada e não existem na versão sadia. O modelo aplicado pode também identificar a ocorrência de grams composto por chamadas existentes mas que não correspodem a nenhum gram presente no modelo base.

Todo o projeto encontram-se na pasta (arquivos de scrpit, arquivos para teste).