Segurança em Redes de Comunicações
Segurança em Redes de Comunicações

Segundo Relatório

Professores:

Paulo Salvador salvador@ua.pt ;
António Nogueira nogueira@ua.pt ;

Objetivo: Defina regras SIEM para detectar comportamentos anômalos de rede e dispositivos possivelmente comprometidos. Teste as regras definidas em um registro de dados de fluxos de tráfego IP, identificando os dispositivos comprometidos.

Descrição:

Uma rede corporativa possui um sistema SIEM com os dados históricos dos fluxos de tráfego na rede. Para implementar um sistema de Cibersegurança confiável é necessária a implementação de regras de alerta, de possíveis ataques, baseadas em comportamentos anómalos.

Considere o conjunto de dados (arquivo datasetX.zip) com os arquivos dataX.parquet, testX.parquet e servidoresX.parquet, onde X é o restante da divisão da soma dos números dos alunos por 10: em python: X=(num_mec1+num_mec2)% 10

Utilizando dados de um dia inteiro (arquivo dataX.parquet) defina o comportamento típico dos dispositivos de rede. Esses dados já foram totalmente analisados e nenhum comportamento ilícito foi detectado. Você pode presumir que o endereço privado IPv4 de cada dispositivo não muda com o tempo e é atribuído ao mesmo usuário final.

O arquivo testX.parquet contém dados de um dia completo e pode conter comportamentos anômalos resultantes de atividades ilícitas dentro da rede, como atividades internas de botnet, exfiltração de dados e C&C remoto de dispositivos. O arquivo serverX.parquet contém dados de um dia inteiro de acessos externos aos servidores da corporação (na rede 200.0.0.0/24) de um pequeno conjunto de clientes na mesma rede, podendo conter usuários externos interagindo com os servidores da corporação de forma forma anômala (dica: não é a quantidade de tráfego ou fluxos).

Cada arquivo de dados *.parquet contém a lista de todos os fluxos de dados IPv4 observados com as seguintes informações sobre cada fluxo (colunas):

- timestamp: horário de observação do primeiro pacote do fluxo, em 1/100 de segundo a partir das 0h do dia;
- src_ip: endereço de origem IPv4 (para arquivos dataX e testX identifica o dispositivo interno, para o arquivo serverX identifica o cliente externo);
- dst_ip: endereço de destino IPv4 (identifica o servidor externo ou interno);
- proto: protocolo de transporte utilizado (tcp ou udp);
- porta: porta de destino;qq
- up_bytes: total de byes enviados;
- down_bytes: total de bytes baixados.

NOTA IMPORTANTE:todos os endereços IPv4 públicos representam redes reais, mas (além das estatísticas de fluxo) apenas o proprietário e a localização são relevantes. A real finalidade/serviços do mesmo não são relevantes!**NÃO REALIZE VERIFICAÇÕES DE SERVIÇO/VULNERABILIDADE NOS ENDEREÇOS IPv4!**

Os dados são estruturados usando pandas e armazenados em formato parquet. Ver:https://pandas.pydata.org/ e https://parquet.apache.org/ . Verifique o script python fornecido (sampleScript.py) com exemplos básicos de como ler e processar o arquivo de dados. A geolocalização baseada no endereço IPv4 deve contar com bancos de dados externos (GeoIP_DBs.zip). Verifique também o script python fornecido com exemplos básicos sobre como realizar geolocalização de IP e consultas DNS.

- Apresentar relatório com as regras SIEM propostas e testes de regras (detecção de dispositivos anômalos). Envio via e-learning, em formato PDF, até 11 de junho. Deve ser feito por um grupo de 2 alunos. Excepcionalmente, pode ser feito individualmente.

- Tarefas:
  - Análise dos comportamentos não anómalos; identificar servidores/serviços internos, descrever e quantificar trocas de tráfego de usuários internos com servidores internos e externos, e descrever e quantificar trocas de tráfego de usuários externos com servidores públicos da corporação (4 pontos).
  - Definição das regras do SIEM e respetiva justificação para deteção de atividades internas de BotNet, exfiltração de dados através de HTTPS e ou DNS, atividades de C&C através de DNS, e utilizadores externos que utilizem os serviços públicos corporativos de forma anómala. (6 pontos).
  - Teste das regras do SIEM e identificação dos dispositivos com comportamentos anómalos (6 valores).
  - Relatório escrito; estrutura e conteúdo (4 pontos).