Universidade de Aveiro

Departamento de Eletrónica, Telecomunicações e Informática

Segurança em Redes de Comunicações

Mini Projeto 1

Eduardo Cruz (93088)

10 de Junho de 2023

Índice

Introdução

Este relatório visa a descrever e realizar a análise da resolução do guião para o miniprojecto dois no âmbito da unidade curricular de Segurança em Redes de Comunicações.
O miniprojecto dois apresenta o seguinte objetivo:

"Com base no registo de dados dos fluxos de tráfego IP, defina regras SIEM para detetar comportamentos anómalos da rede e dispositivos possivelmente comprometidos."

Alguns casos de comportamentos anómalos da rede que serão considerados:

- atividades internas de botnet,
- Exfiltração de dados
- C&C remoto de dispositivos.

O dataset (dataset8.zip file) que contem os ficheiros data8.parquet e test8.parquet, releva-se como sendo o conjunto de dados de fluxo onde fora realizada a analise para definir regras e detetar comportamentos.

Os ficheiros de dados *.parquet contêm a lista de todos os fluxos de dados IPv4 observados com as seguintes informações sobre cada fluxo (colunas):

- - timestamp: hora de observação do primeiro pacote do fluxo, em 1/100 de segundos a partir das 0h do dia;
- - src_ip: Endereço IPv4 de origem;
- - dst_ip: Endereço IPv4 de destino (interno ou externo);
- - proto: protocolo de transporte utilizado (tcp ou udp);
- - port: porta de destino;
- - up_bytes: total de byes carregados;
- - down_bytes: total de bytes descarregados.

O relatório encontra-se divido em 4 tarefas, sendo estas respetivamente:

- Análise e descrição de comportamentos não anómalos
- Deteção de comportamento anómalo, descrição e possíveis causas
- Regras SIEM
- Teste das regras SIEM e identificação dos dispositivos com comportamentos anómalos

o código de configuração que fora desenvolvido para a resolução das tarefas encontra-se disponível no link:
github..

No entanto algum código é exposto no relatório de forma a facilitar a sua leitura.

Análise de Resultados

Non-anomalous behavior analysis and description

Comecemos pela análise de quais os portos utilizados em um dia inteiro, para tal iremos utilizar o ficheiro data8.parquet, o qual será o ponto base para definir o comportamento típico dos dispositivos de redes com nenhum comportamento ilícito.

O código presente no ficheiro 'destinyports.py' permite observar e guardar em um ficheiro .csv denominado 'unique_destination_ports_data' todos os portos únicos utilizados, sendo o resultado os portos seguintes:

- 53: Comunicações Udp para realização de Dns
- 443: Comunicação por protocolo Tcp

A informação sobre protocolos é obtida de forma similar através de análise de dados e o através do programa 'destinyprotos.py' observando assim que os únicos protocolos são UDP e TCP, esta informação encontra-se no ficheiro 'unique_protocols_data.csv'.

Em termos de localizações de destino estes são possíveis de observar pelo código descrito em 'connections.py' que tenta realizar a tradução de todos os ips destino únicos em um organização e pais, sendo os resultados armazenados em 'geolocation_destinations_data.csv'. A partir deste ficheiro e com o código do programa 'destiny_geolocations.py' obtemos dois ficheiros, 'unique_countries_data.csv' e 'unique_organizations_data.csv' com todos os países e organizações únicas respetivamente, sendo esta a base a partir do qual iremos operar regras SIEM com base em países e organizações com o qual o sistema comunica em um dia considerado 'normal', (não apresentado comportamentos maliciosos).

O programa 'data_analysis.py' realiza uma análise mais complexa, calculado para cada conjunto ip fonte, ip destino, protocolo, o, tempo total das conexões, um ratio entre o número de bytes de download e upload, número de tentativas de ligação total, número máximo de ligações que ocorrem numa janela de 5 minutos e o número total de bytes de upload e download, organização e pais se o ip for publico. Os valores encontram-se no ficheiro 'network_analysis_data.csv'

O programa 'ratio_analysis.py' calcula a mediana do ratio de download e upload de dados, desvio padrão e a mediana total dos dados para um ficheiro .csv denominado por 'ratio_analysis_data.csv', no entanto cria outro ficheiro 'top_3_highest_ratio_data.csv' com os três ratios mais elevados para cada ip fonte. A mediana total revela-se de 12.768, com um desvio padrão de 12. 134.Todos os ratios mais elevados revelaram ter como na sua grande maioria, destino serviços de Google ou Microsoft e pais destino maioritariamente nos US. É de realçar o número reduzido de comunicações por minuto.

É utilizado um ratio invés dos valores de download/upload pois permitem atingir valores mais simples e eficazes, sendo que nos permite também obter uma base para detetar quando um ip se comporta ilicitamente como por exemplo, durante o processo de extração de dados onde realizada um upload elevado, ratio significativamente menor a mediana, para um ip que não pertence a rede interna, ou onde realiza downloads proveniente de redes externas, ratio significativamente superior a mediana.

Anomalous behavior detection, description, and possible causes

Realizemos agora a análise ao ficheiro test8.parquet onde se encontra os dados não filtrados da empresa.

Segundo o mesmo código utilizado anteriormente fora, analisado os portos únicos e protocolos, estando estes armazenados nos ficheiros, 'unique_destination_ports_test.csv' e' unique_protocols_test.csv', respetivamente.

Como os resultados apresentam valores semelhantes ao data8, sendo os portos e protocolos utilizados,53 e 443, UDP e TCP, sendo UDP unicamente utilizado para DNS, não existe necessidade de analisar mais profundamente esta vertente, porém caso fossem distintos, uma análise mais rigorosa seria necessário, pois seria indícios de comportamento anormal.
Fora igualmente realizado novamente a analise de organizações e países dos ips de destino das comunicações, sendo os resultados armazenados no ficheiro, 'geolocation_destinations_test.csv'.

Existindo a necessidade de verificar se existira conexões com destino distintas do ficheiro data8 fora alterado o programa 'destiny_geolocations.py' novamente, obtendo todos os países e organizações individualmente em ficheiros diferentes, nomeadamente 'unique_organizations_test.csv' e 'unique_countries_test.csv' e aplicando o programa 'data_notin_csv.py' para determinar quais os novos países e organizações que não se apresentaram presentes quando comparados com os ficheiros relacionado ao data8.
Os resultados desta análise superficial demonstram que existe comunicações com países 'novos', tais como Roménia, Quirguistão, Camboja, entre outros, os quais não se apresentam como o perfil mais indicado para realizar comunicações tendo em conta o perfil de dados data8 o que indica que existe ips fonte comprometidos, o ficheiro 'country_code_not_in_data2.csv' apresenta todos códigos de países encontrados.

O ficheiro 'organization_not_in_data2' revela uma lista extensa de organizações não presentes em data8, sendo que existe a possibilidade de existir organizações não maliciosas, a grande maioria apresenta esse caracter.

O programa 'ratio_analysis.py' revelou uma mediana total de 12.768 e desvio padrão de 12.134 o que é desconcertante devido a ser um valor semelhante ao do data8.

O programa 'src_ips_notin_data.py', revelou a existência de três novos ips fonte ativos, que não existiam em data8, sendo estes, [192.168.108.63; 192.168.108.208; 192.168.108.87], os quais se encontram no ficheiro 'ips_not_in_data2.csv'. Desta forma é necessário confirmar se estes ips se encontram comprometidos, não obstante o facto da possibilidade de ips que já se encontrariam ativos se encontrarem agora comprometidos.

O programa 'non_dsrc_ips.py' permitira observar a mediana do ratio de bytes download/upload e os três ratios mais altos, armazenado estes valores em 'median_top_3_highest.csv' fora igualmente guardado todas as comunicações feitas por estes três ips num ficheiro, 'filtered_connections.csv'

O programa' destiny_search.py' permitira determinar quais os fluxos de ips que se encontravam a a realizar trocas de informação com os países e organizações 'maliciosas' determinadas posteriormente sendo guardada a sua informação no ficheiro 'filtered_connections_countries_organizations.csv'. O programa 'src_ips_find.py' demonstra os resultados de ips unicos no ficheiro 'unique_src_ips_malicious', revelando deste modo 6 ips comprometidos

- 192.168.108.153
- 192.168.108.156
- 192.168.108.158
- 192.168.108.202
- 192.168.108.205
- 192.168.108.44

Estes serão através de regras SIEM bloqueados, até o seu comportamento voltar ao 'normal'.

Para detetar um volume significativo de bytes de download e upload nos fluxos, o que podem indicar atividades de botnet e C&C, o programa 'detection_bytes.py', fora criado no qual iria analisar, data em data8 e cria uma linha base para os valores de download e upload, ao qual iremos aplicar limiar de acréscimo de 25%, expondo no ficheiro 'unique_src_ips_botnets.csv 'todos os ips fonte únicos que apresentavam estas características. Um problema encontrado fora o facto de como ser usado a mediana na linha base, para fluxo ip fonte-ip destino, todos os ips da rede foram enunciados como possivelmente maligno não estando deste modo a ser uma fonte confiável de dados. Através da analise do ficheiro 'top_3_highest_ratio_test.csv' é observável que existe um fluxo massivo de dados na maioria dos fluxos para serviços Google e Microsoft o que destabiliza a analise confiável de fluxos maliciosos. Fora igualmente identificados fluxos com destino serviços CloudFlare porém, não sabendo o seu objetivo mesmo realizando DNS reverso, mesmo não devendo iremos confiar na sua autentacidade destas comunicações, pois não é possível realizar um analise mais profunda á informação que está a ser transmitida.
Tendo isto em conta, fora criado um novo .csv para armazenar todos os fluxos, sem a existência dos fluxos de Organizações Google e Microsoft para verificar os dados sobre um espectro diferenciado, programa 'remove_MG', ficheiro 'filtered_test_network_analysis.csv'.

Realizado novamente o programa 'ratio_analysis.py', fora obtido os resultados nos ficheiros, 'ratio_analysis_noMG.csv' e 'top_3_highest_ratio_noMG.csv', revelando uma mediana para o ratio de download/upload significativamente reduzida de 8.401, e desvio padrão de 1.554. O mesmo processo fora realizado para data8, revelando valores de 8.400 e desvio padrão de 1.517. O que nos permite concluir se existe botnet e atividade C&C será difícil de determinar por estes métodos.

Uma metodologia possível de utilizar seria a analise de ips destino públicos, com uma base de dados com ips conhecidos, pertencentes a botnet. Para estudar, caso o mesmo se passasse internamente fora criado o código 'outliers.py' para detetar os valores anómalos no número de comunicações realizadas internamente. Os fluxos apresentam-se no ficheiro 'outliers_formatted.csv', com essa informação foi possível determinar que

o ip da interface do router é 192.168.108.240 sendo que todos os ips internos comunição extensivamente com este. O processo fora realizado com o ficheiro data8 e a informação descrita anteriormente é novamente reiterada, validando assim a sua veracidade.Outra informação pertinente é de não existir comunicação interna entre ips no ficheiro proveniente de data8 o que nos permite concluir que qualquer comunicação interna é suspeita de ser maliciosa. O código fora atualizado para remover todas as entradas com ip destino do router e guardar no ficheiro 'outliers_formatted_norouter.csv' revelando assim 4 ips únicos:

- 192.168.108.110
- 192.168.108.181
- 192.168.108.76
- 192.168.108.93

Sendo que os dois primeiros comunicam entre si e os últimos dois realizam comunicações com os dois primeiros. Comportamento extremamente suspeito e o qual será bloqueado quando as regras SIEM forem definidas.

SIEM rules

Analisado ambos os datasets é possível deste modo então criar regras de forma a tornar a rede mais segura a atividades maliciosas.
As regras que iremos definir serão com base nos ficheiros que foram obtidos das análises anteriores.

1º Regra - Comunicação Interna

Realização do bloqueio de todas as comunicações internas exceto para o router, visto que os fluxos normais são sempre com serviços de exterior não existe a necessidade de permitir a realização de comunicação interna, exceto com o ip que permite comunicação ao exterior.
Esta regra tem o ponto negativo de quando necessária manutenção ou debug interno este vir a ser difícil de realizar, pois comunicação não é permitida.

2º Regra – Portos e Protocolos

Realização do bloqueio de todas as portas exceto, 53 e 443, visto todas as comunicações realizaram-se através destes dois portos é boa pratica, reduzir o acesso a rede interna por parte de membros externos. Possivelmente seria possível permitir mais um porto para comunicações UDP, sem serem de DNS, porém como não são efetuadas, a sua permissão é sujeita a uma consideração com a empresa sobre que tipo de serviços os utilizadores possam aceder.

3º Regra – Pais de Comunicação Destino

Apesar de não ser a forma mais prática de realizar este bloqueio, realizar o bloqueio de países, 'não seguros' tais como os encontrados no ficheiro 'country_code_not_in_data2.csv' permite reduzir o impacto de comunicações com serviços potencialmente maliciosos. Qualquer país que não se encontre no ficheiro

‘unique_countries_data.csv’ será bloqueado e caso seja necessária comunicação com um serviço de um outro país deveria ser realizado de forma segura ou notificado com a equipa de segurança sobre tal. (Existindo formas de contornar esta regra, como por exemplo uso de VPN, para os membros que quisessem quebrar esta regra.)

4º Regra – Número de acessos por minuto

Bloquear ou reduzir o número de acessos que se podem realizar em um espaço de tempo, iria reduzir significativamente o risco de em caso de um ip interno ser comprometido a quantidade de dano ser significativamente reduzido. O ponto desfavorável seria o atrase em certos serviços, porém caso estes fossem cruciais e críticos seria benéfico as criações de uma regra SIEM para dar prioridade a estes serviços.

5º Regra – Quantidade de dados

Apesar de não ser possível provar a existência de exfiltração de dados pelos métodos utilizados, a prevenção, limitação e bloqueio de fluxos com um elevado número de bytes de upload ou download, quando comparados com o tráfego ‘regular’ do ip e o comportamento da rede, permite a realização de uma boa prática para criar uma segurança acrescida nesta.

6º Regra – Bloqueio de Serviços

Outra regra seria o bloqueio de serviços específicos como pornografia e outros, apesar de não ter sido mencionado nas seções anteriores foram produzidos scripts em bash para realizar a verificação reversa de DNS, no ficheiro ‘scripts.txt’, porém não foram utilizados devido ao barramento periódico das ‘queries’ de DNS que as firewalls apresentavam, diminuindo assim a utilidade deste script.
Não obstante o bloqueio de serviços, permite de uma forma mais minuciosa, bloquear todos os serviços que poderiam infetar ou produzir comportamentos maliciosos para a rede, devido poderem ser não seguros.

7º Regra – Monitorização de utilizadores privilegiados

Caso soubéssemos os ips fonte dos utilizadores com privilégios poderíamos colocá-los numa lista especial para uma monitorização mais minuciosa sendo que estes, apesar de todas as críticas de privacidade que possam efetuar, encontram-se como as entidades mais suscetíveis a serem atacadas devido ao seu grau de importância e o qual não pode acontecer, caso se encontrem comprometidos o poder destrutivo que podem alcançar é extremamente elevado.

SIEM rules test and identification of the devices with anomalous behaviors

As regras descritas anteriormente foram criadas e implementadas pelo programa 'SIEM_Rules.py', o qual irá ler o ficheiro de dados 'data8.parquet', realizando a analise, expondo todos os ips fonte que se encontram comprometidos, pois não respeitam as regras definidas, armazenando-os em um ficheiro, denominado por 'compromised_src_ips.csv '.
A regra 7 não fora implementada pois a informação necessária não existe, poderia ser possível criar um cenário artificial, porém este não iria demonstrar a qualidade ou veracidade da regra mas sim a manipulação de informação para permitir a regra funcionar.
A regra 6 não fora implementada devido a não ser possível verificar quais os serviços que estão atualmente a ser utilizados em cada fluxo, porém esta regra poderia ser testada em um dataset menor e com uma base de dados DNS em que já saberíamos todos os Dns necessários a partida, o mesmo poderia ser feito utilizando uma base de dados de ips botnet, bloqueando quais ips destino que se apresentassem nessa base de dados.

Podemos observar que as regras se demonstram funcionais, tendo em conta que todos os ips comprometidos observados no excerto 'Anomalous behavior detection, description, and possible causes' se encontram identificados no ficheiro.