Monitoring & SIEM - Trabalho Prático 2 de Segurança em Redes de Computadores

Autores

- Ana Vidal (118408)
- Simão Andrade (118345)

Estrutura do Relatório

1. Introdução
2. Objetivo
3. Conteúdo utilizado
4. Implementação
    i. Processo de Análise
    ii. Análise dos comportamentos não anómalos
    iii. Definição das regras SIEM
    iv. Teste das regras SIEM e identificação dos dispositivos comprometidos
5. Conclusão

Introdução

O objetivo principal deste projeto é a definição de regras de Sistema de Gestão de Informação e Eventos de Segurança (SIEM) para a deteção de comportamentos anómalos em redes de comunicação e a identificação de dispositivos possivelmente comprometidos.

A análise será conduzida utilizando um conjunto de dados de fluxos de tráfego IP.

Objetivo

Na realização deste trabalho, as tarefas a serem realizadas são as seguintes:

- [x] Análise dos comportamentos não anómalos (4 valores):
    - [x] Identificar servidores/serviços internos
    - [x] Descrever e quantificar as trocas de tráfego dos utilizadores internos com os servidores internos e externos
    - [x] Descrever e quantificar trocas de tráfego dos utilizadores externos com os servidores públicos da empresa
- [x] Definição das regras SIEM (6 valores):
    - [x] Respetiva justificação para a deteção de atividades BotNet internas
    - [x] Exfiltração de dados usando HTTPS e/ou DNS
    - [x] Atividades de C&C usando DNS e utilizadores externos usando os serviços públicos empresariais de forma anómala
- [x] Teste das regras SIEM e identificação dos dispositivos com comportamentos anómalos (6 valores).
- [ ] Relatório (4 valores)

Conteúdo utilizado

Para a realização deste trabalho, foram disponibilizados os seguintes ficheiros:

- Datasets: `dataset3.zip`
  - Dataset não anómalo: `dataset3.parquet`
  - Dataset anómalo: `teste3.parquet`
  - Dataset apenas c/ comunicação externa: `servers3.parquet`
- GeoIP_DB: `GeoIP_DB.zip`
  - Base de dados para identificar o *Autonomous System* (rede de IPs de uma organização) de um IP: `GeoIP_ASNum.dat`
  - Base de dados para identificar a localização geográfica de um IP: `GeoIP.dat`

Implementação

Processo de Análise

Foi inicialmente nesta fase definido um conjunto de análises a serem realizadas sobre os datasets fornecidos, para obter uma visão geral dos comportamentos a procurar.

- **Inicialmente**:
  - [x] Ip's de origem e destino
  - [x] Portas comuns
  - [x] Protocolos comuns
  - [x] Número de pacotes (por src_ip)
  - [x] Rácio de download/upload (por src_ip)
  - [x] Localização geográfica dos IPs (dos dst_ip para cada src_ip)
  - [ ] Domínios DNS visitados (por src_ip)
  - [ ] Fazer mais análise às comunicações internas (src_ip e dst_ip)
  - [x] Número de conexões por ~~hora~~ (por src_ip)
- **Seguidamente**:
  - [x] Detetar atividades de BotNet (número de conexões por hora)
  - [x] Detetar exfiltração de dados (Taxas anómalas de transferência de dados)
  - [x] Detetar atividades de C&C (número e tamanho de pacotes DNS anómalo)
- **Finalmente**:
  - [x] Identificar dispositivos comprometidos (identificar os IPs que violam as regras definidas)
  - [ ] Tentar identificar o tipo de comprometimento (BotNet, exfiltração de dados, C&C)
  - [ ] Justificar a identificação dos dispositivos comprometidos

Análise dos comportamentos não anómalos

Nesta secção, são analisados os comportamentos não anómalos dos utilizadores internos e das suas comunicações com a rede interna e com as redes externas.

Protocolos utilizados

Na análise dos comportamentos não anómalos, observamos que os pacotes de dados são divididos principalmente entre dois protocolos de transporte: TCP e UDP.

Conforme indicado nos gráficos abaixos, a maioria dos pacotes são do protocolo TCP, representando cerca de 88.06% do tráfego total, enquanto os pacotes UDP correspondem a 11.94%.

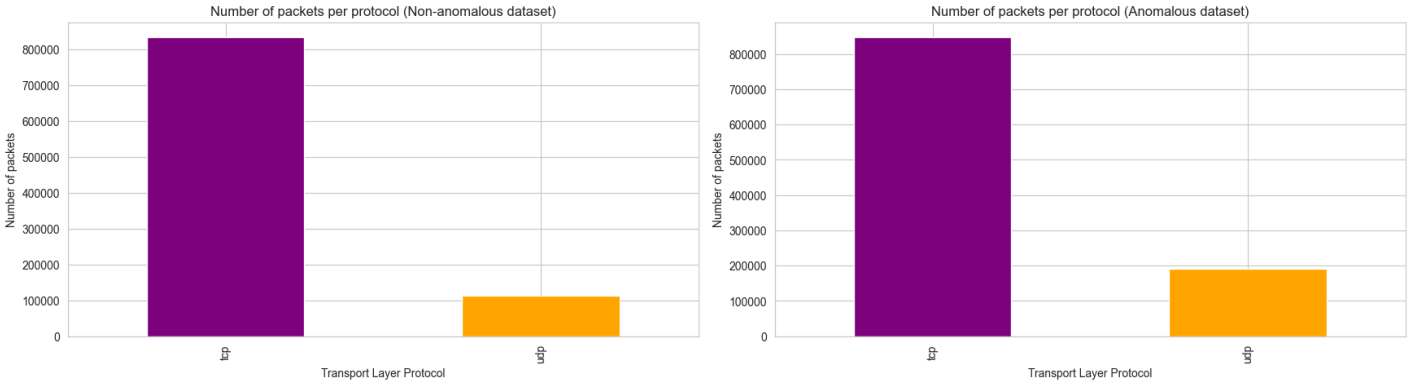

Figura 1: Protocolos utilizados no _dataset_ não anómalo (esquerda) e anómalo (direita)

Esta distribuição é consistente com a utilização típica de redes corporativas.

Portas utilizadas

A análise das portas utilizadas mostra uma predominância do tráfego HTTPS, seguido por DNS. Como podemos observar no gráfico abaixo, quase 90% dos pacotes são HTTPS, indicando um uso intensivo de navegação web segura e serviços relacionados.

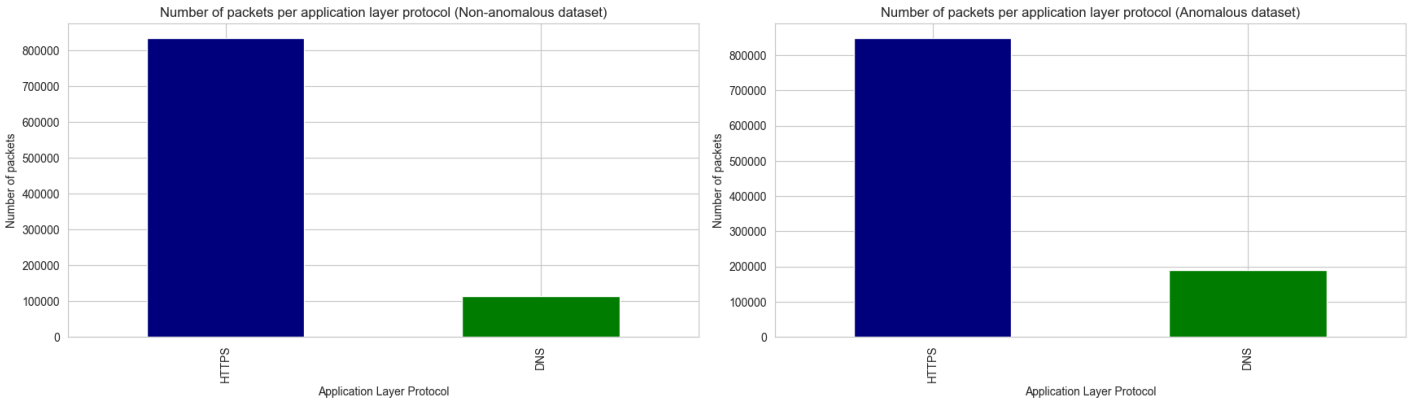

Figura 2: Portas utilizadas no _dataset_ não anómalo (esquerda) e anómalo (direita)

O *dataset* anómalo mostra uma pequena diferença entre o número de pacotes DNS e HTTPS, com um número mais elevado de pacotes DNS (>7% relativamente ao *dataset* não anómalo).

Este facto pode indicar ataques como o *DNS flooding*.

Número de pacotes (por endereço de origem)

Os gráficos abaixo apresentam o número de pacotes enviados pelos 100 principais endereços IP de origem no dataset não anómalo. Os endereços IPs mais ativos foram identificados e contabilizados, tendo dois endereços com um envio maior.

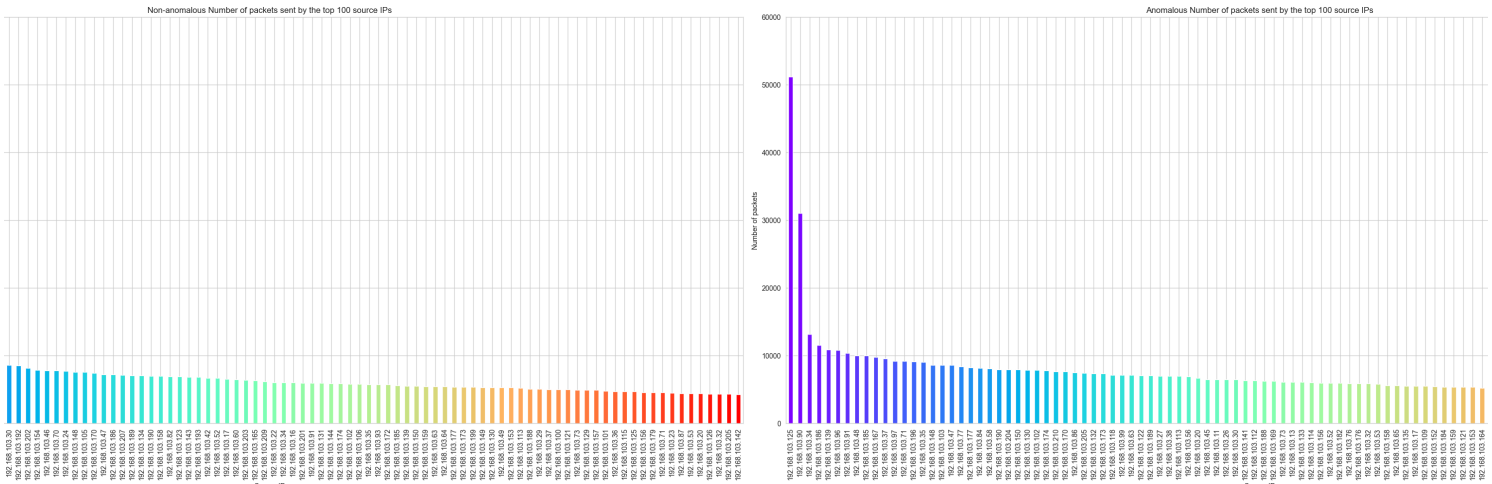

Figura 3: Número de pacotes enviados pelos 100 principais endereços IP de origem no dataset não anómalo (esquerda) e anómalo (direita)

Rácio de *download*/*upload* (por endereço de origem)

Esta análise do rácio revela que, em geral, a quantidade de bytes baixados é significativamente maior do que a quantidade de bytes enviados, como ilustrado nos gráficos abaixo.

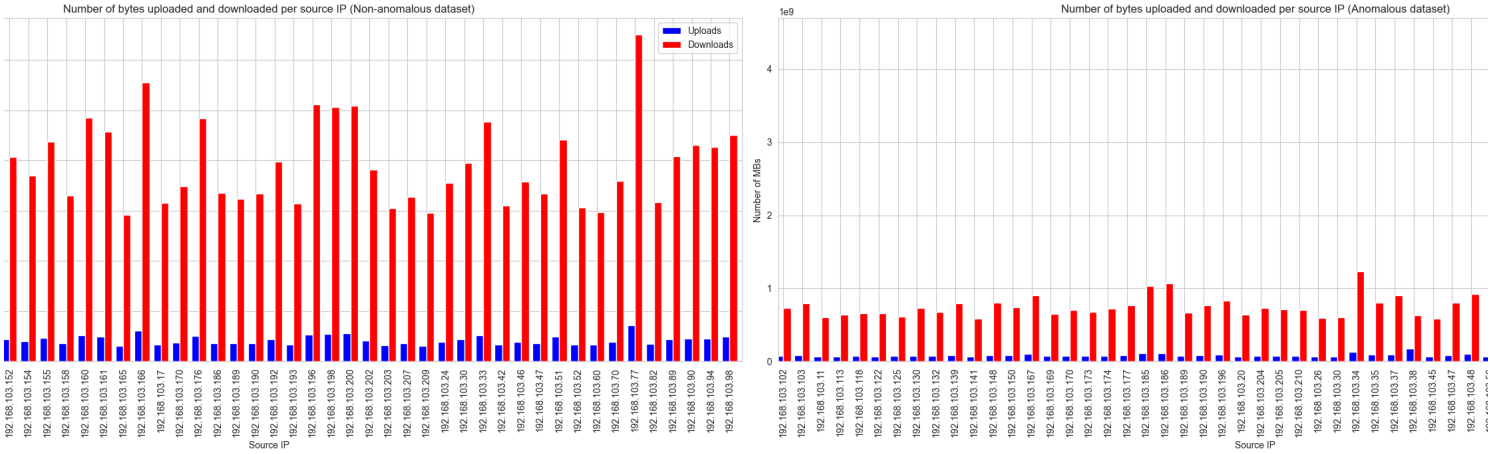

Figura 4: Rácio de _download_/_upload_ por endereço de origem no dataset não anómalo (esquerda) e anómalo (direita)

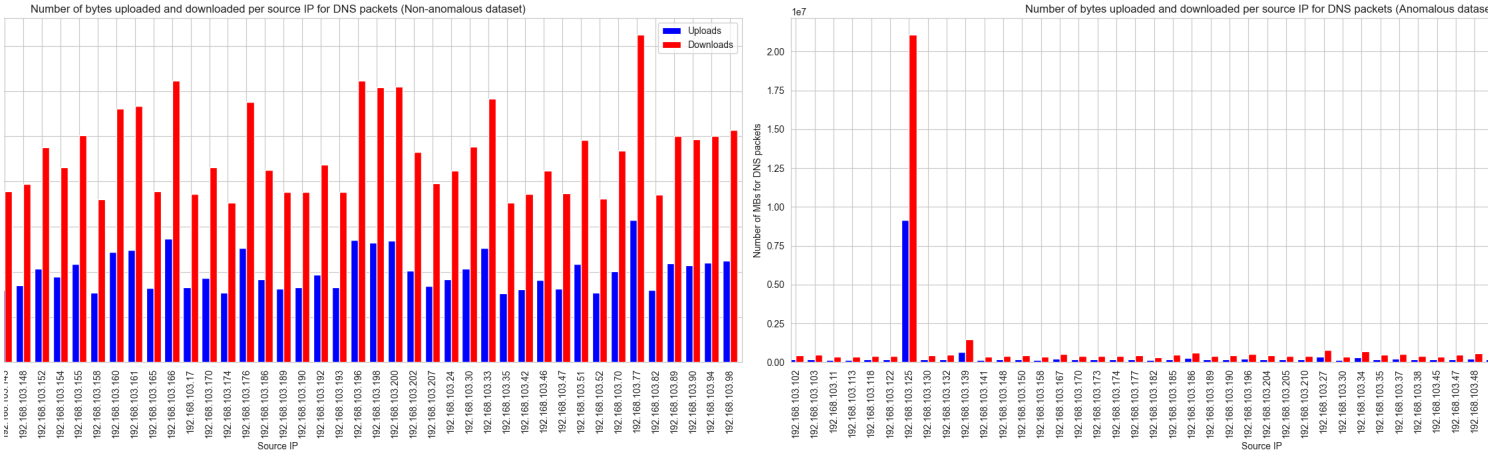

Figura 5: Rácio de _download_/_upload_ por endereço de origem no dataset não anómalo (esquerda) e anómalo (direita) para pacotes DNS

Localização geográfica dos IPs

A localização geográfica dos IPs de destino mostra que a maioria do tráfego é direcionada para os Estados Unidos, seguido por Portugal. Este padrão é esperado, dada a natureza global das comunicações corporativas. Contudo, na análise do dataset anómalo, surgiram **38** novos países no *dataset* anómalo, o que pode indicar atividades suspeitas.

```
Anomalous dataset:
country_code
US    32.510430
PT    24.892034
NA     1.842787
NL     1.799620
DE     1.693245
        ...
IQ     0.000193
IS     0.000096
MD     0.000096
VN     0.000096
MX     0.000096
```

Foi definido um *threshold* de variação de 1% e dentro destes países, destacam-se a Rússia e a Ucrânia:

```
   country  variation  is_new
37      RU   0.273937    True
38      UA   0.011370    True
```

Número de conexões por hora (por endereço de origem)

Definição das regras SIEM

Nesta secção, são definidas regras SIEM para detetar comportamentos de rede anómalos e dispositivos possivelmente comprometidos.

As regras seguem a seguinte estrutura:

1. **Detetar**: Descrição da atividade que se pretende detetar.
2. **Justificação**: Razão pela qual a atividade é considerada anómala.
3. **Regra**: Condição que, se verificada, deteta a atividade.

Regras Identificadas

Regras para fluxos de tráfego interno-interno e interno-externo

1. **Alta percentagem de tráfego enviado por um único utilizador**
   - **Detetar**: Atividades de BotNet, Ataques de DDoS
   - **Justificação**: Um evento de DDoS pode ser detetado se um utilizador enviar uma abundância de tráfego para um (ou multiplos) destino, o que pode sobrecarregar o servidor e causar uma interrupção do serviço.
   - **Regra**: Se um utilizador enviar mais de 2% do tráfego total detetar anomalia.
2. **Taxas anómalas de *upload* de dados (no geral)**
   - **Detetar**: *Data Exfiltration*, Má configuração de serviços, *Malware*/*Ransomware*
   - **Justificação**: A exfiltração de dados é uma técnica comum usada por atacantes para roubar informações sensíveis. Se um utilizador enviar uma quantidade anormalmente grande de dados para fora da rede, isso pode indicar que informações confidenciais são roubadas e a serem enviadas para um servidor externo.
   - **Regra**: Se a variação dos dados enviados em comparação com uma quantidade considerada normal for superior a 1% ou menor que -1%, então detetar anomalia.
3. **Taxas anómalas de *download*/*upload* de dados via DNS**
   - **Detetar**: Atividades de C&C (*Command & Control*) ou Ataques de *DNS Amplification*
   - **Justificação**: Taxas altas de *download* comparadas com *upload* podem indicar respostas DNS excessivamente grandes, típicas de ataques de amplificação DNS ou comunicação maliciosa via DNS.
   - **Regra**: Se o número de pacotes DNS enviados tiver uma variação superior a 0,8 em comparação com os valores normais, então detetar anomalia.
4. **Taxas altas de comunicação com máquinas localizadas em novos países**
   - **Detetar**: Acesso não autorizado, *Data Exfiltration*
   - **Justificação**: A comunicação com máquinas localizadas em novos países pode indicar que uma das máquinas envolvidas foi comprometida e encontra-se a comunicar com um servidor malicioso localizado noutro país.
   - **Regra**: Se a variação de pacotes enviados para um novo pais for superior a 1% em comparação com os valores normais, então detetar anomalia.

Regras para fluxos de tráfego externo-interno

1. **Deteção de Picos Anómalos no Tráfego de *downloads***
   - **Regra**: Configurar alertas para quando o tráfego de *downloads* exceder 400,000 bytes num curto período de tempo.
   - **Justificação**: Picos significativos no tráfego de *downloads* podem indicar a transferência de abundância de dados, possivelmente devido a *downloads* massivos ou sincronizações de dados, o que pode ser um sinal de atividades suspeitas ou não autorizadas.
2. **Deteção de Picos Anómalos no Tráfego de *uploads***
   - **Regra**: Configurar alertas para quando o tráfego de *uploads* exceder 50,000 bytes.

    - **Justificação**: Picos elevados no tráfego de *uploads* podem indicar a exfiltração de dados ou ‘uploads’ massivos, o que pode ser uma atividade maliciosa, como o roubo de dados ou a transmissão de informação sensível para fora da rede.
3. **Monitorização de IPs Específicos**
    - **Regra**: Manter um monitoramento constante do IP 82.155.123.113 devido ao seu comportamento anómalo e picos de tráfego.
    - **Justificação**: O IP 82.155.123.113 mostrou padrões de tráfego anómalos, incluindo picos significativos de dados. A monitorização contínua ajuda a identificar atividades incomuns e a responder rapidamente a possíveis incidentes de segurança.
4. **Alertas de Alta Frequência de Pacotes**
    - **Regra**: Configurar alertas para atividades de alta frequência, como inúmeros pacotes enviados num curto intervalo de tempo.
    - **Justificação**: Um elevado número de pacotes num curto período pode indicar ataques de negação de serviço (DoS) ou outras atividades maliciosas que tentam sobrecarregar a rede ou comprometer a segurança.

Dispositivos comprometidos (Por completar)

A lista seguinte apresenta os dispositivos e o número de regras violadas correspondente:

IP	Regra 1	Regra 2	Regra 3	Regra 4	#
192.168.103.185				X	1
192.168.103.169				X	1
192.168.103.125	X		X		2
192.168.103.90	X		X		2
192.168.103.85		X			1
192.168.103.84				X	1
192.168.103.69		X			1

Pode-se então concluir que os dispositivos com os IPs estão de facto comprometidos.

Conclusão

Com base nos dados analisados, foi possível identificar padrões típicos de comportamento dentro da rede. Estes padrões incluem uma predominância de tráfego TCP e HTTPS, com um rácio *download*/*upload* que favorece o download. Estas observações são consistentes com o uso regular de uma rede corporativa.

Os principais objetivos deste trabalho foram atingidos com sucesso. Conseguimos identificar os servidores e serviços internos principais, bem como descrever e quantificar as trocas de tráfego entre utilizadores internos e externos. A análise detalhada dos comportamentos não anómalos permitiu a definição de regras SIEM fundamentadas, essenciais para a deteção de atividades anómalas.

No entanto, há áreas que poderiam ser melhoradas. A análise de dados poderia ser aprofundada incluindo mais parâmetros, como a duração das conexões e o comportamento ao longo do tempo, o que proporcionaria uma visão ainda mais detalhada dos padrões de tráfego. Além disso, a automação de algumas análises tornaria o processo mais eficiente e menos suscetível a erros humanos.

Em suma, este trabalho proporcionou um aprendizado significativo sobre a análise de tráfego de rede e a implementação de sistemas de deteção de intrusões. Adquirimos habilidades importantes na análise de grandes volumes de dados de tráfego de rede para identificar padrões e anomalias, bem como na definição e implementação de regras SIEM eficazes baseadas em dados concretos. Estes conhecimentos e habilidades adquiridos são fundamentais para futuras análises e implementações em ambientes reais, proporcionando uma base sólida para a prática da segurança em redes de comunicação.

Lista da Ana

Isto não dá para fazer:

- saber se o atacante está dentro ou fora da rede(a usar tunnels);

Determinar Download e upload:

- ver quais os utilizadores enviam tamanhos grandes de informação para fora da rede;
- duração de ‘download’ e ‘upload’ são importantes;
- Broswer != Https, obter o ratio de Upload < Download;
- Maioritáriamente temos mais ‘downloads’ que ‘uploads’ por utilizador;

Dados:

- média do número de conexões, por hora, por dia e criar assim o modelo;
- Anomalias de transferência de dados, o tipo de anomalia;

DNS:

- DNS/NON-DNS, ver domínios estranhos e logs;
- DNS de https, verificar comportamentos, se temos DNS cifrado que não o nosso, então ao histórico dos logs e ver;

Regras SIEM