MAC6958
Tópicos Avançados em Ciência de Dados para Redes de Computadores

- **Prof. Daniel Macêdo Batista**
  Prof. Roberto Hirata
- **DCC-IME-USP**

- 14/4/2020

Roteiro

- **Tráfego de telefonia celular** (Lucas e Giovana)
- **Intrusão em redes de Internet das Coisas** (Gustavo e Kétly)
- **Ataques de XSS** (Caio Lente)
- **Ataques de negação de serviço distribuídos** (Alan e Caio Martinelli)

Tráfego de telefonia celular

- Lucas e Giovana

  “The dataset describes in two files (traffic and topology) the hourly traffic per Base Station in some city in China with 1625680 and 13296 records each. The first file (Traffic) has the number of packets trafficked and the corresponded number of users connected per hour per base station. In the second file (topology) it is described for each Base Station what is the Latitude and Longitude.”

- https://github.com/caesar0301/city-cellular-traffic-map
- Distribuição por dia da semana costuma ser informação importante
- **O que os autores fizeram com latitude e longitude para garantir privacidade?**
- **Qual a unidade de usuários por hora no segundo gráfico (Question 2)?**
- **Sobre as BS com menos usuários, há várias, mas na saída do relatório só é mostrada a BS 5?**
- **Sobre os dados com retorno 0 (base stations com menos bytes e pacotes), como investigarão?**

Tráfego de telefonia celular

- Sobre o documento entregue
  - **Não era possível copiar e colar do .pdf**
  - **Usem identificadores para as figuras (Figura 1, Figura 2, …)**
  - **Revisem os títulos dos eixos x e y (pg. 6, pg. 8, pg. 22 , etc...)**

Tráfego de telefonia celular

- Precisam da explicação de algum conceito?
- Sobre a proposta do artigo
  - Pensaram em alguma ideia? Há sugestões de trabalhos futuros pelos autores?
  - Tem algum conhecimento sobre redes 5G? O dataset informa qual a tecnologia da malha de telefonia celular de onde os dados foram capturados?
- Duas sugestões:

1) Qual a melhor localização de novas BS? $ envolvido?
2) Como fazer alocação de recursos antecipadamente por hora do dia e dia da semana?

Intrusão em redes de Internet das Coisas

- Gustavo e Kétly

  “It contains 42 ... (pcap)... The goal of the authors were to create several types of network attacks in Internet of Things (IoT) environment ... All attacks except Mirai Botnet category are packets captured while the authors simulated attacks using tools such as Nmap. In the case of Mirai Botnet, the attack packets were generated on a laptop and then manipulated to make it appear as if it was generated by the IoT devices.”

- https://ieee-dataport.org/open-access/iot-network-intrusion-dataset
- **Quais foram as outras ferramentas usadas para gerar o tráfego?**
- **Houve anonimização dos dados? É informado o que era cada IP?**
- **Como vocês decidiram quais informações dos pcap (features) eram importantes no pré-processamento? Algo não foi usado?**
- **Que protocolo é o UDT? É usado para quê? Não há ataque com ele?**
- **“We found odd that the Denial of Service (DoS) category has less packets than MITM, since its goal is to flood the network” → Depende do ataque de DoS usado**
- **Como é a relação de categorias e subcategorias de ataques?**

Intrusão em redes de Internet das Coisas

- Sobre o documento entregue
  - **pg. 2: “Those features received the values “Normal” and “Normal”...” → Fazer uma última revisão antes da entrega**
  - **Como o acesso à Internet tem sido problemático de vez em quando, tentem sempre colocar a maior parte do contéudo no .pdf entregue**

Intrusão em redes de Internet das Coisas

- Precisam da explicação de algum conceito?
- Sobre a proposta do artigo
  - A ideia é propor algum mecanismo de aprendizado que antecipe os ataques com mais precisão, mas usando alguma especificidade de IoT?
  - Já encontraram algum outro dataset para avaliar se o atual é realista?
  - Entraram em contato com os autores?
  - Duas sugestões:

1) Avaliar o quão boa é a anonimização de datasets de IoT

2) Melhorar a antecipação de ataques em termos de precisão (desde que o dataset seja realista)

Ataques de XSS

- Caio Lente

  “...há dois arquivos com dados: xssed.csv e normal_examples.csv. Apesar de estarem salvos no formato CSV, ambos os arquivos con- tém apenas uma entrada por linha, a saber, o URL utilizado para acessar uma página da internet (sem nenhuma informação sobre o domínio do site). O primeiro CSV contém URLs de ataques XSS, en- quanto o segundo contém URLs normais.”

- https://ieee-dataport.org/open-access/detecting-xss-attacks-combining-cnn-lstm
- **“Durante a redação deste texto, ficou claro que até o motor JavaScript de conversão Markdown-LaTeX era vulnerável a ataques XSS.” → Como você gerou o documento? Há explicação no fim do documento mas não ficou claro**
- **Os URLs de ataques foram com sucesso?**
- **Os autores comentam sobre as técnicas atuais para evitar esses ataques? WAF (Web Application Firewall) por exemplo?**
- **É possível deduzir o domínio a partir dos URLs?**

Ataques de XSS

- Sobre o documento entregue
  - **Como o acesso à Internet tem sido problemático de vez em quando, tentem sempre colocar a maior parte do contéudo no .pdf entregue**

Ataques de XSS

- Precisa da explicação de algum conceito?
- Sobre a proposta do artigo
  - Os autores comentam algo sobre trabalhos futuros com base nos dados?
  - Importante pensar nos falsos positivos. São muito similares?
  - Duas sugestões:

1) Um mecanismo que detecte o ataque e não deixe a requisição chegar no processo do servidor e que consuma menos tempo do que um WAF
2) Similar ao anterior mas que tenha melhor precisão

Ataques de negação de serviço distribuídos

- Alan e Caio Martinelli

  “Esse dataset foi gerado conforme a descrição deste paper que buscava trazer uma nova forma de classificar ataques DDoS e gerar um dataset moderno e descritivo desse tipo de ataque. Toda a motivação partiu da inexistência de um dataset satisfatório.”

- https://www.unb.ca/cic/datasets/ddos-2019.html
- Dependência do software CICFlowMeter que converte pcap → csv por fluxo (quais parâmetros? **Porque é complexo? Outras ferramentas similares? Conversem com Gustavo e Kétly**)
- Tentar reproduzir com outro dataset (não há – todos muito antigos)
- Entrar em contato com os autores (A resposta não ajudou. Apontaram um README)
- **Análise do realismo do tráfego?**

Ataques de negação de serviço distribuídos

- Sobre o documento entregue:
  - Contaminação dos fluxos com mais de um ataque por CSV (difícil saber quando começa e quando termina os ataques)
  - **O que seria o ataque LDAP que foi apresentado na entrega? É um ataque DDoS contra um servidor LDAP? Seria bom descrever**
  - **Seria bom descrever o que são as features no .pdf que foi entregue. O que é Fwd Packet Lenght Min por exemplo?**
  - **Sobre propor um modelo online, o tcpreplay poderia ser usado para 'regerar' o tráfego**

Ataques de negação de serviço distribuídos

- Precisam da explicação de algum conceito?
- Sobre a proposta do artigo, sugestão de ordem de preferência (o primeiro abaixo seria o melhor)
  1) Explorar desbalanceamento das classes (importante pois no mundo real haverá a "contaminação")
  2) Novas features (importante pois isso estaria sendo feito com um dataset recente)
- Evitaria:
  1) Algoritmos mais modernos (fica parecendo mais do mesmo)
  2) Modelo online (precisaria de muito recurso computacional e de tempo)