THE DEVELOPER'S CONFERENCE

Trilha: Inteligência artificial e machine learning

Machine Learning e *Internet* das coisas: privacidade e conveniência

Antônio Janael Pinheiro

Agenda

1. Introdução
2. Privacidade
3. Internet das Coisas
4. Análise de tráfego criptografado
5. Implementação
6. Avaliação
7. Conclusão

Quem sou eu?

- Doutorando em Ciência da Computação pelo Centro de Informática da UFPE;
- Pesquisador em privacidade para IoT;
- Entusiasta do aprendizado de máquina.

Uma questão importante

- É possível identificar dispositivos IoT a partir do tráfego de rede?

Privacidade

- "Privacidade é a reivindicação dos indivíduos em determinar por si mesmos quando, como e em que medida as suas informações são comunicadas a terceiros."(Alan Westin)

Autonomia e liberdade

Anonimização

Intimidade

Internet das Coisas que te vigiam

- Microfones, câmeras e todo tipo de sensor;
- Dispositivos IoT restritos computacionalmente;
- Fabricantes negligentes.


BIG BROTHER
IS WATCHING YOU

Mais dúvidas surgem

- Como identificar dispositivos e eventos IoT sem violar a privacidade dos indivíduos?

Tráfego de rede criptografado

- A criptografia protege o conteúdo do tráfego;
- Variações nos tipos de dados refletem nos metadados;
- A criptografia é insuficiente para ocultar metadados.

Side-channel information

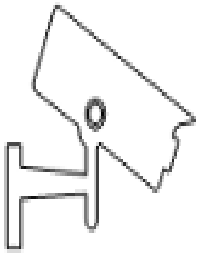
- Tamanho do pacote;
- Intervalo entre *frames*;
- Duração do fluxo de dados;
- Sentido do tráfego.

Racional

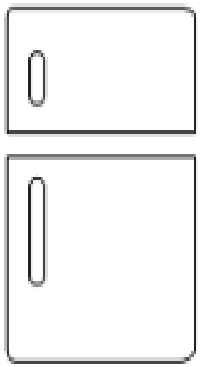
- Diferentes aplicações produzem pacotes com tamanhos distintos;
- Implementações distintas de uma aplicação geram padrões diferentes;
- Tipicamente, dispositivos IoT têm um propósito específico.












Machine Learning e classificação de tráfego

- Identificar padrões expostos por metadados;
- Empregar atributos do tráfego como *features* para a construção de modelos;
- Construir um perfil do comportamento normal de um dispositivo IoT.

Implementação – atributos do tráfego IoT

- Janelas de um segundo:
  - Tamanho médio;
  - Desvio padrão;
  - Número de bytes.

Implementação – ferramentas

- *Python* 3;
- *Scikit-learn*;
- *Pandas*;
- *Numpy*;
- Classificadores:
  - *k-Nearest Neighbors*;
  - *Decision Tree*;
  - *Random Forest*;
  - *Support Vector Machine*.

Avaliação

- 24 dispositivos IoT;
- *Stratified* 10-*fold cross-validation*;
- *Chronological* 10-*fold cross-validation*.

Avaliação – métricas

- Acurácia;
- Precisão;
- *Recall*;
- F1-*score*;
- *Specificity*;
- *Geometric mean*.

Avaliação – oportunidades

- Distinguir entre dispositivos IoT e não IoT;
- Identificar dispositivos específicos;
- Determinar os eventos que produziram o tráfego.

Resultados – alguns números...

- Resultados com o *Random Forest*:
  - Precisão de 99% na distinção entre dispositivos IoT e não IoT;
  - Acurácia de 96% na identificação de dispositivos IoT;
  - Acurácia de 99% na identificação de eventos IoT.

Resultados – possibilidades

- Identificar comandos de voz do *Amazon Echo Dot*;
- Determinar a presença de indivíduos em suas residências;
- Identificar dispositivos ligados/desligados;
- Detectar movimentos no campo de visão de uma câmera de segurança.

Potenciais aplicações

- Isolamento de dispositivos comprometidos por *malwares*;
- Priorização dos dados de determinados equipamentos;
- Identificação de comportamentos atípicos.

Uma faca de dois gumes

- Criptografia é insuficiente para proteger a privacidade dos indivíduos;
- Identificação de dispositivos e eventos IoT a partir do tráfego criptografado;
- Contramedida: ofuscação de tráfego.

Vamos conversar?

- LinkedIn: linkedin.com/in/janael-pinheiro/
- *WhatsApp* e *Telegram*: (81) 995897449
- E-mail: ajp@cin.ufpe.br;

Agradecimentos

Agradeço ao prof. Divanilson Campelo, Jeandro Bezerra e Caio Burgardt, co-autores do artigo "Identifying IoT devices and events based on packet length from encrypted traffic".

Conclusão

- Privacidade é potencialmente um direito natural dos indivíduos;
- A IoT tem potencial para destravar valor em vários setores da economia;
- As ameaças à privacidade introduzidas pela IoT são imensas;
- É possível prover conveniência para os usuários IoT sem violar a sua privacidade.

Obrigado!

Questões?