Smart Spying via Deep Learning: Inferring Your Activities from Encrypted Wireless Traffic

Tao Hou[†], Tao Wang[‡], Zhuo Lu[†] and Yao Liu[†]

[†]University of South Florida
[‡]New Mexico State University

UNIVERSITY OF SOUTH FLORIDA

Outline

- Motivation
- System Design
- Evaluation
- Conclusion

Outline

- Motivation
- System Design
- Evaluation
- Conclusion

Wireless is Ubiquitous

Wireless Technologies

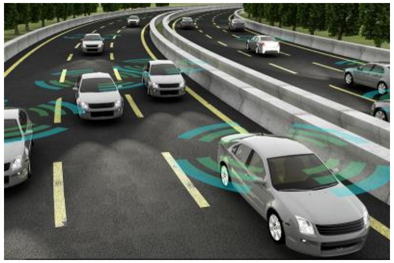
4G Lte 5G WiFi Bluetooth wimax ZigBee RFID NFC

Wireless Devices



Applications





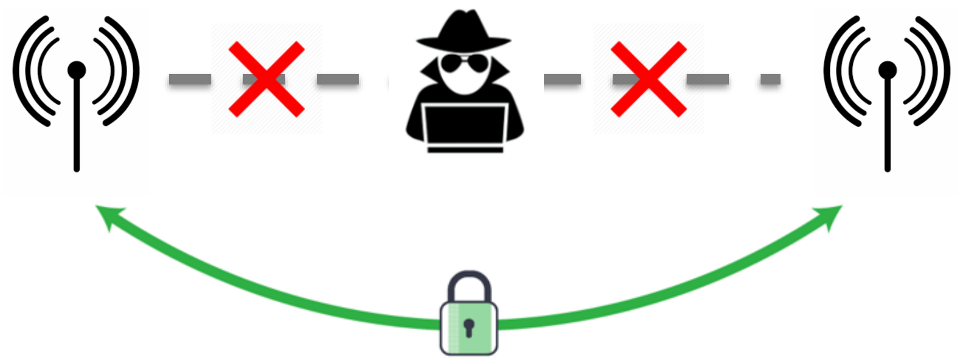
Wireless Eavesdropping

- Open nature of wireless medium

- Wireless eavesdropping

Wireless Encryption

- Anti-Eavesdropping



- Is it still available to spy user activities?

Existing Methods

- Traffic analysis on statistic patterns
  - APP usages
  - spoken phrases
  - behaviors

- Limitations:
  - low accuracy
  - specific domains

Outline

- Motivation
- System Design
- Evaluation
- Conclusion

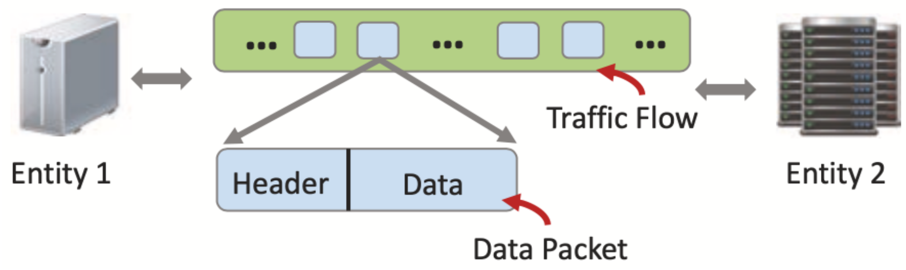
Main idea

Design a smart spying strategy: SS-Infer

1. Improve the data representativeness
   - statistic data
   - encrypted raw data
2. Develop a fusion Deep Neural Network model
   - Convolutional Neural Network (CNN) : spatial features
   - Long Short-Term Memory (LSTM) : temporal features
   - Extract flow features from network traffic

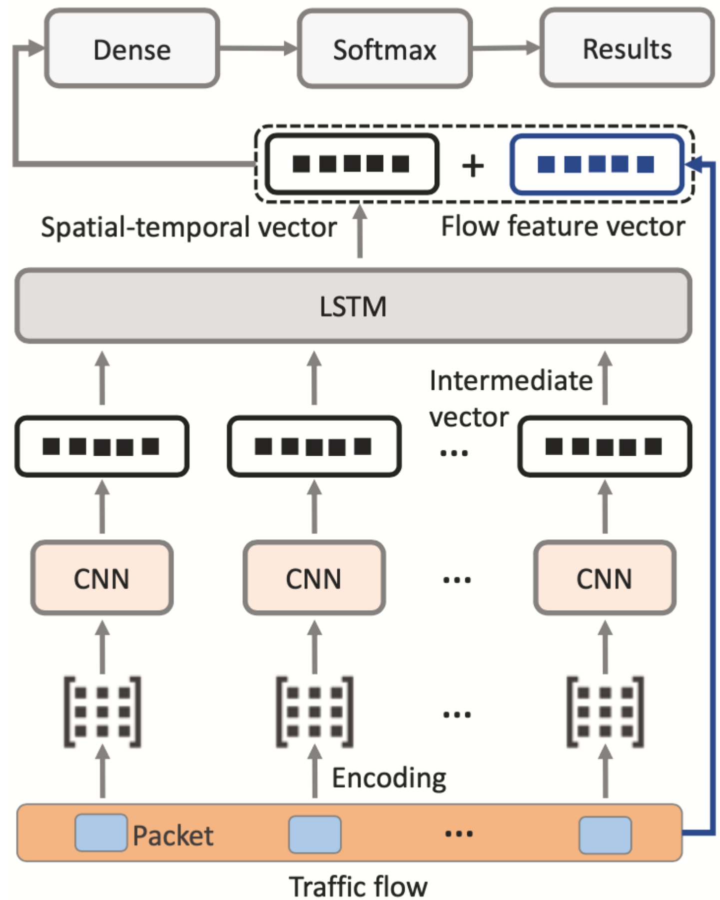
Traffic Flow

- Infer activities in the level of traffic flows



Intercepted packets will be aggregated and grouped into traffic flows

System Architecture



Extract flow features from network traffic

Learn spatial-temporal dependencies by CNN and LSTM

Spatial-temporal Features

- One-Hot Encoding (OHE)

- Learn spatial dependencies through CNN

Convolution: $$\mathbf{m}_i = f(\mathbf{w} \cdot \mathbf{c}_{i:i+s-1} + \mathbf{b})$$

Pooling: $$\hat{\mathbf{m}} = max\{\mathbf{m}\}.$$

$$\{\hat{\mathbf{m}}_1, \hat{\mathbf{m}}_2, ..., \hat{\mathbf{m}}_p\}$$

- Learn temporal dependencies through LSTM

$$\{\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, ..., \hat{\mathbf{y}}_p\}$$

Flow based Features

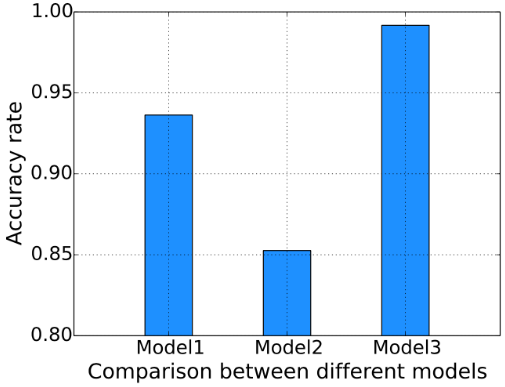
Category	Feature	Description
Header Information	Source Port	Port number at source
	Destination Port	Port number at destination
	Source Address	IP address of source
	Destination Address	IP address of destination
Statistical Information	Forward Inter-arrival Time	(mean, min, max, std) Inter-arrival time for forward packets in a traffic flow
	Backward Inter-arrival Time	(mean, min, max, std) Inter-arrival time for backward packets in a traffic flow
	Packet Length	(mean, min, max, std) Number of bytes for packets in a traffic flow
	Active Time	The time a flow was active
	Idle Time	The time a flow was idle
	Out of Order Count	The total number of packets that arrive destination out of order in a traffic flow
	Bytes per Second	The number of bytes transmitted per second in a traffic flow
	Packets per Second	The number of packets transmitted per second in a traffic flow

Outline

- Motivation
- System Design
- Evaluation
- Conclusion

Evaluation

- UNB ISCX network traffic dataset
  - Web Browsing, Email, Chat, Streaming, File Transfer, VoIP, P2P
- Evaluation results:



Number	0	5	10	15	21
Accuracy (%)	93.36	98.53	99.10	99.15	99.17
Time	1.00	1.02	1.05	1.10	1.18

Outline

- Motivation
- System Design
- Evaluation
- Conclusion

Conclusion

We design a smart spying strategy, named SS-Infer, which can accurately and efficiently infer a user's activity from encrypted wireless traffic.

THANK YOU

Any Questions