I Know Your Activities Even When Data Is Encrypted: Smart Traffic Analysis via Fusion Deep Neural Network

Tao Hou†, Tao Wang‡, Zhuo Lu† and Yao Liu†
†University of South Florida, Tampa, FL, USA, {taohou@mail., zhuolu@, yliu@cse.}usf.edu
‡New Mexico State University, Las Cruces, NM, USA, taow@nmsu.edu

***Abstract*—Network transmissions are vulnerable to Man-In-The-Middle (MITM) attacks. Through decoding intercepted data, attackers may infer victims' sensitive activities or even steal their private information (e.g., password). Though transmitted data can be encrypted against eavesdropping, attackers can still infer user activities via traffic analysis. Nevertheless, previous inference methods usually have the limitation that they can only achieve a relatively high accuracy in a specific domain (e.g., app usages, spoken phrases, motion and behaviors). In this research, we propose a smart traffic analysis strategy to overcome this limitation. By developing a fusion deep neural network, our design can infer a user's activities of multiple domains with a higher accuracy. We also implement a prototype tool on top of this design to conduct experiments. The preliminary evaluation results show our strategy works effectively in activity inference on encrypted data, with an accuracy rate as high as 99.17%.**

I. Introduction

Through providing a convenient and express way to connect to the world, smart devices (e.g., computers, smart phones, IoT devices, etc) nowadays are ubiquitous in our daily life. But their connections via wired or wireless networks are potentially vulnerable to Man-In-The-Middle (MITM) attacks [1]–[3]. Once the transmitted data is obtained by an attacker, it can further decode the message to infer the user's sensitive activities (e.g., voice or video chatting) or even steal the user's private information (e.g., password and personal information), which is carried in the eavesdropped data.

A simple yet efficient method to prevent such information leakage is to encrypt the transmitted data [4], such that it is difficult for the eavesdropper to decode useful information. Multiple encryption schemes have been proposed to preserve the confidentiality of the data traffic during transmission, including WPA2, HTTPS, PGP, MSP and etc. However, data encryption does not stop the attacker from exploring new ways to spy on users. Through traffic analysis [5] on patterns or statistic of side-channel information, attackers can successfully infer user activities on encrypted data.

Nevertheless, these attacks usually only utilized the statistic results of features from a specific domain to perform activity inference. Consequently, they can only achieve a relatively high accuracy in a corresponding domain. In this research, we aim to overcome this limitation by proposing a smart traffic analysis strategy. The core idea is two-fold: 1) besides the statistic results, encoding the encrypted data to improve the data representativeness. In this way, our design can capture the characteristics concealed in the data payload, which are ignored by previous methods. 2) developing a fusion Deep Neural Network (DNN) model which integrates multiple traditional neural networks to improve learning abilities.

In particular, we utilize Convolutional Neural Network (CNN) to learn the spatial dependencies among the encoded data; and then adopt the Long Short-Term Memory (LSTM) to learn the temporal dependencies on the results from the first step. Finally, we combine the spatial-temporal features from previous steps with the flow features directly extracted from network traffic to improve the classification accuracy. With this proposed architecture, our preliminary evaluation results [6] show that we can achieve a classification accuracy rate as high as 99.17% when identifying a user's real-world activities.

In the future, we will collect more real-world datasets to validate the proposed smart traffic analysis strategy and therefore to further refine our design. In addition, we will build an efficient and effective traffic analysis system and conduct a comprehensive experiment evaluation.

II. System Design

A. Overview

Different network activities (e.g., chatting, streaming) indicate different behaviors in the level of traffic flows, rather than the behaviors of single packets. Usually, a traffic flow is composed of multiple data packets in transmission. As Figure 1 shows, there is a connection between Entity 1 and Entity 2, a traffic flow is transmitted through the connection. We consider an MITM attacker that can intercept data packets of different connections. According to the TCP/IP protocol, the intercepted packets will be then aggregated and grouped into traffic flows of different connections. We further assume the transmitted data is encrypted to ensure the attacker cannot infer user activities by directly decoding the intercepted traffic.

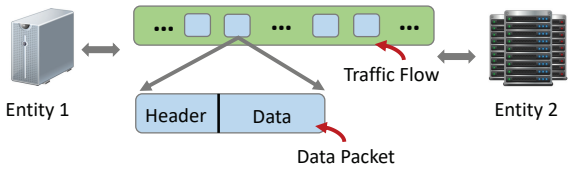

Fig. 1. Network traffic flow.

In this research, we develop a smart traffic analysis strategy to infer user activities from encrypted data. The core idea is to infer activities on top of a deep learning based traffic classifier,

which is a fusion DNN model. As Figure 2 shows, our design integrates both the internal layers of CNN and LSTM, such that it can capture not only spatial dependencies for the data in each packet, but also temporal dependencies among different packets. Finally, it makes the classification based on the combination of spatial-temporal vector and directly extracted flow feature vector.

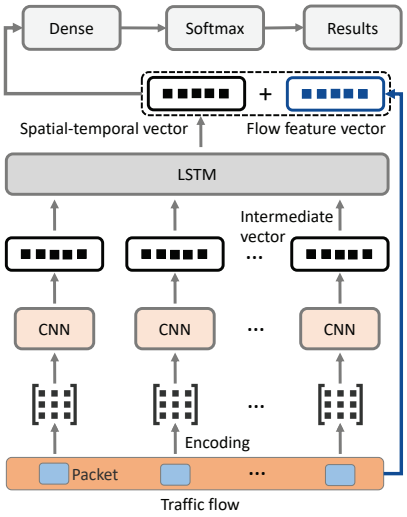

Fig. 2. System design of SS-Infer.

B. Learning Spatial-temporal Features

1) Encrypted Data Encoding: Though encrypted, the data payload in a packet still contains information indicating a user's activities. In particular, the spatial and temporal correlation relationships among different packets can both contribute to the classification for different activities. We adopt One-Hot Encoding (OHE) to represent the encrypted data, and we only take into consideration of packets with data payload larger than 300 bytes. The OHE vector is the binary code of each byte, i.e., an 8-dimensional vector.

2) Learning Spatial Dependencies through CNN: CNN is designed with the ability to learn the spatial dependencies. Here, we adopt internal layers of CNN to extract the spatial dependencies from each packet as the intermediate vector. This vector is the inputs of LSTM. Assume filter $\mathbf{w}$ works with a window size of s, $\mathbf{m}_i$ is the i-th generated feature, $\mathbf{c}_i$ is the i-th column of the data encoding matrix, $\mathbf{b}$ is a bias, and f is ReLUs. We get:

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

Then, a max-over-time pooling is applied to feature map $\mathbf{m} = [\mathbf{m}_1, \mathbf{m}_2, ..., \mathbf{m}_{300-h+1}]$ to get the intermediate vector:

$$\hat{\mathbf{m}} = max\{\mathbf{m}\}. \tag{2}$$

3) Learning Temporal Dependencies through LSTM: LSTM is suitable for learning temporal features, especially for the long-term temporal dependencies. We apply internal layers of LSTM after the convolution layers to learn the spatial-temporal dependencies. It takes the intermediate vector of each packet in order as inputs. Assume there are p valid packets in a flow, the input sequence is therefore denoted as $\{\hat{\mathbf{m}}_1, \hat{\mathbf{m}}_2, ..., \hat{\mathbf{m}}_p\}$. Through a series of transitions by a set of adaptive multiplicative gates in these internal layers, we get the output $\{\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, ..., \hat{\mathbf{y}}_p\}$, which is the spatial-temporal vector.

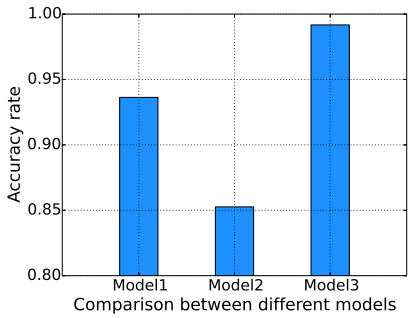

Fig. 3. Classification performance.

III. Evaluation

In our experiments, we use a high-performance workstation with four NVIDIA GeForce RTX 2080 GPUs to perform traffic classification on top of TensorFlow. UNB ISCX Network Traffic Dataset [7] is used for evaluation. It is captured from real world networks and the traffic flows are labeled manually as ground truth. Meanwhile, the dataset also contains the original encrypted data. We consider three models: Model1 is a neural network only using the spatial-temporal features for classification; Model2 only using the flow features for classification, and Model3 combines the spatial-temporal features and the flow features, which is adopted in our design. Figure 3 shows the evaluation results. We can see that when only using spatial-temporal features or flow features to infer a user's activities, the accuracy rates are 93.63% and 85.26% for Model1 and Model2, respectively. Though the accuracy rate is already relatively high for activity inference, our design (i.e. Model3) can achieve a more accurate result, with an accuracy rate being as high as 99.17%.

References

[1] Yuanyu Zhang, Yulong Shen, Hua Wang, Jianming Yong, and Xiaohong Jiang. On secure wireless communications for iot under eavesdropper collusion. *IEEE Trans. on Automation Science and Engineering*, 2015.

[2] Tao Wang, Yao Liu, Qingqi Pei, and Tao Hou. Location-restricted services access control leveraging pinpoint waveforming. In *ACM CCS*, 2015.

[3] Tao Wang, Yao Liu, Tao Hou, Qingqi Pei, and Song Fang. Signal entanglement based pinpoint waveforming for location-restricted service access control. *IEEE Trans. on Dependable and Secure Computing*, 2016.

[4] William Stallings. *Network and internetwork security: principles and practice*, volume 1. Prentice Hall Englewood Cliffs, NJ, 1995.

[5] Mauro Conti, Qian Qian Li, Alberto Maragno, and Riccardo Spolaor. The dark side (-channel) of mobile devices: A survey on network traffic analysis. *IEEE Communications Surveys & Tutorials*, 2018.

[6] Tao Hou, Tao Wang, Zhuo Lu, and Yao Liu. Smart spying via deep learning: inferring your activities from encrypted wireless traffic. In *IEEE GlobalSIP*, 2019.

[7] Gerard Draper-Gil, Arash Habibi Lashkari, Mohammad Saiful Islam Mamun, and Ali A Ghorbani. Characterization of encrypted and vpn traffic using time-related features. In *ICISSP*, 2016.