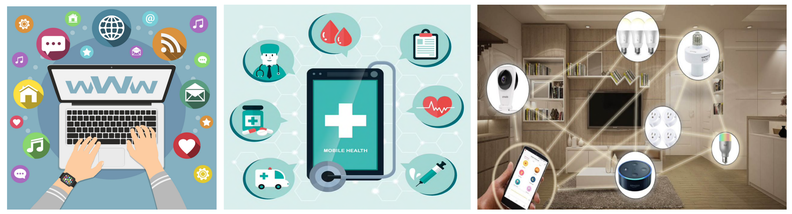
I Know Your Activities Even When Data Is Encrypted: Smart Traffic Analysis via Fusion Deep Neural Network

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

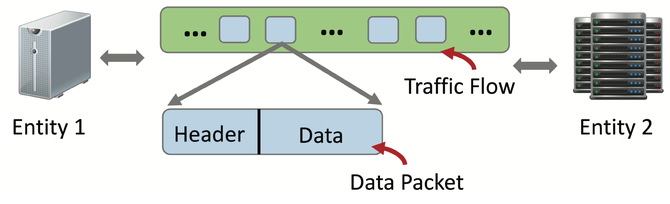
†University of South Florida ‡New Mexico State University

MOTIVATION

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 MITM attacks. A simple yet efficient method to prevent such information leakage is to encrypt the transmitted data, such that it is difficult for the eavesdropper to decode useful information.



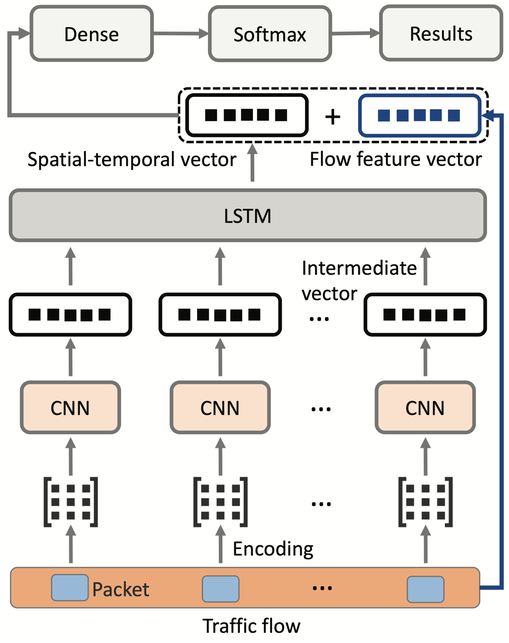
In this research, we aim to overcome this limitation by proposing a smart traffic analysis strategy. The core idea is two-fold: 1) besides statistic results, encoding the encrypted data to improve the data representativeness. In this way, our design can capture the characteristics concealed in the data payload. 2) developing a fusion Deep Neural Network model which integrates multiple traditional neural networks to improve learning abilities.



This work was supported in part by NSF CNS-1553304 and CNS-1717969.

SYSTEM DESIGN

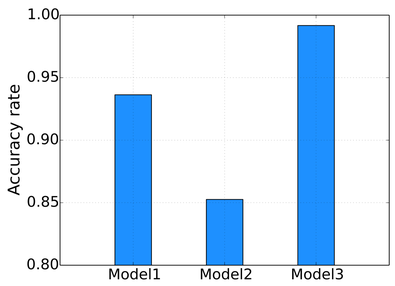
Different network activities (e.g., chatting, streaming) indicate different behaviors in the level of traffic flows. Usually, a traffic flow is composed of multiple data packets in transmission. We consider an MITM attacker that can intercept data packets of different connections. According to the TCP/IP protocol, the intercepted packets will be then aggregated and grouped into traffic flows of different connections.



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.

RESULTS

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 is used for evaluation. 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%.



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

CONCLUSION

Network transmissions are vulnerable to MITM attacks. Though transmitted data can be encrypted against eavesdropping, attackers can still infer user activities. In this research, we propose a smart traffic analysis strategy to achieve this goal. By developing a fusion deep neural network, our design can infer a user's activities with a high accuracy. The preliminary evaluation results show our strategy works effectively in activity inference on encrypted data, with an accuracy rate as high as 99.17%.