Poster: Feasibility of Malware Traffic Analysis through TLS-Encrypted Flow Visualization

Dongeon Kim, Jihun Han, Jinwoo Lee, and Heejun Roh
Cyber Security Major, Division of Applied Mathematical Sciences
Korea University Sejong Campus
Sejong, Republic of Korea
{dekim99, smallstone00, jwoo619, hjroh}@korea.ac.kr

Wonjun Lee
School of Cybersecurity
Korea University
Seoul, Republic of Korea
wlee@korea.ac.kr

***Abstract*—With the wide adoption of TLS, malware's use of TLS is also growing fast. However, fine-grained feature selection in existing approaches is too burdensome. To this end, we propose to visualize TLS-encrypted flow metadata as an image for better malware traffic analysis and classification. We discuss its feasibility and show some preliminary classification results with high accuracy.**

***Index Terms*—Transport Layer Security (TLS), Malware, Malware Family, TLS Flow Metadata, Visualization**

I. Introduction

With the increasing demand of end-to-end security, Transport Layer Security (TLS) is being widely adopted for network traffic encryption. According to Google's Transparency Report [1], for instance, more than 95% of traffic across Google is encrypted. Similarly, Zscaler reports that 80% of enterprise traffic on the Zscaler cloud in 2018 is SSL/TLS-encrypted [2]. Unfortunately, recent studies [3], [4] show that malware's use of TLS is also rapidly increasing. This trend of TLS-encrypted flows is problematic in threat detection, since traditional approaches such as deep packet inspection and signatures have limited effectiveness for encrypted traffic [4].

To this end, there are several research efforts [4]–[6] on malware traffic detection/classification have been conducted. Anderson and McGrew [4] observes that features from TLS handshake metadata and related but unencrypted DNS/HTTP flows (so-called *contextual flows*) can be utilized to classify TLS-encrypted malicious flows accurately, in a binary manner. Furthermore, another work from the same authors [5] shows that $l1$-logistic classifiers with carefully chosen TLS flow features (*e.g.*, TLS flow metadata, the sequences of packet lengths and inter-arrival times, distribution of bytes, and client/server information) can perform malware family classification with 90.3% of 10-fold accuracy. In addition, [6] revisits the previous approach with incremental learning algorithms, considering more practical scenarios.

978-1-7281-6992-7/20/$31.00 ©2020 IEEE

However, these approaches require fine-grained feature selection conducted by experts. In [4], for example, defining contextual flows for TLS flows requires detailed understanding of several Internet standards. Furthermore, the above approaches need to conduct field-specific preprocessing for message field values. That is, understanding each field is essential to enable the approaches.

Instead, we propose a coarse-grained feature selection approach for TLS flow metadata called *TLS-encrypted flow visualization*, useful for malware traffic analysis. Our approach is inspired by malware image visualization [7] widely used for analysis and classification of malware binaries. Our approach has several advantages:

- We utilize the common lesson from [4]–[6]: *TLS flow metadata which is exchanged at the beginning of TLS flow without encryption have fruitful information to classify encrypted malware traffic*. It implies that we can perform early classification of malware family without monitoring the whole flow.
- We assign different color to each message type in TLS flow metadata. To this end, different messages of a flow can be easily observed as a colored image.
- Similar to the discussion in [7], images can capture small changes yet retain the global message exchange pattern. It implies that malware variants in the same family could have similarity as images, and distinct patterns would be observed among different malware families.

II. TLS-Encrypted Flow Visualization

In our approach, we generate TLS flow metadata images from several messages at the beginning of TLS connections, by regarding each byte as a pixel. In offline setting with `pcap` files containing sniffed packets in a network, the messages can be extracted with `tshark`, as an example. Our image format is shown in Figure 1. We colored the messages to distinguish its type for better visualization and added padding to cope with different message sizes. Note that for each pixel, the closer to `0xff`, the lighter it is.

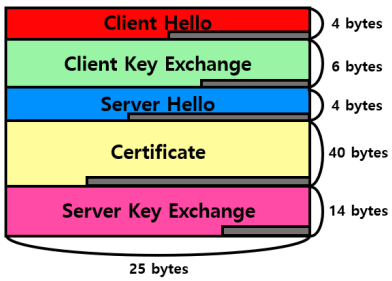


Fig. 1: TLS flow metadata image format.

III. Malware Traffic Analysis via Images

To show the feasibility of our approach in malware traffic analysis, we analyze images of several malware families (Dridex, Gootkit, Hancitor, IcedID, and Trickbot) which are generated from `pcap` files offered by [8]. As shown in Figure 2, we observe that different malware families have distinct image patterns.

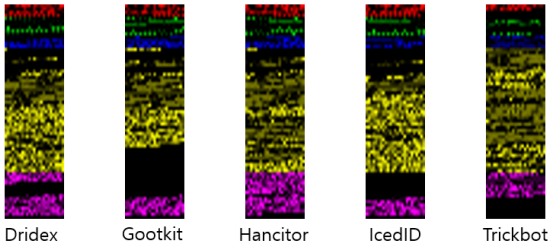
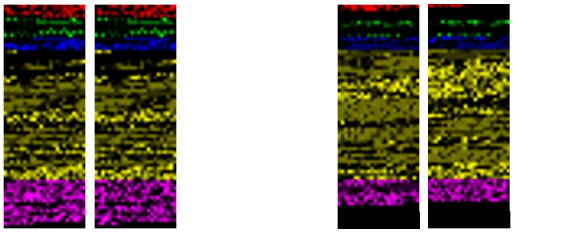


Fig. 2: Representative images of malware belonging to different malware families.

From Figure 2, one may think that a distance measure such as Euclidean distance could be applied to the images for malware family classification. Actually, Figure 3a shows some similarity in different samples in the same malware family. However, unfortunately, a counterexample in Figure 3b is available, each of which tries to access to different server website.

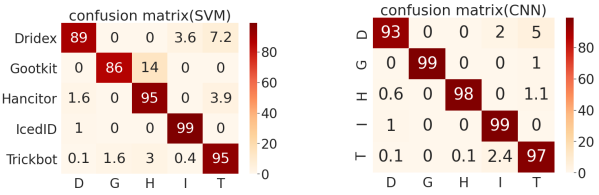
IV. Malware Family Classification via Images

We also conduct malware family classification using the images, without careful selection of TLS metadata



(a) Hancitor Sample Images (b) Trickbot Sample Images

Fig. 3: Similarity and dissimilarity in the same family.



(a) SVM (b) CNN

Fig. 4: The resulting confusion matrices.

features. Two machine learning classifiers, a Convolutional Neural Network (CNN) model and a Support Vector Machine (SVM) model, are applied using the `pcap` dataset [8] with 1,358 TLS flows, where 70% of the flows are chosen for training. Figure 4 shows the resulting confusion matrices. Surprisingly, the CNN classifier shows 97% accuracy and the SVM classifier shows 93% accuracy in total.

V. Conclusion

We discussed the feasibility of malware traffic analysis through TLS-encrypted flow visualization. We observed that different malware families have different patterns in their images but another kind of differences could be available in the same malware family. Surprisingly, nevertheless, malware family classification via images using SVM and CNN models has high accuracy. In future work, we will further discuss why such high accuracy could be possible.

Acknowledgment

This research was supported by Korea Institute of Science and Technology Information (KISTI). Prof. Heejun Roh is the corresponding author.

References

[1] Google. HTTPS encryption on the web. [Online]. Available: https://transparencyreport.google.com/https/overview

[2] Zscaler. SSL-based threats — security report. [Online]. Available: https://info.zscaler.com/whitepaper-ssl-traffic-threats

[3] L. Nagy. Nearly a quarter of malware now communicates using TLS. [Online]. Available: https://news.sophos.com/en-us/2020/02/18/nearly-a-quarter-of-malware-now-communicates-using-tls

[4] B. Anderson and D. McGrew, "Identifying encrypted malware traffic with contextual flow data," in *Proc. of AISec'16*, Vienna, Austria, October 2016.

[5] B. Anderson, S. Paul, and D. McGrew, "Deciphering malware's use of TLS (without decryption)," *Journal of Computer Virology and Hacking Techniques*, vol. 14, no. 3, pp. 195–211, August 2018.

[6] I. Lee, H. Roh, and W. Lee, "Encrypted malware traffic detection using incremental learning," in *Proc. of IEEE INFOCOM'20 Poster Session*, July 2020.

[7] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, "Malware images: Visualization and automatic classification," in *Proc. of IEEE VizSec'11*, Pittsburgh, PA, USA, July 2011.

[8] B. Duncan. Malware traffic analysis. [Online]. Available: http:/malware-traffic-analysis.net/