Cross-layer Device Identification for Smart Grid Substation Networks

IEEE CNS 23 Poster

Venkat Sai Suman Lamba Karanam, Fahmida Afrin, Byrav Ramamurthy and Nirnimesh Ghose
School of Computing
University of Nebraska-Lincoln
Lincoln, NE, USA
saisuman@huskers.unl.edu, fafrin2@huskers.unl.edu, ramamurthy@unl.edu and nghose@unl.edu

***Abstract*—Smart grid (SG) substations are responsible for the grid's smooth operation. SG substations consist of several interacting components that communicate over the network. Attacks on the SG substation communication can lead to dire consequences like loss of power or even worse. Device fingerprinting can help identify malicious devices and communications. Although device fingerprinting has been studied for cyber-physical systems, mostly in wireless scenarios, there is little to no work for SG substation networks. We develop a device fingerprinting framework for SG substations using a cross-layer approach. We specifically developed our approach for the IEC68150 standard, a commonly used communication standard in SG substations. We took a cross-layer device fingerprinting approach with three models- the Link layer model, Transport layer model and the stacking-based ensemble cross-layer model using logistic regression. We analyzed the accuracy in device identification of our cross-layer approach on a real world SG substation-like dataset *4SICS*. The results show that our cross-layer approach is a feasible option to fingerprint SG devices. To the best of our knowledge, our work is the first device fingerprinting work done on SG substation networks.**

I. Introduction

*a) **Motivation**:* Smart Grid (SG) substations are the crux of the SG networks and are responsible for energy transmission and distribution. SG substation networks employ the IEC68150 standard [1] that defines the communication protocols and thus enables automation and control of operations (see Fig. 1). Securing SG substation communication from malicious attacks is important because the consequences can range from minor to catastrophic (like large scale power failure).

Although IEC68150 standard-based protocols employ authentication, the communication is vulnerable to attacks similar to most traditional authentication (spoofing, eavesdropping, device compromise, power analysis, etc.). Device fingerprinting techniques can identify devices based on their communication patterns and thus helps recognize devices exhibiting abnormal behavior. Device fingerprinting has unique advantages for smart grid scenarios because it is possible to fingerprint and identify devices continuously, even after the initial authentication. *Currently, little to no work employs device fingerprinting in SG communications, especially for substation communication. Designing device identification techniques via fingerprinting specifically for SGs is much needed to secure SG networks from future attacks.*

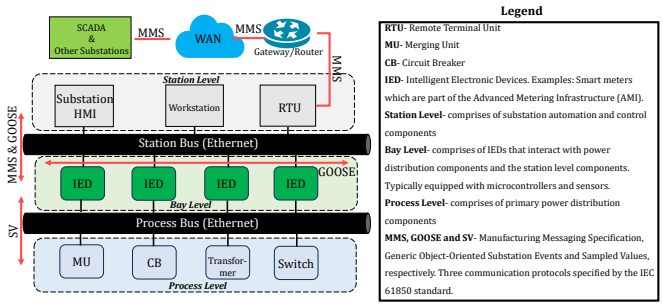

Fig. 1: IEC68150 standard for smart grid substation automation and control.

*b) **Limitations in Current Studies**:* Existing works have two severe limitations. (i) Most fingerprinting techniques were developed for cyber-physical systems (and IoTs), and they mostly considered RF/wireless mediums at the physical layer. Merely grouping SGs into cyber-physical systems ignores the complexities inherent in SG data communications. For example, these works do not consider the substation networks that form the core of SG data communication. The link layer offers rich information that can be exploited for fingerprinting but is rarely given attention. (ii) Device fingerprinting using Transport layer statistics such as the packet inter-arrival time (IAT), the probability distribution of packet header types, etc., have shown promising results. Link layer device fingerprinting can be further improved using Transport layer awareness, i.e., using a cross-layer device fingerprinting technique that combines Link and Transport layers. Unfortunately, no work exists to fingerprint devices on smart grids using Transport layer information. Consequently, no work exists for smart grid device fingerprinting using cross-layer techniques.

*c) **Our Idea**:* We propose a cross-layer device fingerprinting approach for smart grids that applies separate deep learning techniques on Ethernet and Transport layers (see

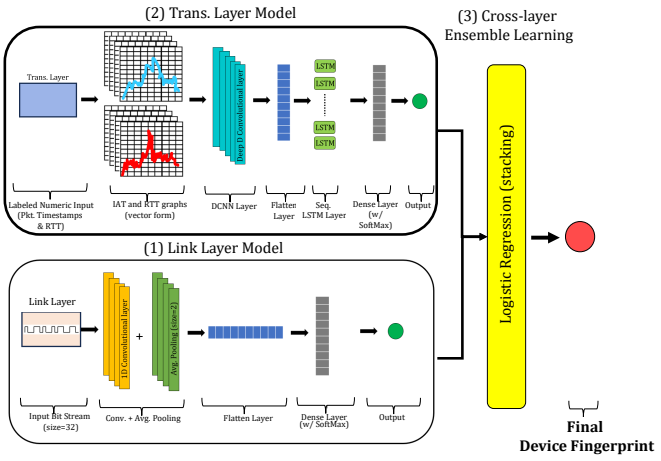

Fig. 2: Visual illustration of our cross-layer framework to fingerprint devices in a smart grid substation network.

Fig. 2). (1) The Link layer model is a convolutional and densely connected neural network with vector embeddings. The input to the first layer is the vectors of binary streams representing the Ethernet frames. This first layer converts the input to dense vectors and feeds it to the convolutional layers with *ReLU activation*, followed by the pooling layers and then the dense layer with *SoftMax activation*. (2) The Transport layer model is a deep convolutional neural network plus LSTM technique (DCNN+LSTM) that learns from the inter-arrival time (IAT) and round trip times (RTT) at the Transport layer. First, the IAT and RTT graphs are generated from the measurements made at the Transport layer. IAT and RTT of each device depend on the underlying hardware, such as the CPU, clock frequency, and cache configuration, and the circuitry imperfections compound this during the manufacturing process. The IAT and RTT graphs in vector forms are fed separately to the DCNN as a second step. The DCNN subnetwork consists of 4 *(2D convolutional layer + MaxPooling layer)* with *ReLU activation function*. The DCNN subnetwork output is flattened and then fed to a sequential LSTM network. The encoded output of the LSTM network is then fed to a fully connected (*dense*) layer with *ReLU activation function* followed by another fully connected (*dense*) layer with *SoftMax activation function*. (3) The Link and Transport layer models are then stacked using a logistic regression model to create our cross-layer device fingerprinting model as an ensemble.

II. Evaluation

We evaluate our framework on a real world publicly available dataset representative of an SG substation network, the *4SICS* dataset from the "Geek Lounge Lab" [2], which contains network traffic data collected from industrial equipment. It contains traffic exchanged between Remote Terminal Units (RTUs), controllers, and other components found in a smart grid network. The *4SICS* dataset contains three sub-datasets that were collected on three different days and times. First we combined the sub datasets and then split it into 70% training+validation and 30% testing sets, respectively. The accuracy results of our model on the *4SICS* dataset are presented in Table I. We defer the cross-day training/testing of our model using the sub-datasets in *4SICS* for the future. The hyper parameters for the Link, Transport and Cross-layer models were chosen through trial and error, no through cross validation (see Table II).

Model	Training	Validation	Testing
Link Layer Model	63	73	69
Trans. Layer Model	57	67	71
Cross-layer Model	66	72	74

TABLE I: Accuracy results (%) in device identification using each of the three sub-datasets in the *4SICS* dataset.

Model	Epochs	Learn. Rate	Batch Size	Optimizer
Link Layer Model	30	*1e-5*	64	*Adam* and *RMSprop*
Trans. Layer Model	75	*1e7*	32	*Adam*
Cross-layer Model	75	*1e5*	64	*Adam*

TABLE II: Hyper parameters that were found to work for higher accuracy in device identification using the Link layer and Transport layer models

III. Conclusion and Future Work

We demonstrated that our cross-layer approach is a feasible technique to identify devices on the SG substation networks. Our immediate future work is to improve the model through rigorous performance analysis as well as designing more efficient ensemble techniques. For the former effort we would like to explore alternate representations of the input data for each of the layers and for the latter effort we would like explore boosting approach to combine the individual models instead of the stacking approach we considered. Additionally, we will extend our framework to fingerprint devices connecting to the smart grid network outside the substation perimeters, such as the PLC modems, smart meters, and vendor-specific commercial-off-the-shelf (COTS) devices. We are building a SG sandbox to emulate a smart grid network, including the substation, AMI, and other communicating components.

Acknowledgement. This work is sponsored in part by a grant from the Nebraska Energy Sciences Research (NCESR) center in collaboration with the Nebraska Public Power District (NPPD).

References

[1] D. Baigent, M. Adamiak, R. Mackiewicz, and G. Sisco, "IEC 61850 communication networks and systems in substations: An overview for users," *SISCO Systems*, 2004.

[2] Capture Files from 4SICS Geek Lounge. Accessed August 22, 2023. https://www.netresec.com/index.ashx?page=PCAP4SICS.