Payload-based Packet Classification using Deep Learning

Michael Shell[1], Homer Simpson[2], James Kirk[3], Montgomery Scott[3], and Eldon Tyrell[4], *Fellow, IEEE*
[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA
[2]Twentieth Century Fox, Springfield, USA
[3]Starfleet Academy, San Francisco, CA 96678 USA
[4]Tyrell Inc., 123 Replicant Street, Los Angeles, CA 90210 USA

The abstract goes here.

***Index Terms*—Packet Classification, Deep Learning, CNN, RNN**

I. Introduction

THIS demo file is intended to serve as a "starter file" for IEEE Transactions on Magnetics journal papers produced under LaTeX using IEEEtran.cls version 1.8b and later. I wish you the best of success. Next subsection will mention xxxxxx II-A

mds
August 26, 2015

A. Subsection Heading Here

Subsection text here.

1) Subsubsection Heading Here

Subsubsection text here.

II. Related Work

Recently, there are many researches and technologies for packet classification in networks. In the existing research, there is a rule-based packet classification method. Recently, research on deep learning has been developed and research on packet classification using deep learning has been actively carried out. Packet classification using deep learning is a method of automatically classifying packets without human intervention. In this chapter, we study the rule-based packet classification research and the packet classification studies that utilized deep-learning.

A. Classify Packets Using Deep Learning

Several studies using deep learning have been conducted recently. Also, researches to utilize machine learning in networks are actively being conducted. Accordingly, studies are being actively carried out to perform packet classification using deep learning of network.

Wei Wang et al. uses CNN model to classify malware traffic and general traffic. First, if 5-tuple (source IP/port, destination IP/port, protocol) are the same among the packets, one flow is defined as one dataset. In addition to the flow dataset, packets are defined as a session set as a dataset. The session dataset is a case where 5-Tuple is paired with the same flow and the same source IP/port and Destination IP/port cross each other. When data is divided into actual flow and session, data set is constructed by removing information of IP and MAC address. The reason is that IP and MAC address can show certain characteristics. The flow and session data sets constructed above are composed of 28 * 28 data similar to the MNIST dataset. The constructed dataset is used to learn CNN model to classify malware traffic and general traffic. In this paper, the result of classification of malware traffic and general traffic is 100

M. Lopez-Martin et al. used packet classification using CNN and RNN combination of deep learning model. The packet data is extracted from the header information and the payload data in the packet using the DPI Tool and used as learning data. The extracted learning data was used as input data to the combined model of CNN and RNN. They showed that CNN and RNN combined better than CNN and RNN models. In both of the above papers, learning was performed on the deep learning model by adding the header information of the packet to the dataset. In such a case, the classification accuracy may be high because the header information can be certain information that characterizes the data to some extent. Therefore, in this paper, we will perform learning only with payload data of application layer except header information of packet.

B. Rule-based Packet Classification

Rule-based packet classification is a method of classifying packets entering the network according to predefined rules. The rule-based packet classification method is classified by using the header information of the network packet. Therefore, rule-based packet classification is performed based on the source and destination IP and port of the packet header.

So, there are limitations to the rule-based packet classification method. Since the packet is classified using the information of the packet header, if the information of the packet that doesn't match the packet is received, the packet can't be classified or classified differently. In addition, because of rule-based packet classification with IP and port information of packet header, there are local limitations. Therefore, when a new network accesses or packets of a new network occur, there is a problem that a new rule must be redefined.

fangfan Li et al. [] At least it was used to lower the dependency of packet header information. Using this approach,

Manuscript received December 1, 2012; revised August 26, 2015. Corresponding author: M. Shell (email: http://www.michaelshell.org/contact.html).

they found that our packets haping device uses HTTP and TLS-handshake fields in their matching rules, and only for the first packet in each direction. If there is similar information in the header information of the incoming packet using the header information found in the first packet, it is classified as the packet of the same type. Although the IP and port number of the packet is used less, the method of classifying the subsequent packets by using the header information of the first packet also depends on the header information.

III. Conclusion

The conclusion goes here.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

Michael Shell Biography text here.

John Doe Biography text here.

Jane Doe Biography text here.