CS540 - Paper Review Report # 20

Hailu Belay Kahsay - 20155624

Title: A Survey of Techniques for Internet Traffic Classification using Machine Learning

Author: Thuy T.T. Nguyen and Grenville Armitage

Internet traffic could be inspected: by ISP for the purpose of effective network network management or automatic intrusion detection (to detect patterns indicative of denial of service attacks, trigger automated re-allocation of network resources for priority customers, or identify customer use of network resources that in some way contravenes the operator's terms of service), or due to government request(policy) to ISP. .

One way of doing it is by performing direct inspection of each packet's contents,at some point on the network, or controlling applications identity assuming that most applications use "well-known port numbers". As most application these days are not working in this way, this approach is not efficient; hence some researchers suggested that a deep packet inspection technique that relies on two assumptions: payload visible and payload can be interpreted, need to be used. But still these approach are not effective as most end users encrypt their data before they send it over the net and government may also impose privacy regulations, and the heavy operational overhead needed. Hence, a number of researchers are suggesting that using Machine Learning (ML) techniques for IP traffic classification.

The survey paper talks about some of machine learning based IP traffic classification research works done during 2004-2007. The paper starts by first discussing about the importance of IP traffic classification in operational networks: QoS and lawful inspection, Traffic classification metrics: Positives, negatives, accuracy, precision and recall, Byte and Flow accuracy, and then talk about the limitations of traditional description port- and payload-based classification. The authors then had a survey on how a number of ML algorithms can be used for offline analysis, such as AutoClass, Expectation Maximisation, Decision Tree, NaiveBayes with their accuracy levels. Finally authors have done comparison the research works done sofar.

Discussion points:

- Computational complexity of ML algorithms?
- Traffic Inspection Vs Privacy?
- Accuracy and processing speed concerns with the increase in the size of dataset?
- Scalability: Internet Infrastructure is growing .i.e. real-world applications of traffic classification require tools to work online.
    - For Online traffic classification requires accuracy and performance with this comes cost issue?
    - Latency and Computational Overhead issues

CS540 - Paper Review Report # 20

Hailu Belay Kahsay - 20155624

Title: The Applications of Deep Learning on Traffic Identification

Author: Zhanyi Wang

With the growth in today's network and variety of protocols used network traffic an accurate mapping of traffic to protocols or applications is important for network management, anomaly detection and etc. The author argue that some traffic identification techniques such as: Port-based, signature-based and statistical-features-based techniques are not efficient as these approaches are inaccurate, over-dependent on experiences of experts(labour-intensive), time-consuming and they do not work well when ports are new or changed. Hence, the author proposed an ANN and deep learning based network traffic identification.

The author started by introducing Artificial Neural Network (ANN) and deep learning with some of the most successful deep learning methods that involve artificial neural networks, such as Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Deep Belief Networks (DBN) and Stacked Auto-Encoder (SAE). Detailed description of the Key applications of the approach. such as Data set, automatic feature learning(Feature Extraction and Selection), Protocol Classification and Anomalous Protocol Detection, and Unknown Protocol Identification, is also included in the paper.

Finally, using the the real data that are collected from an enterprise network, the author claims that his approach works very well on the applications of feature learning, protocol classification, anomalous protocol detection and unknown protocol identification. He also claims that the same approach solves the problem of non-automation and poor adaptation in traditional ways. Applying conventional neural networks model and analysis of encrypted traffic have been left as a future work.

Discussion points:

- Lack of comparative graphs and analysis? .
- Latency and Computational Overhead issues?