

Visualization for Frequent Itemset Mining for Streaming Data

Sophie Qiu

May 16, 2016

1 Introduction

Visualising communication log data is useful in many cases. When large amount of data come in, it is hard to tell what is happening by looking at numbers and logs. Therefore, visualising data makes it easier for human beings to detect abnormal traffic.

There are some techniques presented in the literature on how to visualise network traffics. However, these algorithms are applied on static data rather than streaming data. There are some significant benefits if streaming data can be visualised at real time. It is helpful to detect malicious connection immediately. It also helps human beings to understand the flow of the traffic and how they evolve over time. In this project, I explored how to apply visualisation techniques on streaming data.

2 Related Work

The first work that is useful to my project is by Glatz et al.[?]. In this work, Glatz et al. introduces a schema to visualise network traffic data using hyper graphs. The schema first uses Frequent Itemset Mining (FIM) to identify the most frequent 5 connections. Each connection is represented by a circle, whose size is proportional to the frequency. Each circle contains the frequency of this connection in the examined period. Then the attributes of these connections, such as the source and destination IPs and Ports, are added to the corresponding circles.

However, their FIM is performed after the traffic data is collected and after the period is ended. Mozafari et al. propose an algorithm using sliding window technique to fast and accurately identify frequent items. In their work,

- 3 Method**
- 4 Result and Discussion**
- 5 Conclusion and Future Work**