Automated Analysis and Aggregation of Packet Data

Samuel Oswald Hunter

25th February 2010

1 Principle Investigator

Samuel Oswald Hunter
g07h3314@campus.ru.ac.za
Supervised by: Mr Barry Irwin

2 Background

Network telescopes provide us with a sampled view of the internet, more specifically they provide us with empirical data created by nefarious network traffic. A network telescope listens in and captures traffic destined for un-used address space, which means that they should receive little or no legitimate traffic [4]. There for the network telescopes sole purpose is to capture packets targeted towards the address range it operates on. These packets are stored and then used for further analysis.

The analysis of these packets may yield many interesting results, for instance they allow us to monitor small and large scale events on a network. An event such as a Distributed Denial of Service (DDoS) attack might be recorded, port scans and finger printing probes are also detected alongside traffic from automated worm propagation. All this traffic could be used to produce interesting reports and allow us to better understand the current state and inheritance of a network.

The information we wish to attain is determined by the security metrics we choose, choosing the right metrics will give us insight into the data we have [5]. The metrics determine what we are looking for in the data and thus determine how we will be analysing the data. Examples of metrics might be what are the 10 most targeted TCP or UDP ports or a traffic metric that determines what networks are responsible for the most traffic. The data obtained from the traffic metric might then be used to determine which countries are responsible for the

most traffic towards our network telescope.
The data obtained from the network telescopes may also be used in a more practical application such as generating real-time black hole lists (RBL), which can be used by mail servers to flag or reject spam originating from the addresses contained in the RBL. On internal networks the telescopes could act as an early warning system to worm propagation or other malicious activity.

3 Requirements and Objectives of Research

My project has two main sections the telescope data aggregation framework and the dashboard application. The dashboard application will generate graphs based on the analysis of captured packets while the telescope data aggregation framework will alow for easy management and aggregation of captured packets between multiple network telescopes and some central management node. A large portion of this project will be dedicated to the development of the dashboard application and data aggregation framework, further portions of research include security metrics, data visualization, mashup techniques and efficient transportation of packet data over a network.

3.1 Telescope Data Aggregation

There is a need for these distributed network telescopes to communicate and aggregate the packet data they have collected, so that the combined data maybe analysed at some central management node. This central management node will also be responsible for the secure aggregation of data between nodes. The framework that needs to be designed must allow for the different network telescopes to perform basic processing on their own nodes to transform the packet data into a suitable format (if required). One possibility to decrease the size of the data transfer is to serialize the packet data using MessagePack [2] a binary based highly efficient object serialization library.

Once at the central management node the aggregated data must be well organised and categorized, with options to export the data to various file formats such as xml and csv. In addition to these two file formats the central management node will also be responsible for creating sql scripts from the aggregated data to easily insert this data into a database. the central management node will also be responsible for producing other outputs such as Border Gateway Protocol (BGP) maps and Real-time black hole lists.

3.2 Dashboard Application

A dashboard application is then required to display reports that will be generated by the analysis of the captured packets. The dashboard will contain the security metrics that will be computed and then displayed to the user. The dashboard application will also allow for limited ad hoc queries to the packet database. The generated reports should contain relevant statistical information concerning the packets that have been analysed and should be generated at various predetermined periodic intervals.

It would be preferable for this information to be displayed in a highly interactive and visual manner to allow the user to more easily navigate and interpret the information. Some of the current metrics to create graphical representations from are:
Source to target IP address also indicating density (number of packets) on a world map by geographic location. Traffic type (port scanning, spam, unknown) by density as well as by geographic locality. Protocols used (TCP, UDP, ICMP) as percentages to each other (proportionate). Source/destination port numbers, total packets captured over time, time intervals between packet arrivals. The above metrics could also be displayed in correlation with each other, for example: A graph showing total ICMP or TCP SYN-ACK packets received on the x-axis and time intervals in 5 milliseconds on the y-axis. This graph could indicate a possible Distributed Denial of Service attack if the density of packets suddenly increased beyond the median for a certain period of time.

4 Further Research

There will be a strong focus on visual data representation in the dashboard application. Besides using traditional means of data visualization such as linegraphs, histograms, pie charts and bar charts. I would like to investigate dynamic means of representing data that may be interacted with by the user, possibilities of this include navigating a 3D environment that has been populated with packet data. I will also be looking at using Edward Tufte's sparkline graphs as they appear to work very well when contrasting multiple streams of information over a period of time and thus allowing the observer to more easily comparethem. Edward Tufte describes sparklines [6] as data-intense, design-simple wordsized graphics.

The reports generated by the dashboard application will be based on security metrics, although some metrics have already been identified I will be doing further research to identify more metrics that might yield useful results. Research in the secure and efficient transportation of data will also be done to help lower the bandwidth and processing power required for the transportation and serialization/deserialization of packet data in the data aggregation framework. The transportation research could also prove useful in the dashboard application if I decide to send data reports to an Adobe AIR [1] frontend.

5 Development

The Following languages and tools will be used during the duration of my project

- php
- python
- Adobe AIR
- Ajax, html, flash
- C #
- Rational Rose

Php and python will be used for the development of the data aggregation framework that will be running between the network telescopes and the central management node. Ideally I would like to use Adobe AIR with ajax, html and fash to develop the dashboard application. I will however first develop the dashboard as a web application using php and then later extend it by adding an interface to pull data from and display that data in a frontend created in Adobe AIR .Some further test applications may be written in c# using .NET and the SharpPcap [3] library, which is a .NET implementation of WinPcap. Adobe AIR development will take place in Aptana studio and Rational Rose will be used to produce design documents such as class diagrams, sequence diagrams, statediagrams and use cases.

6 Project Deliverables

The project will have two main deliverables, the first will be the data aggregation framework for the network telescopes and central management node. Thisframe work will allow the network telescopes to securely send data to the central management node. Pre-processing and some post processing will also behandled.

The second deliverable will be the dashboard application which will be able to securely interact with the packet database, handle the security metrics, generate and display the various automated graphs and reports to the user. The dashboard application should also be able to act as an early warning system if it suspects something like a DDoS attack and enable the user to perform basic ad-hoc queries.

7 Significance of the project

This project will yield two useful applications that will reduce the time spent analysing network telescope traffic. It will provide an early warning system to large events on the network and provide a tool for future research on network traffic analysis. Hopefully the project will also yield additional useful security metrics as well as new and powerful ways of visualising large amounts of quantitative data.

8 Time Line

February
Project Research
Project Planning

March
Project Research
Project Planning
Application Prototyping

April
Project Research
Application Prototyping

May
Literature Review
Design Documentation
June
Project Development (coding)

July
Project Development (completed)
Oral Presentation

August
Final Draft
Poster Presentation

September
Short Paper submitted
Second Draft

October
Project Complete
Final Oral Presentation

9 Future Extension

The dashboard application may be extend to include more security metrics and data visualisations. The reports generated by the dashboard my then also be serialized and retrieved by an Adobe AIR [1] application interface installed on a users machine. Further mashups may be created with the dashboard output and other technologies.

References

[1] Adobe air - rich internet applications that run outside the browser.

[2] Messagepack - ninary based efficient object serialization library. http://msgpack.sourceforge.net/.

[3] Sharppcap - cross platform packet capture framework. http://sourceforge.net/projects/sharppcap/.

[4] caida.org, 2010.

[5] Jaquith, A. *Security Metrics, Replacing Fear, Uncertainty and Doubt.* Addison Wesley, 2007.

[6] Tufte, E. *Beautiful Evidence.* Graphics Press, 2004.