Project Proposal

David Yates

February 28, 2014

1 Principle Investigator

David McWilliam Yates
Flat 1 One-On-Luke
1 Luke Street
Grahamstown 6139
076 811 3908
g11y1408@campus.ru.ac.za
Supervised by Prof. Barry Irwin

2 Project Title

A System for Characterising Internet Background Radiation

3 Statement of the Problem

Every year, millions of packets of data are sent to targets not set up to receive them. Although some sources of this data are simply misconfigured network adapters, other causes include malicious activity such as DDOS attacks, computer worms and botnets (Pang *et al.*, 2004).

A computer worm is a kind of virus that, once executed, duplicates itself and spreads to other computers over a network[1]. The most well-programmed worms are able to propagate through a network within minutes and can do so indefinitely (Staniford *et al.*, 2002).

A distributed denial-of-service (DDOS) attack occurs when numerous systems (usually compromised) all attack a single system, flooding it with packets to the point where it is too overloaded to serve ordinary users[2].

[1] http://searchsecurity.techtarget.com/definition/worm

[2] http://searchsecurity.techtarget.com/defininiton/distributed-denial-of-service-attack

Both of these phenomenons are harmful to users and content providers on the internet, and both can happen very suddenly, leaving victimised systems with little time to react. This project aims to create a system for characterising IBR such that it will be able to detect DDOS and worm activity in real-time, to aid in reaction to these events as they happen and further research on the subject.

4 Objective of the Research

Since 2006, network telescopes at Rhodes University have collected a total of almost 350 million unsolicited data packets. These packets include active traffic such as probes (such as port scans) and reflected, or passive, traffic, which is another term for IBR.

The primary goal of this research will be to categorise this set of historical Internet Background Radiation (IBR) data and to create a working prototypical system for doing similar characterisation on new data as it is recorded.

A secondary goal will be the creation of a system to generate traffic typical of the IBR data described for use in further research.

5 History and Background

Large volumes of IBR traffic were a recent phenomenon as of Pang *et al.* (2004). They result from two sources: computers running harmful code (such as worms and distributed denial of service (DDOS) attack tools) and computers with an incorrectly set up network configuration (Barford *et al.*, 2006).

Wustrow *et al.* (2010) explain that between 2004 and 2010, the volume and nature of IBR had grown and changed respectively, largely due to the advent of botnets, which are malevolent executables on compromised systems that communicate with and follow the directives of a central botnet controller program (Cooke *et al.*, 2005). Botnets are often used to perform DDOS attacks, which cause a backscatter of packets sent by the victimised systems in response to spoofed DDOS packets. These packets, being addressed to spoofed IP addresses, form part of IBR (Moore *et al.*, 2006).

Previous research in this area has focused on gathering packets for study via network telescopes, optimising these network telescopes (Pemberton, 2007), characterising the gathered packets (Barford *et al.*, 2006), and finding ways to reduce the background radiation (Wustrow *et al.*, 2010). Zou *et al.* (2005) proposed a method for early detection of internet worms via analysing patterns in IBR, placing an emphasis on the methodology of detecting trends in internet traffic rather than focusing on bursts of activity.

This project will focus on characterising the IBR portion of the data gathered by a network telescope at Rhodes University over a period of eight years. From this analysis a prototypical system for classifying IBR in real-time can be developed, which can be used as a framework for simulating IBR traffic to fuel further research.

General methods of characterising the data will include analysing the difference between daytime and night-time traffic, noting the common ports the traffic is directed to, analysing and comparing the contents of the packets in the data and finding points where significant changes occur in all or part of the data.

Kohler *et al.* (2002) showed that the arrangement of active addresses in the address space can be effectively modelled by a Cantor Dust multifractal system with two parameters. These structures remain constant over short periods of time and are able to act as unique identifiers for individual websites, but change significantly when worm activity enters the traffic. Barford *et al.* (2006) builds on this to show that IBR source addresses can also be modelled by a multifractal system, in this case a random cascade model.

Similarly, Border Gateway Protocol (BGP) routing tables have a general stable historical structure that can be compared against changes to the routing structure to detect spoofed routes, as in Qiu *et al.* (2007).

A system for detecting and characterising internet background radiation would be capable of sorting the large amounts of historical data into usable information. It would also be able to classify data as it is captured, and could therefore find application as an early warning system for DDOS attacks and worm activity, or as a tool for investigating developing botnets. The data could also be used to find ways to reduce the amount of IBR on the internet, reducing this unnecessary strain on bandwidth for internet service providers and their customers.

6 Approach

1. A literature survey will be performed, to gain a broader background in the research area and ascertain what has been done before in similar projects.

2. Time will be spent gaining familiarity with the data captured by the Rhodes University network telescopes and the *pandas* (Pandas, n.d.) data analysis library.

3. The system will be designed. This design will be incorporated in the plan of action.

4. The system will be implemented using Python and tested thoroughly.

5. Results of running the system will be recorded and compared to the results gathered by similar systems.

6. The project thesis and short paper will be written.

7. The project research website will be completed.

7 Requirements and Resources

- A modern personal computer running a distribution of Linux with Python and PostgreSQL.

- Access to the data gathered by the Rhodes University network telescopes.

8 Progression Time-line

Deadline	**Activity**
19 March	Literature survey
18 April	Design of data characterisation system
30 May	Literature survey and plan of action completed
11 July	System implementation completed
15 September	Short paper completed
31 October	Project thesis completed
7 November	Research website completed

References

Barford, Paul, Nowak, Rob, Willett, Rebecca, & Yegneswaran, Vinod. 2006. Toward a model for source addresses of internet background radiation. *In: Proc. of the Passive and Active Measurement Conference.*

Cooke, Evan, Jahanian, Farnam, & McPherson, Danny. 2005. The zombie roundup: Understanding, detecting, and disrupting botnets. *Page 44 of: Proceedings of the USENIX SRUTI Workshop*, vol. 39.

Kohler, Eddie, Li, Jinyang, Paxson, Vern, & Shenker, Scott. 2002. Observed Structure of Addresses in IP Traffic. *Pages 253–266 of: Proceedings of the 2Nd ACM SIGCOMM Workshop on Internet Measurment.* IMW '02. New York, NY, USA: ACM.

Moore, David, Shannon, Colleen, Brown, Douglas J., Voelker, Geoffrey M., & Savage, Stefan. 2006. Inferring Internet Denial-of-service Activity. *ACM Trans. Comput. Syst.*, **24**(2), 115–139.

Pandas. Python Data Analysis Library. Online. Available from: http://pandas.pydata.org/. [Accessed on 27 February 2014].

Pang, Ruoming, Yegneswaran, Vinod, Barford, Paul, Paxson, Vern, & Peterson, Larry. 2004. Characteristics of Internet Background Radiation. *Pages 27–40 of: Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement.* IMC '04. New York, NY, USA: ACM.

Pemberton, DS. 2007. *An Empirical Study of Internet Background Radiation Arrival Density and Network Telescope Sampling Strategies.* Ph.D. thesis, MSc Thesis, School of Mathematics, Statistics and Computer Science, Victoria University of Wellington, New Zealand.

Qiu, Jian, Gao, Lixin, Ranjan, Supranamaya, & Nucci, Antonio. 2007. Detecting bogus BGP route information: Going beyond prefix hijacking. *Pages 381–390 of: Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007. Third International Conference on.* IEEE.

Staniford, Stuart, Paxson, Vern, Weaver, Nicholas, *et al.* 2002. How to Own the Internet in Your Spare Time. *Pages 149–167 of: USENIX Security Symposium.*

Wustrow, Eric, Karir, Manish, Bailey, Michael, Jahanian, Farnam, & Huston, Geoff. 2010. Internet Background Radiation Revisited. *Pages 62–74 of: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement.* IMC '10. New York, NY, USA: ACM.

Zou, C.C., Gong, W., Towsley, D., & Gao, Lixin. 2005. The monitoring and early detection of Internet worms. *Networking, IEEE/ACM Transactions on*, **13**(5), 961–974.