Literature Review: Network Telescope Dashboard and Telescope Data Aggregation

Samuel Oswald Hunter

20 June 2010

1 Introduction

The purpose of this chapter is to convey to the reader a basic understanding required to interpret and understand some of the most important components of this thesis. To achieve this, an understanding between the contents of the thesis and past work should be established. The goal of this thesis is to develop a network telescope information dashboard that will need to analyse telescope traffic to produce useful representations of the data. The second part of this thesis is concerned with the development and implementation of a data aggregation framework that will enable information sharing amongst dispersed network telescopes and simplify the data management process between them.

By introducing topics that will be encountered in the rest of the thesis a clear distinction is drawn between past research and its applicability to current research. Past research contains a wealth of knowledge and it is important to consider other findings which could increase current understanding of a topic and lend explanation to new findings.

Section 2 starts by introducing the concept of network telescopes as a means to gather data on nefarious traffic found on the internet. It provides an introduction to the architecture of a network telescope, the various issues that face network telescopes and results from existing network telescope research. Some of the big contributors to network telescope research is mentioned, the different types of network telescopes are also shown along with similar technologies.

Section 3 of this chapter provides an overview of the two main types of traffic that is recorded by network telescopes, namely backscatter from distributed denial of service (DDos) attacks and the automated propagation of worms. These two malicious traffic types are explored to show their causes and how they are generated. Analysis techniques are revealed and enough information is covered to convey an understanding of what one should look for in raw traffic to infer any of these traffic types.

Section 4 introduces the concept of information dashboard by looking at the underlying principles of information dashboards and aspects of information visualization. Existing traffic visualisation techniques are looked at and the considerations and uses of metrics are explored. The use of static information dashboards in other fields such as that of finance and management are then explained.

Section 5 will address security metrics and measurement techniques. By showing what metrics are and why they are important and then explaining some of the most commonly used statistical methods for analysing and aggregating data.

2 Network Telescopes

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 captures all traffic destined for a range of un-used address space, which means it should receive little or no legitimate traffic. The traffic captured by the network telescope can provide information regarding denial of service attacks [19], automated worms and virus propagation [14].

When looking at traffic observed by a network telescope various research papers refer to the term backscatter [14] [19] [2] which refers to residual traffic observed from other hosts that have been the target of distributed denial of service attacks and are responding to spoofed source addresses. Other traffic such as network scanning from worms and malicious users also amount to backscatter while a very small portion of backscatter is the result of miss-configured hardware [2].

This chapter starts by exploring the various types of network telescope configurations which are required to produce a sampled view of malicious activity on the internet. The first two configurations of network telescopes discussed are active and passive, a brief introduction to similar technologies then follows such as honeypots and intrusion detection systems. Much research [25] [4] [10] has gone into the collaboration and aggregation of these various technologies to produce a more holistic view of nefarious internet topology. The chapter then continues by explaining the probabilities associated with the detection of internet events. The architecture and required configuration of network telescopes is then looked at, this includes bandwidth and storage requirements when operating different sized network telescopes and also introduces the topic of packet filtering. The section will then highlight some of the problems that network telescopes encounter while monitoring network traffic. This includes the geographic placement of network telescopes and the various assumptions that need to be made about the recorded traffic. Another issue that will be introduced is that of traffic poisoning and how network telescope operators try to avoid it. Lastly this section will explore existing network telescopes projects and what they have accomplished.

2.1 Types of network telescopes and similar technology

Network telescopes are also known as darknets or blackholes, the dark or black referring to the address space that is empty and thus not in use by devices. The basic setup for a network telescope is to have a server to which traffic that would normally be destined for un-used address space is forwarded. Different configurations exist for network telescopes, some respond to incoming traffic (Active Telescopes) such as the IMS which makes use of a lightweight responder [3] and others simple capture all traffic forwarded to them (Passive Telescopes). Similar technologies also exist such as honeypots which attempt to lure malicious traffic and intrusion detection systems (IDSs) that monitor live network traffic.

2.1.1 Passive Network Telescopes

The paper inferring Internet Denial-of-Service Activity by Moore et al. [19] shows how it is possible to infer Denial-of-Service (DoS) activity from a passive network telescopes. A passive network telescope is only capable of receiving incoming traffic and has no means of responding to any packets. The lack of response capability means the telescope is unable to complete the 3-way TCP handshake required to receive TCP payloads, this however is not required to infer some DoS attacks as will be shown in section 3. Passive network telescopes can however collect data from UDP and ICMP packets as they do not require active responses [10] and contain a wealth of information in the initial packet.

It has been shown that worm and virus attacks may also be inferred using a passive network telescope, the paper Observing Internet Worm and Virus attacks with a Small Network Telescope by Harder et al. [14] examines different methods of analysing worm and virus traffic on a small network telescope. Alternative methods also exist to detect the spread of malware such as using honeypots that attempt to attract malicious traffic and are able to respond to and in some instances capture malware [23] [1].

While passively configured network telescopes are unable to record TCP based exploit data or details of miss configured application requests [2], they are still capable of detecting source addresses and packet header information. Passive network telescopes are also able to record earlier worms such as Witty [24] and Slammer [18] which propagated over UDP and were able to be delivered to their target via a single packets payload.

The network telescope used for research in this thesis (RUscope) is a passive network telescope and will be introduced towards the end of this section.

2.1.2 Active Network Telescopes

Actively configured network telescopes are capable of issuing responses to certain requests, by doing so they are able to receive application level data that may lead to a better understanding of an exploit attempt. For example a telescope might be configured to reply to a TCP SYN request with a TCP SYN-ACK reply and in so would at least receive the first data packet [2] which is enough to identify a threat such as the blaster worm.

The Internet Motion Sensor (IMS) [3] project is an example of a distributed darknet monitoring system that makes use of a lightweight responder to elicit the initial packet of each TCP connection. This allows the IMS to retrieve more information than a passively configured network telescope, by implementing a very simple stateless TCP responder. The IMS also employs a novel payload storage technique, for each packet it receives it creates a MD5 checksum of the payload. [3] The checksum is then compared to a signature database for that day, if the signature does not exist in the database (it new) the payload is stored and the signature added to the database [10]. If the signature had already been captured that day it, the signature is logged but the payload discarded. This catching technique allows the IMS to reduce its storage requirements but does not hinder its ability to monitor new payloads or payload frequency.

2.1.3 Passive Network Telescopes

Honeypots are very similar in nature to network telescopes, they do however serve a more specific purpose in that they are used to emulate a vulnerable service or host in order to attract malicious traffic. They can be used to monitor individual addresses or function over a range of addresses [1].

One of the key differences between a honeypot and a network telescope is the resource requirements, because honeypots interact at greater depths with a threat (such as a worm) it requires more resources. This allows honeypots to characterize the vulnerability thats has been exploited and its affect on a machine [4]. Honeypots can be deployed in many different flavors, a single physical host could act as a high interaction honeypot alternatively that same host could emulate 10 virtual honeypots thus creating a virtual honeynet or honeyfarm. Honeypots may also be categorized as either a low or a high interaction honeypot. This refers to the level of interaction allowed with a honeypot, low interaction honeypot might only emulate venerable services, while a high interaction honeypot might be a complete physical machine with a vulnerable operating system. Banks of honeypots are sometimes used to handle traffic that was initially detected by network telescopes, filtered and deemed important enough for further analysis [4].

By analyzing the data captured from a honeypot it is possible to learn more about attack patterns such as how an attacker might elevate their privileges to gain root access on a box. Other information could also be determined like what type of attacks are currently the most prevalent, which ports they are targeting and which services they try to exploit. Lastly honeypots such as nepthenses [1] are used to collect worms and could provide insight to their spread and underlying architecture.

2.1.4 Intrusion Detection Systems

Intrusion detection systems provide automated monitoring and analysis of events that attempt to compromise the confidentiality, integrity, availability, or to bypass the security mechanisms of a computer or network [17]. Intrusion detection systems employ one of two methods to detect threats, the first makes use of known signatures of bad events, it compares traffic patterns to these signatures and logs the event if the traffic matches a known signature [17]. The second method involves traffic patterns and the deviation there of, when anomalies in traffic are detected the second method logs the event. Anomaly detection is a topic still undergoing further research and is employed in only a few intrusion detection systems in a limited form [17]. While intrusion detection systems are effective at identifying known threats, it comes at a price of resources as the traffic comparison against known signatures involves considerable resources especially as the signature database grows.

2.1.5 Hybrid Systems

A passive network telescope is not able to respond to requests, which might be needed to identify certain threats. It should be noted that even the use of an active network telescope might not guarantee that enough information is captured. While a distributed collection of network telescopes would be able to register global events, they lack the ability of assessing how exactly the threat is spreading. On the other side of the spectrum honeypots, antivirus software and intrusion detection systems posses the ability of collecting [1] or at least providing more detailed information on a threat. However due to the relatively small scale of addresses that they operate on, they are incapable of interpreting events on a global scale, such as those required to identify the early growth stages of a worm [4].

For this reason hybrid architectures have been developed that take advantage of the scale provided by network telescopes and the detailed threat interpretation provided by honeypots. A particular system explained by [4] provides a method whereby a collection of distributed network telescopes are used to track initial threats. The network telescope architecture then uses packet filtering techniques too select particular sources of the traffic that is then routed from the network telescopes to honeyfarms. These honeyfarms consist of multiple honeypots that are then able to interpret the threats at a lower level [4]. While the thesiss current research is not concerned with the combination of honeypots and network telescopes it is a field of notable importance and may one day form as an extension of the proposed aggregation framework.

2.2 Event Probability

The amount of traffic a network telescope observes is proportional to the size of the address block it is monitoring. The CAIDA Network Telescopes: Technical Report refers to the analogy of network telescopes as astronomical telescopes and explains how having a larger size (fraction of address space or telescope aperture) increases the quantity of basic data available for processing. [20] The size of the address block also influences the probability of that network telescope observing a given event. When looking at the Internet Protocol version 4 (IPv4) address space which allows a host a 32 bit address, there are a total of 2^{32} possible IP addresses. Address blocks are commonly assigned according to the number of leading bits that uniquely identify that address, a /8 address block would describe a range of 2^{24} addresses who all share the first 8 bits of their IP address. A /32 would then describe a single host, the probability of a network telescope monitoring a unique host in IPv4 address space is thus given by $p(x) = 1/2^x$ where p is the probability of monitoring the host and x is the size of the address block. For a telescope monitoring a /8 network, $p8 = 1/2^8 = 1/256$ which means the telescope has a 0.39% chance of observing a packet from a single unique host. A formula for calculating the probability of 1 of m packets being observed can then be calculated by $E(X) = nm/2^32$ where n is the number of unique addresses being monitored [19].

In conjunction with event probabilities it should be noted that there appears to be a disproportionate distribution between traffic observed and source address distribution. Research by [4] found that nearly 90% off all traffic observed by network telescopes were sent by less than 10% of the total observed source addresses. The distribution is illustrated in figure 1. below.

It has also been shown that the traffic observed by a network telescope does correlate to its locality in IPv4 address space, which results in different network telescopes observing separate and repeated events [4] [2]. The concept of locality and traffic will be explored more in section 3.

2.3 Architecture and Configuration

Multiple considerations need to be addressed when configuring a network telescope, for instance the method of forwarding traffic to the telescope that would have been destined for un-used address space needs to be decided upon. Two common formats that are used to collect telescope traffic include pcap and NetFlow [2], these formats record raw packet data that will need to be processed before analysis.

Storage and processing is also an important factor, as traffic that has been captured needs to undergo filtering to extract the desired packet information and then needs space to be stored for later analysis. It would be inefficient to analyse entire raw packets as they contain more information than would be used, instead we apply a filter to the packets and insert they attributes we wish to keep into a database for further analysis. An example of database tables for network telescope traffic is depicted below.

2.3.1 Packet Forwarding

When dealing with a small network telescope such as one capturing traffic from only a few IP addresses, [2] suggests configuring the network telescope to send ARP replies to the router for each un-used address. This however is not scalable for monitoring large blocks of un-used address space and may be improved by configuring an upstream router to statically route entire address blocks to the network telescope. This requires an entire address block to be dedicated to the network telescope and while this is easiest it might not be ideal, for a more flexible solution [2] suggests routing all packets that would have been dropped by the router to the network telescope. With the above examples it is assumed that the un-used IP addresses are in fact globally addressable and reachable. There are however methods that allow monitoring of unused non-routable addresses [9], such as those found inside service providers and large organisations internal networks. Three common network telescope configurations are illustrated below.

2.3.2 Storage, Filtering and Bandwidth requirements

The once a network telescope has captured traffic, that traffic needs to undergo a filtering process by which information pertaining to whichever study is being done is extracted from the packets. This information then needs to be stored into a database for later analysis. For instance, if a study only requires source and destination ports, there is no need to insert packet header and address information the database. The filtering of packets uses allot of processing power and research has been done into more efficient ways of filtering packets such as using FPGAs and GPU for processing the packets. Once the packets have been filtered, they need to be stored, storage provisions need to be made and storage capacity needs to be decided on.

The storage requirements of network telescopes is dependent on the amount of traffic that will be received, [2] has found that on a /24 sensor the average traffic rate can be approximated to 9 packets per second, while a /16 sensor would receive roughly 75 packets per second and a /8 sensor around 5000 packets per second. The average packet size from the Rhodes University Network telescope which contains just over 40million packets is 101 bytes, using the average packet size and estimated packet arrival rate according to sensor size one is able to make a approximation of the size of storage that will be required. Average bandwidth requirements according to [2] is displayed in the table below.

Sensor Size	Bandwidth Requirements
/24	7 Kbps
/16	60 Kbps
/8	4Mbps

The above bandwidth requirements and packet arrival rates were calculated from data collected by the globally deployed Internet Motion Sensor (IMS) [3] which is a distributed darknet monitoring system and will be looked at in more detail later in this chapter.

2.4 Problems facing network telescopes

Due to the inherently distributed architecture of the internet there is always a chance that traffic does not end up where it should be. Routers may fail, dns services may be miss configured and firewalls might drop legal traffic. These are just a few examples of the hurdles that traffic over the internet face. The public nature of the internet also means that any host may attempt to send data to any other accessible host on the internet, this intern could result in the poising of network traffic. It has also been shown [10] that the traffic collected at individual address blocks monitored by network telescopes may vary greatly from other address blocks that exist on separate IP ranges.

This section will highlight some of these inconsistencies encountered during traffic analysis and why certain considerations should always be kept in mind when conducting research with network telescope traffic. It will start by introducing the topic of network telescope placement and the variance of results obtained by monitoring geographically displaced network telescopes that are part of the Internet Motion Sensor project. A brief introduction to internet topology and architecture will follow to explain some of the problems encountered during traffic routing from source to destination. The topic of result poisoning follows which is one of the key reasons that telescope specific data such as actual IP addresses should never be published in research. Lastly the section will highlight ethical considerations pertaining to traffic captured by network telescopes and similar technologies such as honeypots.

2.4.1 Placement of Network Telescopes

Although the internet threats observed by network telescopes such as denial of service attacks and worm propagation are globally scoped, data from the IMS indicate widely different trends between separate network telescopes [10]. These differences where noted across three dimensions namely, over all protocols and services, a specific protocol and port and lastly signatures of known worms. Another publication [25] shares this view that multiple points of monitoring are required coupled with a collective interpretation to provide a more comprehensive view of nefarious network traffic. Thus observing traffic from only a single point would provide very little, if any information about the background activities [25]. Contrasting this view there are however still important information that can be learnt from a single vantage point, [14] showed how a small class C telescope was used to identify and distinguish between port scans, host scans and DDoS attacks. While the results found in [14] might not correlate strongly with global findings, their findings were still useful in understanding current threats.

2.4.2 Internet Topology

The internet is dependent on an ever expanding, interconnecting network of physical devices such as routers and switches, which are ultimately responsible for getting packets from point A to point B. These devices however can be overcome by traffic and the result of which is a loss of packets. Packets delivery from point A to point B can be slowed down by processing delays, queuing delays, transmission delays and propagation delays [16]. These delays can obscure the actual arrival rate of packets that may biased some of the results obtained from network telescope traffic. Packet loss can also be a serious problem, in the event of a large scale Distributed Denial of Service attack routers may queue packets and eventually drop [19] packets if the queue becomes too long. A good description of the vulnerability of traffic on the internet is presented by [16] as follows unfortunate that the physical laws of reality introduce delay and loss as well as constrain throughput. Miss configured of hardware and software can result in arbitrary but legitimate packets ending up in a network telescope, an example of this could be a NetBIOS configuration that sends small numbers of unsolicited packets to a monitored address range [19].

2.4.3 Result Poisoning

Due to the nature information that can be learnt from the analysis of network traffic, it would be in the best interest of certain nefarious entities to try and obscure the data obtained from network telescopes. According to [25] Knowledge of a monitors sensor location can severely reduce its functionality as the captured data may have been tampered with and can no longer be trusted. For this reason findings obtained from network telescope should never include actual address ranges used by the networks telescopes. The power of network telescopes lies in their capability to collect traffic from sources that believe they are sending data to legitimate hosts.

2.4.4 Ethical Considerations

Depending on the configuration of network telescopes and honeypot technology detailed information may be extracted from the traffic they obtain. Coupling the information that could be extracted and the underlying internet architecture that traffic depends on to reach its information there is always a chance that legitimate traffic might be captured. Even knowledge of illegitimate traffic might harm certain entities such as the naming of organisations that have been under attack from DDoS attacks or ISPs that have not put in place corrective measures such as ingress filtering to hinder DDoS attacks. As such service providers and content providers consider such information private and confidential [19]. As such care should always be taken in the publication of findings and the collection of data.

2.5 Existing network telescopes and big contributions

As a result of the importance and usefulness of network telescopes, various projects and collaborative efforts have developed to help manage the knowledge base and steer research in the field of network traffic analysis. This section of the chapter will introduce some of the big contributions to the field, not by specific authors but rather by collective effort of many. Caida is introduced first and then followed by team Cymru and the then the Internet Motion Sensor project. Lastly a brief introduction to the Rhodes University Network Telescope is given.

2.5.1 Caida

Caida, the Cooperative Association for Internet Data Analysis is a collaborative undertaking among organizations in commercial, government, and research sectors aimed at promoting greater cooperation in the engineering and maintenance of robust, scalable global internet infrastructure. [5] Amongst their various contributions they also supply data sets, produced by monitoring locations in several large Internet Service Providers (ISPs) and several other network telescopes to produce large datasets of network traffic. Caida also produces various important reference material that was used in the thesis such as the Network Telescopes: Technical Report [20].

2.5.2 Team Cymru

Team Cymru is a non-profit internet security research firm that is dedicated at making the internet more secure [27]. They provide in-depth information regarding setting up and maintain network telescopes. In addition to network telescope construction and use they provide secondary services that may be used during analysis of captured traffic, such as a malware hash registry that provides a look up service for captured malware.

2.5.3 IMS

The Internet Motion Sensor project that is run by the university of Michigan consists of a heterogeneous set of sensors and data aggregators which can be divided into two main categories [10]. Blackhole sensors (network telescopes) that collect treat data and topology sensors that provide context regarding the data collected by the blackhole sensors. While the blackhole sensors passively collect traffic over ranges of IP address space they include an active component that responds to so TCP SYN requests in an attempt to illicit more data [10].

The IMSs novel method of traffic storage by use of filtering, hashing and categorising has already been explored in section 2.1.2.

2.5.4 Rhodes University Network Telescope

The Rhodes University network telescope (RUscope) is the core provider of captured network traffic used for this thesis. The network telescope monitors a class c address block, /24 and is passively configured so that it on captures traffic targeted at it and does not respond to any requests. After packets are filtered they are inserted into a Postgress database for further analysis such as that achieved by this project.

3 Analysis of Network Telescope Traffic

Traffic observed by network telescopes could be explained as either a miss configuration of host hardware, backscatter produced by spoofed source addresses (most likely the event of a Distributed Denial of Service (DDoS) attack), scanning from worms or other types of probing [10]. The raw packet data obtained by network telescopes would be of little use if there were no method to distinguish the specific threats. There are however various techniques [19] [14] that are used to infer these types of traffic and they will be the introduced in this section. By understanding the traffic and how it is generated it become easier to identify it.

3.1 Distributed Denial of service attacks

The purpose of Denial of Service attacks is to consume the resource of its target, that being a host or network [19]. The result of which denies legitimate users access to that resource. DoS attacks may be divided into two main categories, that of logic attacks and that of flooding attacks. Logic attacks such as the Ping of Death exploit existing software vulnerabilities to crash or severely downgrade the service/availability of a remote server [19]. Logic attacks are often executed by sending a few well crafted packets to a vulnerable operating system or application and if the correct combination of packets is sent the target service or host could stop or crash [16]. Flooding attacks on the other hand focus on consuming as much of the targets CPU, memory or network resources. This is achieved by sending large numbers of spurious requests and botnets are often used for this purpose. Flooding attacks may be rather easily achieved by sending large volumes of small packets as quickly as possible as this can overwhelm routers and NICs packet processing capabilities.

One of the best known DoS attacks is that of the SYN flood, the SYN flood is a type of flooding attack that aims at immobilising a server by initiating as many TCP connections with the target as possible. For every SYN request packet the victim receives, it has to process that packet by going through a list of connections, if no match is found resources need to be allocated to the new connection. According to [19] even a small SYN flood can overwhelm a remote host. This shows how a single host may cause significant damage to a target, however often nefarious agents on the internet with make use of multiple hosts to perform attacks. Attackers compromise multiple hosts and leverage their combined bandwidth and processing power to mount more powerful attacks. These compromised hosts are referred to as zombie hosts and together make up is known as a botnet.

In an attempt to conceal their location and create multiple connections in the case of a SYN flood, attackers forge or spoof the source addresses of each packet they send [19]. Due to the spoofed nature of source addresses, they are also the cause of backscatter detected by a network telescope. During a DDoS attack, the victim attempts to send SYN-ACK or RST [21] replies to the Spoofed source addresses and if these addresses happen to fall into the same address range of a network telescope they are observed. Targeted hosts are not the only cause of backscatter from DDoS attacks, occasionally network devices between the spoofed address and the target send their own ICPM messages [26] to the spoofed address.

Packet Sent	Response from victim
TCP SYN (to open port)	TCP SYN/ACK
TCP SYN (to closed port)	TCP RST (ACK)
TCP ACK	TCP RST (ACK)
TCP RST	No response
TCP NULL	TCP RST (ACK)
ICMP Echo request	ICMP Echo reply
ICMP TS request	ICMP TS reply
UDP pkt (to open port)	protocal dependent
UDP pkt (to closed port)	ICMP Port Uncreachable

The table above illustrates common packet requests and their responses and was taken from [19].

3.1.1 Inferring DoS attacks

Before attempting to infer DoS traffic one needs to make the following assumptions. Addresses spoofed by attackers need to be uniformly distributed across the entire IP address space, that is the attacker needs to spoof the IP addresses at random. This assumption is often effected by ISPs that employ ingress filtering [12]. Ingress filtering monitors source addresses and drops packets with source addresses outside its client address range. This can cause neither that all packets may arrive at their target nor that the IP addresses are uniformly distributed. Its possible to check if a set of observed addresses are uniform to a network telescope range by calculating the Anderson-Darling(A2) test [11]. However as has been discussed earlier in section 2.2 and 2.4 distributed network telescope results vary from different address ranges. Although [19] still affirms that if the distribution of source addresses is not random, then it would be impossible to calculate an un-biased attack rate from the arrival rate.

The next assumption is that of reliable delivery, all attack traffic is assumed to have been delivered to a victim and all backscatter to the network telescope without issue [19], included in this assumption is that all packets elicit a response. These statistics are needed to correctly interpret the rate of packet arrival and estimate the size of an attack. Problems with this assumption include the volatile topology of the internet as discussed in section 2.4.2. Again ingress filtering may hinder the packet arrival as could intrusion detection software and firewalls that rate limit the packet arrival. These issues could result in the under-estimation of results.

According to [19] the last assumption needed is that all unsolicited packets received by the network telescope represent backscatter. Any host on the internet is free to send packets to any other host, this includes host addresses in a network telescopes range. Another issue that faces the last assumption is that of nefarious agents and the poisoning of telescope results as introduced by section 2.4.2. These assumptions where used by [19] to infer DoS attack traffic from network telescope data. They go on to say that their approach provides at worst a conservative estimation of current denial-of-service activity, as would be expected from the uncertainty presented by traffic delivery on the internet.

3.1.2 Attack Classification and Metrics

Due to the vast number of approaches one could take to analyse network telescope traffic, [19] has decided to focus on two classifications of traffic, flow-based and event-based. By categorising denial of service attacks within one of these two categories, [19] is able to determine the severity of attacks on short time scales with event-based. While making use of flow based classification it is possible to learn the following metrics; how many, how long and what kind [19].

A flow was defined by [19] as a series of consecutive packets sharing the same target IP address and IP protocol. A flow is then considered to exist from the 1st packet received until the last packet received within a 5min gap from the second last packet to be received, this timeout variable can affect end results. To ignore insignificant backscatter any flow with less than a 100 packets or minimum duration of 60 seconds was ignored. Flows also need to contain packets that have been sent to multiple addresses in the network telescopes range [19].

The following data was used by [19] for flow based analysis:

- TCP flags, to determine what flows consist of.

- ICMP payload, from ICMP packets with TTL expired.
- Address uniformity, whether they pass the A2 test and are uniformly distributed.
- Port Settings, whether port range is fixed, for A2 uniformity test.
- DNS information, DNS address of the victims source address.
- Routing information, prefix mask and origin according to a local BGP table.

Event-based classification makes use of the victims IP address and notes details over a fixed time window to examine time-domain qualities [19]. These qualities include the distribution of attack rates or simultaneous attacks. The event-based analysis was achieved by dividing the captured traffic into 1 minute intervals and noting the each attack event during that 1 minute [19]. Attack events consisted of a victim sending at least 10 backscatter packets to the network telescope in that minute.

These two methods, flow and event categories are just 2 examples of classification used during the analysis of network telescope traffic. Many other metrics may be extracted with different means. Arrival times discrepancy is introduced by [14], they show that inter arrival time of backscatter and normal traffic differ due to distinctive peaks around 10 and 120 milliseconds are missing [14].

3.2 Worm Propagation

Worms operate by spreading over a network to different hosts by exploiting vulnerabilities in the host operating system or in application level software [29], exploiting there vulnerabilities allows the worm to execute code and become self replicating. Worms scan ranges of network addresses for vulnerable hosts that they can exploit and in so doing are able to propagate and spread by themselves.

At any time there is a steady flow of background traffic on the internet caused by worms and viruses this traffic flow also includes port scans and backscatter from DDoS attacks [14]. This traffic will be visible to any routable, un-firewalled host connected to the internet. Furthermore it has been shown that many worm propagation (target selection) strategies are biased towards local addresses [29]. By targeting local addresses the worm is able spread at a faster rate, due to smaller network distances, common administrative practices and also exploitation of an already breached firewall. Examples of such worms include Code Red II [6], Nimda [7] and Blaster [8].

The existence of this never ceasing background traffic caused by worms does bring concern, that this traffic can pose a serious threat to the usability of the internet and do in fact cause instability during their epidemic stages. With increasing bandwidth speeds [30] and more devices being connected to the internet the platform for worms to spread is increasing. The main purpose of network telescopes when it comes to the observation of worms is defined by their configuration. Passive network telescopes only record packet traffic from worms and do not respond to any requests, their port scans can be detected and a good visual example of detected port scans are illustrated by [28] with the InetVis research. Actively configured network telescopes on the other hand are able to respond to some worm requests to obtain additional information and as such are able to identify TCP and UDP based worms.

Some concepts and terminology surrounding computer based worms include: worm virulence refers to the extent to which a worm generates traffic and which paths (routers) become most congested by it [3]. Worm demographics refer to the number of hosts infected the geographic and topological placement of hosts. The worm demographics also refer to infected host attributes such as operating systems, available bandwidth and application level software that was exploited (if any). Below is a graph indicating 3 phases of the Blaster worm life cycle.

4 Information Dashboards

A Dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance. [13]

As an integral part of this thesis is concerned with the production of an information dashboard for a network telescope it is important to outline a clear and concise definition thereof. The network telescope information dashboard will have to provide a concise and high level visual representation of relevant information derived from the analysis of the network telescope traffic.

A dashboard should be displayed on a single screen [13], this severely limits the amount of information that could be displayed and thus relevant and complementing information should be chosen for the dashboard. It has been shown in previous sections that the amount of information that can be obtained from a network telescope is large and can easily lead to an information overload. Thus [13] explains how a dashboard should scale the available information as to provide an overview which allows relevant information to be conveyed and indicate what would require further drilling down.

The concept of an information dashboard is not new and has been used extensively in the fields of finance and management. Any field where mission critical choices depend on large combinations of constantly updated information will benefit from the use of dashboards.

4.1 Data visualization

The volume of data obtained from network telescopes can easily overwhelm observers and allow information to become lost within the data, to avoid this different mechanisms are used to interpret and display data. Some data can easily and very effectively be represented by purely numerical means, other information however is much better interpreted when presented in a visual manner.

The following guidelines were taking from [15] and should always be considered when generating visual representations such as graphs and charts: Keep the graphic simple, colours might imply extra meaning or spark biases so try to avoid them. 3D graphics and shadows often distract from the actual data. Always clearly indicate empirical data and graphs should always be observed in context of their data. It should be noted that the negative perception of colour and its ability to distract, from [15] is contrasted by [13], who explains that colour can more easily help distinguish attributes such as those used in a pie graph with its legend. Both of these view points are correct and care should be taken to choose neutral colours unless the aim is to biased or more strongly outline a certain piece of information

By using the correct graphing techniques in the creation of a dashboard it is possible to create a concise summary of information that is easily understood and conveyed to the observer with little effort on their part.

5 Security Metrics

A Sans security reading described security metrics as metrics derived by comparison of two or more measurements that had been taken over a time to a predetermined baseline [22]. They further define metrics by stating that metrics are generated from analysis, as opposed to measurements which are generated by counting.

Metrics quantify particular characteristic of data to facilitate insight in the chosen subject area. Good metrics should be consistently measureable, repeatable, specific, cheap to gather, expressed as a cardinal number or percentage and have at least one unit of measurement [15] [22]. Security metrics thus determine what information one is interested in learning and providing empirical data and understanding on. The outputs of which should provide useful information and interpretation that can be used in conjunction with risk analysis and threat mitigation. Network telescopes are used to gather data and if there is any consideration for obtaining sensible information from that data useful and practical security metrics need to be determined that will be used to analyse the data.

5.1 Analysis Techniques

Raw data needs to be processed to produce information, information then needs to be analyzed to yield insight. Insight into the data is what is accomplished by using correct security metrics visualization techniques. A multitude of analytical methods exist that can be used to transform data into something that explains itself and this section will attempt to explain some of them as listed below.

- Average (Mean)
- Median
- Standard deviation
- Grouping and aggregation
- Time series analysis

5.1.1 Average (Mean)

Calculating the average value from a dataset is a standard aggregation technique [15] used to give an overview of a data set. The average is calculated by adding all the values from a set of elements and then dividing that sum by the total number of elements. While using averages is a good method of aggregating results it does however cause problems for some data sets. Outliers are obscured and the richness of underlying data is lost. For this reason [15] states that means are a poor choice for aggregating highly variegated data sets. he continues by adding elaborating the means have a tendency to obscure hidden insights and often steamroll over spikes and valleys that an analyst might consider interesting. Arithmetic means are however acceptable to use with certain data sets when the scope of data represented is narrow [15].

5.1.2 Median

The median is different from the mean as it denotes a value and proportion. The median represents the number that separates the top 50% of elements from the bottom elements and is calculated by sorting a dataset in descending order and then choosing the element in the middle of the dataset. The median is often used instead of the mean, as it can give better insight to data that has strong outliers and has said to offer significant advantages [15] over means in respect to measuring performance.

5.1.3 Standard deviation

Standard deviation reflects the dispersion of a dataset from its mean. A high standard deviation could indicate irregular or unpredictable data, while a low standard deviation would indicate a high degree of data clustering around the mean. The standard deviation is calculated by first calculating the mean for the data set. Then square the difference of each element from the mean, add together all the squared differences and divide the result by the number of elements and this produces what is known as the variance, the square root of which gives the standard deviation.

5.1.4 Grouping and Aggregation

Grouping and aggregation of data is one of the methods used to transform a large quantity of raw data into useful information [15]. Grouping refers to the organization of similar and complementing data from the same scope of analysis onto a combined unit. While aggregating refers to the calculation of summary statistics for each group of data, examples of which could include the sum, mean, standard deviation, minimum and maximum values. Grouping and aggregation can break down large data sets into meaningful chunks that are easily understood and might even bring to light underlying relationships between data.

5.1.5 Time Series Analysis

A very important analysis technique that is used in conjunction with network telescope traffic is that of Time Series Analysis as it is used during the inference of denial of service attacks [19], worm propagation [14] and load calculations [20]. Time series analysis is explained by [15] as attempting to understand the evolution of a dataset over time and will contain a series of observations for a particular attribute that have been taken at regular intervals.

5.2 Automation of Metric Calculations

By automating the generation of security metrics, it is possible to improve the aspects of repeatability and increased measurement frequency with minimal if any ongoing effort. This is achieved by automating the process of data gathering, computation and presentation. As part of this thesis is concerned with the development of a network telescope dashboard that will be capable of performing automated periodic analysis, it is important to explore the automated process of metric calculations.

Some of the benefits from automating metric calculations include added accuracy, repeatability, increased measurement frequency, reliability, transparency and audit ability [15]. Characteristics and considerations of an automated metric system includes the following.

A real time focus, this could be based in days, weeks, months [15]. If the information is represented over to fine an interval too much time will pass before it could be made use of, too large and it abstracts details. Automating the metric calculations also allow for the aggregation of aggregated results, that is observed events could be further aggregated.

6 Conclusion

Network telescopes have proven to be a useful tool in the capture of nefarious network traffic. The data they collect provides an indication of large and small scale network events such as worm lifecycles and denial of service attacks. By analysing raw traffic obtained from network telescopes using reputable metrics, one is possible to extract useful and accurate information. Using that information and transforming it into a visual representation to provide insight is an invaluable asset. The purpose of a dashboard is to convey meaning and understanding at a glance, by implementing the visual data created onto a dashboard it creates a platform from which to expand knowledge and drill down where necessary.

Network security is of increasing concern and by utilizing network telescopes too get a glimpse of the nefarious side of network activity it allows for a better understand and even preparation to face the ever increasing onslaught of malicious entities circulating the Internet.

References

[1] Nepenthes - detecting malware. http://nepenthes.carnivore.it.

[2] Bailey, M., Cooke, E., Jahanian, F., Myrick, A., and Sinha, S. Practical darknet measurement.

[3] Bailey, M., Cooke, E., Jahanian, F., Nazario, J., and Watson, D. The internet motion sensor: A distributed blackhole monitoring system. In *In Proceedings of Network and Distributed System Security Symposium (NDSS 05* (2005), pp. 167–179.

[4] Bailey, M., Cooke, E., Jahanian, F., Provos, N., Rosaen, K., and Watson, D. Data reduction for the scalable automated analysis of distributed darknet traffic. In *IMC '05: Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement* (Berkeley, CA, USA, 2005), USENIX Association, pp. 21–21.

[5] caida. caida.org, 2010. http://www.cert.org/advisories/CA-2001-19.html.

[6] CERT. Cert advisory ca-2001-19 ”code red” worm exploiting buffer overflow in iis indexing service dll. http://www.cert.org/advisories/CA-2001-19.html.

[7] CERT. Advisory ca-2001-26 nimda worm. Advisory, September 2001. http://www.cert.org/advisories/CA-2001-26.html.

[8] CERT. Cert advisory ca-2003-20 w32/blaster worm. Advisory, August 2003. http://www.cert.org/advisories/CA-2003-20.html.

[9] Cooke, E., Bailey, M., Jahanian, F., and Mortier, R. The dark oracle: perspective-aware unused and unreachable address discovery. In *NSDI'06: Proceedings of the 3rd conference on Networked Systems Design & Implementation* (Berkeley, CA, USA, 2006), USENIX Association, pp. 8–8.

[10] Cooke, E., Bailey, M., Mao, Z. M., Watson, D., Jahanian, F., and McPherson, D. Toward understanding distributed blackhole placement. In *WORM '04: Proceedings of the 2004 ACM workshop on Rapid malcode* (New York, NY, USA, 2004), ACM, pp. 54–64.

[11] D'Agostino, R. B., and Stephens, M. A., Eds. *Goodness-of-fit techniques.* Marcel Dekker, Inc., New York, NY, USA, 1986.

[12] Ferguson, P., and Senie, D. Network ingress filtering: Defeating denial of service attacks which employ ip source address spoofing, 2000.

[13] Few, S. *Information Dashboard Design: The Effective Visual Communication of Data.* O'Reilly Media, Inc., 2006.

[14] Harder, U., Johnson, M. W., Bradley, J. T., and Knottenbelt, W. J. Observing internet worm and virus attacks with a small network telescope. *Electron. Notes Theor. Comput. Sci. 151*, 3 (2006), 47–59.

[15] Jaquith, A. *Security Metrics: Replacing Fear, Uncertainty, and Doubt.* Addison-Wesley Professional, 2007.

[16] Kurose, J. F., and Ross, K. W. *Computer Networking: A Top-Down Approach.* Addison-Wesley Publishing Company, USA, 2009.

[17] Mell, R. B. P. ntrusion detection systems. NIST special publication, November 2001. http://purl.access.gpo.gov/GPO/LPS72073.

[18] Moore, D., Paxson, V., Savage, S., Shannon, C., Staniford, S., and Weaver, N. Inside the slammer worm. *IEEE Security and Privacy 1*, 4 (2003), 33–39.

[19] Moore, D., Shannon, C., Brown, D. J., Voelker, G. M., and Savage, S. Inferring internet denial-of-service activity. *ACM Trans. Comput. Syst. 24*, 2 (2006), 115–139.

[20] Moore, D., Shannon, C., Voelker, G. M., and Savage, S. Network telescopes: Technical report. http://www.cs.unc.edu/ jeffay/courses/nidsS05/measurement/moore-telescopes04.pdf.

[21] Pang, R., Yegneswaran, V., Barford, P., Paxson, V., and Peterson, L. Characteristics of internet background radiation. In *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement* (New York, NY, USA, 2004), ACM, pp. 27–40.

[22] Payne, S. C. A guide to security metrics:. http://www.sans.org/reading$_{r}oom/whitepapers/auditing/guide - security - metrics_{5}5, 19June$2006.

[23] Provos, N., and Holz, T. *Virtual honeypots: from botnet tracking to intrusion detection*. Addison-Wesley Professional, 2007.

[24] Shannon, C., and Moore, D. The spread of the witty worm. *IEEE Security and Privacy 2*, 4 (2004), 46–50.

[25] Shinoda, Y., Ikai, K., and Itoh, M. Vulnerabilities of passive internet threat monitors. In *SSYM'05: Proceedings of the 14th conference on USENIX Security Symposium* (Berkeley, CA, USA, 2005), USENIX Association, pp. 14–14.

[26] Song, D. X., Song, D., Perrig, A., and Perrig, A. Advanced and authenticated marking schemes for ip traceback. pp. 878–886.

[27] team cymru. Team cymru - about. http://www.team-cymru.org/About/.

[28] van Riel, J.-P., and Irwin, B. Inetvis, a visual tool for network telescope traffic analysis. In *AFRIGRAPH '06: Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa* (New York, NY, USA, 2006), ACM, pp. 85–89.

[29] Weaver, N., Paxson, V., Staniford, S., and Cunningham, R. A taxonomy of computer worms. In *WORM '03: Proceedings of the 2003 ACM workshop on Rapid malcode* (New York, NY, USA, 2003), ACM, pp. 11–18.

[30] Zou, C. C., Gong, W., and Towsley, D. Worm propagation modeling and analysis under dynamic quarantine defense. In *WORM '03: Proceedings of the 2003 ACM workshop on Rapid malcode* (New York, NY, USA, 2003), ACM, pp. 51–60.

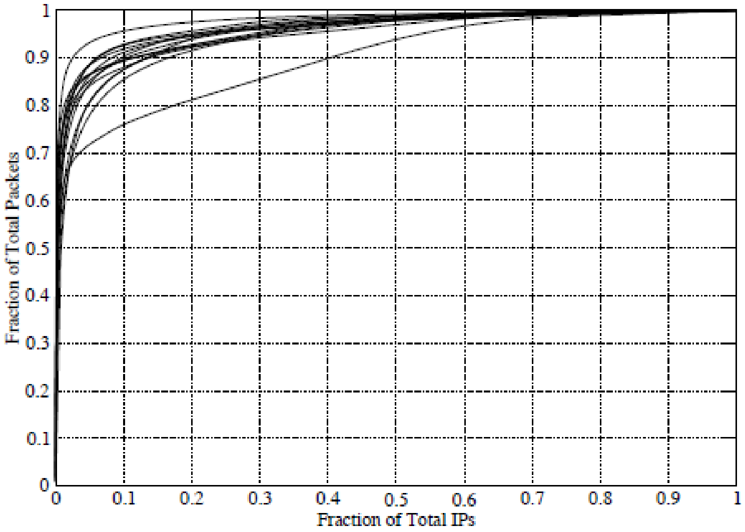
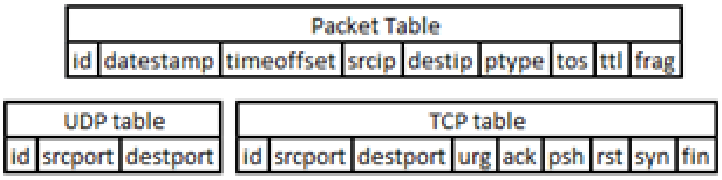


Figure 1: Contribution of individual IP addresses to the total number of packets collected from 14 darknets. Shows that 10% of source addresses are responsible for 90% of traffic observed [4] .

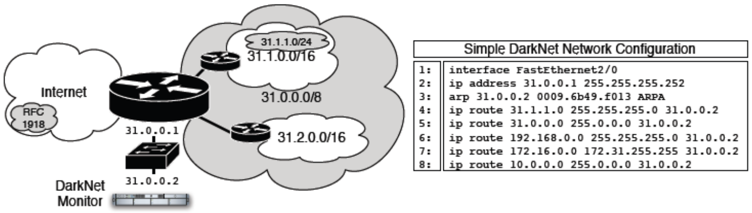


Packet Table								
id	datestamp	timeoffset	srcip	destip	ptype	tos	ttl	frag

UDP table		
id	srcport	destport

TCP table								
id	srcport	destport	urg	ack	psh	rst	syn	fin

Figure 2: Example of packet tables used in a database to record packets from a network telescope .



Simple DarkNet Network Configuration

```
1: interface FastEthernet2/0
2: ip address 31.0.0.1 255.255.255.252
3: arp 31.0.0.2 0009.6b49.f013 ARPA
4: ip route 31.1.1.0 255.255.255.0 31.0.0.2
5: ip route 31.0.0.0 255.0.0.0 31.0.0.2
6: ip route 192.168.0.0 255.255.255.0 31.0.0.2
7: ip route 172.16.0.0 172.31.255.255 31.0.0.2
8: ip route 10.0.0.0 255.0.0.0 31.0.0.2
```

Figure 3: 3 Basic deployement configurations for network telescopes: Capture out-bound traffic to reserved space (lines 6-8), traffic detined to a statically configured unused subnet (line 4) or capturing all unused address space within an allocation. (line 5) [2]

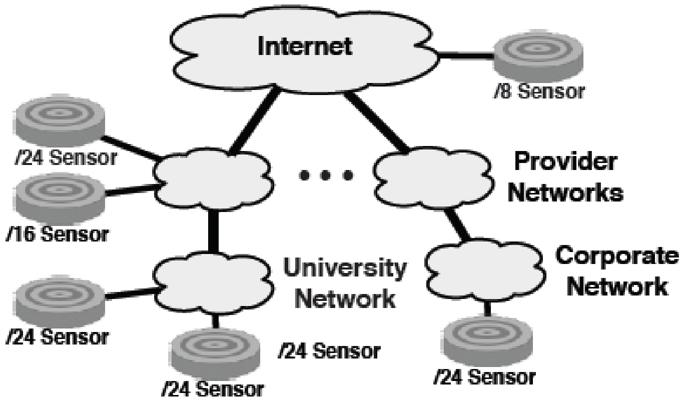


Figure 4: Internet motion sensor Architecture [10].

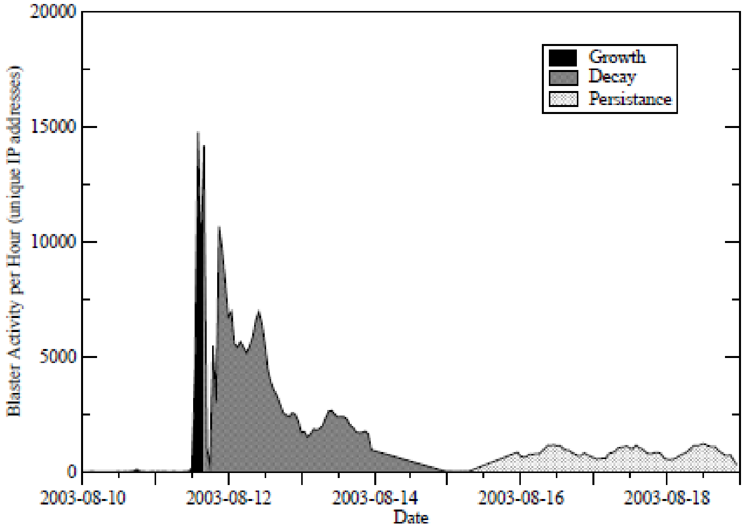


Figure 5: Blaster worm through its growth, decay and persistance lifecycles [3].