2013 5th International Conference on Cyber Conflict
K. Podins, J. Stinissen, M. Maybaum (Eds.)
2013 © NATO CCD COE Publications, Tallinn

A Baseline Study of Potentially Malicious Activity Across Five Network Telescopes

Barry Irwin
Security and Networks Research Group, Department of Computer Science
Rhodes University
Grahamstown, South Africa
b.irwin@ru.ac.za

Abstract: This paper explores the Internet Background Radiation (IBR) observed across five distinct network telescopes over a 15 month period. These network telescopes consisting of a /24 netblock each and are deployed in IP space administered by TENET, the tertiary education network in South Africa covering three numerically distant /8 network blocks. The differences and similarities in the observed network traffic are explored. Two anecdotal case studies are presented relating to the MS08-067 and MS12-020 vulnerabilities in the Microsoft Windows platforms. The first of these is related to the Conficker worm outbreak in 2008, and traffic targeting 445/tcp remains one of the top constituents of IBR as observed on the telescopes. The case of MS12-020 is of interest, as a long period of scanning activity targeting 3389/tcp, used by the Microsoft RDP service, was observed, with a significant drop on activity relating to the release of the security advisory and patch. Other areas of interest are highlighted, particularly where correlation in scanning activity was observed across the sensors. The paper concludes with some discussion on the application of network telescopes as part of a cyber-defence solution.

Keywords: *network telescope, darknet, internet radiations, scanning*

1. INTRODUCTION

This paper explores the Internet Background Radiation (IBR) [1], [2], [3] observed across five distinct network telescopes over a fifteen month period. These five network sensors each monitored a block of 256 IP version 4 (IPv4) addresses, with a combined size of 1 280 addresses. No live services or hosts existed in this address space, and as such one would expect relatively little traffic to have been observed. In contrast nearly 100 million errant, unsolicited datagrams were observed across the sensors, recorded from over 14 million sources. Of particular interest is the degree in the similarity of traffic observed across portions of the observed traffic, despite the monitored address blocks being numerically distant in terms of the IPv4 addressing scheme.

An advantage using smaller blocks is that one can attain a wider view of what trends are occurring with IBR, than one would with the same address space in a contiguous block. Greynets [4], [5] are a related implementation using smaller slices of address space than have traditionally been used for the operation of network telescopes, and may be of increasing value in the future. The case studies presented serve to illustrate some of the value in running distributed network sensors, as traffic can be correlated for an extended period, and responses to events such as security advisories observed in the collected data.

While a detailed analysis of all aspects of this observed traffic is beyond the scope of this paper, several interesting observations are presented, and analysed. Conventional wisdom relating to the sizing of network telescopes [1], [2], [6] has agreed that a large address space, such as that utilised by CAIDA[1] is needed in order to obtain meaningful data but, as shown in the following sections, viable results have been obtained using a significantly smaller aggregate sensor size than the /8 used by CAIDA, or /16 typically used by other researchers [3]. This work is also novel in terms of the correlation of activity and observed hosts across different network telescopes over a fairly lengthy period.

A. STRUCTURE

The remainder of the paper is structured as follows. Section 2 provides a brief introduction to the use and history of network telescopes. The data sets used in this paper are disclosed in Section 3 along with the collection methods used. A high level analysis of observed traffic is explored in Section 4. Two case studies are presented in Section 5, considering the application of a network telescope toward

[1] Cooperative Association for Internet Data Analysis http://www.caida.org/

the monitoring and identification of network threats. These are presented with a focus on the similarities and differences in traffic targeting these ports across the different monitored network address blocks. Section 6 concludes this paper, providing a discussion on the findings presented, their potential application, and future research relating to this area of study.

2. NETWORK TELESCOPES

Network telescopes are a class of network security sensors, which have been used by security researchers in recent years. The basis of a network telescope is to monitor portions of unused IP address space [7], [8]. Specifically a network telescope makes use of unallocated IP addresses which are not being used for running services. Based on this premise, any incoming traffic observed as destined for the monitored IP address range can be viewed as unsolicited, as no clients or servers are operating using these addresses. This allows researchers to focus on the traffic commonly termed Internet Background Radiation (IBR) [1], [8] without having to worry about distinguishing it from potentially legitimate traffic to servers or client systems. From a research perspective, no legitimate traffic should be arriving at the sensor, which can ease privacy concerns relating to traffic capture.

Care is taken to filter traffic so as to ensure that no response traffic is sent so as to appear to remote hosts as indistinguishable from an unallocated address. A more detailed discussion of the varying modes of operation for network telescopes and related analysis methods can be found in [9].

What is important to bear in mind when analysing the data collected using the passive means of collection such as that used in this research, is that one of the shortcomings of this type of network telescope setup is that only the first packet of the potential TCP 3-way handshake is actually captured. Since the handshake, by design, cannot complete due to filtering of any return traffic, no data payload can be captured. Due to this limitation it can only be inferred, albeit with a high level of certainty that traffic targeting a given port is related to particular protocols or malware such as the Conficker [10], [11] and Morto [12] worms considered in the case studies.

In essence a network telescope is a passive sensor system that collects incoming traffic or ‘radiation’ from the Internet. This radiation is constituted from multiple source systems and traffic types. The analysis of this collected data can provide useful insight into the operation of the Internet, or even particular events such as worms or distributed denial of service (DDoS) attacks. Over the last few years researchers have focussed on using telescopes for DDoS analysis as discussed in Moore et al. [7], [13]. Data collected has been successfully utilised for worm analysis, particularly that of Code Red [14] (the first worm observed on a Network Telescope) [15], [16]; Witty [17], [18], [19]; SQL Slammer [20] as well as more generically [21], [22]. In analysis reported by Kumar et al. [19], the researchers were able to perform detailed analysis of the Witty worm based on the traffic observed, to the extent of evaluating the number of physical drives present in infected systems and the probable identification of 'patient zero'. This was achieved through the analysis of data collected by a network telescope.

While all traffic received at the network telescope monitoring node can be seen to be unsolicited, the collected backscatter can be further classified under a number of categories. Strictly speaking backscatter can be regarded as traffic that is passive, and as such distinct from the active traffic recorded on the sensor. The term is, however, often misused in the sense of referring to all traffic that is not directly associated with communications of hosts on a network. This section builds on the view that traffic can be divided into the two broad classes of active and passive. Further discrimination is performed within these categories. Passive traffic can be defined as traffic from which no legitimate response can be expected from a system's TCP/IP networking stack when received [9]. As such it is unlikely that a potential attacker or instance of malware will be able to determine anything about the target system. The backscatter traffic observed can be seen to be the result of activities which result in the reflection of traffic from the originating machines to the telescope sensor. This requires that the source address of datagrams be spoofed to be within the IP address range monitored by a telescope sensor. Active traffic is defined as traffic which is expected to elicit a response of some kind when processed by a target system's TCP/IP stack. This differentiation is only easily possible for TCP and ICMP traffic where a clear structure is present as to what are a 'request' and a corresponding 'response'. UDP traffic is near impossible to classify as active and passive without doing payload analysis. To this end some sensors such as CAIDA's backscatter datasets filter it out [15].

3. DATA SOURCES

The data collected for this research was sampled from five network telescopes over a continuous period of 464 days from February 10th 2011 to May 20th 2012. The telescope's sensors each consisted of a /24 netblock (comprising 256 individual addresses) routed to a collection server. Packets were logged to disk on the sensor systems using libpcap with appropriate filters, to only log traffic destined to a specific netblock. The collection system host firewalls were configured to prevent any response to incoming traffic being generated, so as to avoid potentially disclosing their presence. The collector systems were located at two points on the TENET network. A sample setup of a collection system is shown in Figure 1Capture files were subsequently collated, and processed on a separate system using an analysis platform as described in [9].

The blocks of IPv4 address space being monitored were distributed across three distinct top-level IP version 4 network address blocks (netblocks) – 146/8, 155/8 and 196/8. These networks are all contained within the TENET[2] (AS2018) network. Three /24 blocks of addresses were contained in 196/8 block, in two separate /16 netblocks. Datasets are referred to by the /8 netblock in which they reside. A summary of the data sets collected is shown in Table I, along with the naming of the datasets. These have been sampled from the much larger datasets collected from these sensors, some of which have been running since 2005. Combined, the datasets used in this study comprise 99 007 576 packets.

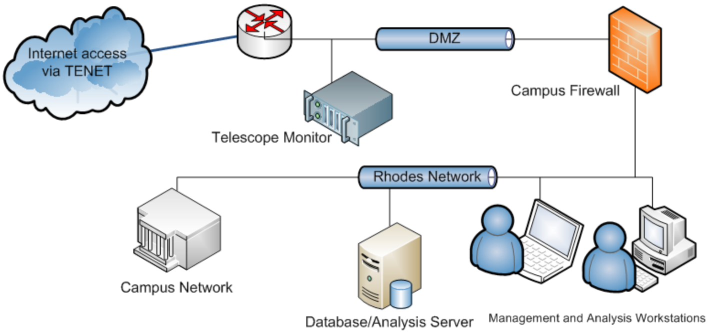

Figure 1. Sample Network Telescope Setup

Table I. Dataset overview

Dataset Name	Total packets	Protocol %			Sources /32
		TCP	UDP	ICMP	
146	4 768 524	62.11	25.26	12.62	467 419
155	7 547 605	77.18	14.45	8.36	498 134
196-1	20 071 795	91.42	6.25	2.30	3 304 445
196-2	32 479 388	92.87	4.87	2.25	5 036 684
196-3	34 140 264	90.63	7.43	1.92	5 103 316

[2] TENET is the Tertiary Education Network in South Africa http://www.tenet.ac.za/

The sensor collecting traffic for dataset 196-1 experienced a significant 171 day outage from May 20th to November 9th 2011 due to a failed network card[3]. The other sensors recorded traffic on all days in the period of study. Factoring in the outage, based on the recorded data, it's reasonable to expect that similar traffic levels would have been observed for this dataset as with the other two sensors in 196/8. What is particularly interesting is that despite this outage, the composition of traffic on 196-1 is very similar to the other two netblocks being monitored in 196/8 as shown in Table I.

4. ANALYSIS

The focus of this paper is to consider the similarity of the activity observed across the five network telescope systems. A summary of the traffic composition across TCP, UDP and ICMP for each dataset is shown in 0 TCP is seen to be the predominant protocol observed, particularly for the three blocks in 196/8 where it accounts for more than 90% of packets. Over 99.97% of traffic observed was accounted for as being one of TCP, UDP or ICMP. Traffic falling outside of these protocols is not being considered as part of this research, and consists predominantly of what appear to be damaged or corrupted, or otherwise nonsensical datagrams. In all cases TCP was the dominant protocol observed followed by UDP and ICMP. The higher proportion of UDP traffic present in 146 and 155, is most likely due to the decreased dilution caused by lower volumes of traffic destined to 445/tcp as compared to the sensor blocks in 196/8. This is discussed further in Section 5A. Datasets 146 and 155 also experienced significantly less traffic than the sensors in 196/8, with both host counts and packet counts being an order of magnitude less.

The remainder of this section comprises a closer look at TCP and UDP traffic observed, and the origins of some of traffic from a network perspective. A summary of the top ten ports for TCP and UDP as observed for the monitored address blocks are shown in Tables II and III respectively. The final area of analysis is a brief discussion considering the hosts which have been observed across multiple sensors.

[3] At this stage three sensors were operated by an external party with traffic only periodically analysed. Post outage the author took over the data collection and management of these sensors.

Table II. Top 10 TCP ports (SYN flag set)

	146		155		196-1		196-2		196-3	
Rank	Port	%TCP	Port	%TCP	Port	%TCP	Port	%TCP	Port	%TCP
1	**445**	21.92	**3389**	8.02	**445**	68.68	**445**	68.82	**445**	69.21
2	**3389**	15.51	**1433**	7.81	**22**	2.33	**22**	2.20	**22**	2.05
3	**1433**	11.95	**445**	6.76	**80**	1.84	**1433**	1.74	**1433**	1.90
4	**80**	10.89	**80**	5.95	**1433**	1.66	**80**	1.64	**80**	1.49
5	**22**	6.12	57471	5.86	**3389**	1.46	**3389**	1.54	**3389**	1.22
6	**8080**	5.12	**22**	4.34	**23**	1.13	49787	1.15	10300	1.13
7	139	4.26	**8080**	2.63	135	0.98	**23**	1.08	135	1.01
8	**23**	3.84	**23**	2.10	39459	0.84	135	0.89	**23**	0.87
9	135	3.59	3072	1.56	**8080**	0.81	**8080**	0.74	**8080**	0.76
10	3306	1.57	135	1.54	25	0.48	5900	0.48	5900	0.46
$\sum_{\text{Top10}}$		84.78		46.62		80.22		80.28		80.11

A. TCP

Traffic reported on in this section related only to those TCP datagrams received that had the SYN flag set. As such these packets were deemed to be ‘active’ in the sense that they would likely generate a response, and potentially establish a TCP session with a target host. Specifically excluded is traffic not matching this criterion which is determined to be backscatter. Only active traffic was considered, as it is felt that this provides a better indication of potentially malicious activity targeting hosts. Backscatter traffic can arise from a number of situations, such as the monitored address space being used in spoofed packets generated as part of a decoy scan or denial of service attack. These typically present as packets arriving with the ACK flag set if ports are open or RST otherwise.

TCP traffic observed across the five datasets is fairly consistent, being dominated by traffic targeting 445/tcp. Seven ports, 22/tcp (ssh), 23/tcp (telnet), 80/tcp (http), 445/tcp (microsoft-ds), 1433/tcp (ms-sql-s), 3389/tcp (rdp) and 8080/tcp (http/proxy), were present in the top ten actively probed ports across all sensors. These ports have been highlighted in bold in Table II. Dataset 155 is somewhat of an anomaly with the top ten ports representing only 46.62% of the TCP packets received, in contrast to the others which are over 80%. In this case the top twenty ports only accounted for 56% of the TCP data received. Other commonly targeted ports observed are 25/tcp (smtp), 135/tcp and 139/tcp which are used by older Microsoft RPC and file sharing implementations, 3306/tcp (mysql) and 5900/tcp (vnc). Scanning for hosts with open ports on common services such as these, particularly at his kind of volume and scale, is a typical pre-cursor to future possible exploitation attempts.

The port 3072/tcp and the ‘high value’ ports of 10300/tcp, 39459/tcp, 49787/tcp and 57471/tcp, are not commonly used by established protocols and couple possibly be scans for backdoors. Without TCP payloads this is difficult to determine with any certainty they do however warrant further exploration beyond the scope of this paper.

Table III. Top 10 UDP ports

	146		155		196-1		196-2		196-3	
Rank	**Port**	**%UDP**	**Port**	**%UDP**	**Port**	**%UDP**	**Port**	**%UDP**	**Port**	**%UDP**
1	**5060**	20.27	**5060**	22.27	**5060**	30.82	**5060**	36.11	**5060**	21.87
2	24003	12.66	**1434**	6.30	21566	8.32	19416	5.62	22549	2.86
3	**1434**	5.57	**137**	2.11	53	6.88	**1434**	4.32	**1434**	2.69
4	**137**	2.14	6257	2.02	**1434**	4.50	**137**	2.65	41560	2.20
5	5159	2.12	32737	1.71	**137**	2.25	6257	1.38	**137**	1.54
6	6257	1.75	53	1.71	6257	1.09	473	1.36	41559	1.21
7	41511	1.64	6568	1.48	9115	0.77	31683	1.20	6257	0.84
8	18261	1.62	60505	1.32	17762	0.63	38834	1.19	53	0.82
9	30989	1.55	43815	1.02	1046	0.63	53	0.87	15401	0.76
10	4375	1.54	39455	0.90	48170	0.57	6655	0.74	64578	0.71
$\Sigma_{\mathbf{Top10}}$		50.86		40.85		56.46		55.44		35.49

B. UDP

Observed traffic destined to ports using the UDP protocol on the sensor networks was found to be much more diverse than the case with TCP previously discussed. Only three ports 137/udp (netbios-ns), 1434/udp (ms-sql-m) and 5060/udp (sip) are common across the top ten on all sensors. These ports have been highlighted in bold in Table III. Common exploits exist for the 1434/udp service, in many cases scanning activity is attributable to the SQL Slammer worm, which has been around since January 2003. In this case the attribution can be performed with certainty for a given packet as the payload is present, and can be matched against known samples. The top ten ports in each sensor accounted for more than half the observed traffic in the cases on 145,196-1, and 196-2 and a significant portion in excess of a third in the case of the other two sensors. Port 53/udp is used by DNS and present in four of the sensors and rank in 15th place in dataset 146. Traffic commonly consists of spurious requests, often in the quest for open resolvers, which can be utilised maliciously in a number of ways. As with TCP, there are a significant number of ‘high’ ports, including a number from the ephemeral range (>49152), although Microsoft Windows systems typically use ports in the range 1025-5000 for this purpose.

Exploitation of Voice over IP (VoIP) services including those using SIP as a transport has become popular in recent years. Compromised systems are monetised by on selling calls. This is evidenced by it being the top ranked port in each sensor and accounting for more than 20% of UDP traffic in each case. As seen in the following subsection, significant proportions (19.27%) of the hosts targeting this port have been seen across all sensors. The relatively low host counts would indicate a well-co-ordinated scanning network, rather than random independent systems or attackers. Further investigation would serve as another interesting area to explore in the future.

Table IV. Observed sources across sensors

Target	3389/tcp		445/tcp		5060/udp	
Sensor Count	Hosts	%Sources	Hosts	%Sources	Hosts	%Sources
1	145 609	65.52	7 046 086	76.45	1 475	24.58
2	35 850	16.13	1 611 847	17.48	1 075	17.92
3	18 649	8.39	555 218	6.03	1 401	23.35
4	11 903	5.35	1 905	0.02	893	14.88
5	10 210	4.59	1 527	0.02	1 156	19.27
	222 221		9 216 583		6 000	

C. OBSERVED CROSS SENSOR HOSTS

Table IV contains a summary for three of the most popular ports observed in the datasets; 445/tcp, 3389/tcp and 5060/udp. In each case an analysis has been performed looking at the number of hosts observed targeting a port for each sensor. These lists were then combined, and common hosts enumerated, and a classification done based on the number of sensors on which a remote IP was recorded. Of these, 5060/udp was found to be the most interesting, despite the small total number of hosts, due to the fairly high proportion having been observed on multiple sensors. In the case of 3389/tcp more than ten thousand hosts were seen across all five sensors. Further exploration of this mode of analysis will explore this prevalence across sensors over shorter temporal periods.

5. CASE STUDIES

Two specific case studies have been chosen from the datasets. These are an analysis of the active (scanning) traffic destined for ports 445/tcp and 3389/tcp respectively. In both cases the similarity in observed trends across datasets is found to be significant. These also were two of the top TCP ports by packet count.

A. RPC/DCOM (445/TCP)

Port 445/tcp is used by the Microsoft Windows family of operating systems for providing distributed RPC services. This service has a long history of vulnerabilities and associated exploitation. One of the earliest of these was detailed in MS03-026 [23] and later in MS03-039 [24] – and subsequently exploited by the Blaster and Welchia worms in August 2003 [25]. A further vulnerability in the RPC stack was exploited by Sasser in April 2004, taking advantage of the vulnerability disclosed in MS04-011 some seventeen days previously [26]. The problems with the RPC/ DCOM stack as implemented in the Microsoft Windows Family operating systems continued and MS06-040 was released in September 2006 [27]. This was further widely exploited following MS08-067 [28], most notably by the Conficker worm. Work has previously been published relating to the observation of Conficker on single network telescopes [29], [30].

What is of interest in this research is the grouping of the datasets into two sets, based on the activity observed relating to this. As discussed in Section 4, only 'active' TCP traffic was analysed, as this was seen to be a more reliable indicator of actual malware activity, and associated scanning for the presence of the vulnerability, either by the malware itself or other sources. The observed counts of packets and distinct sources observed by each target address in the datasets are plotted in Figures 2 and 3 respectively.

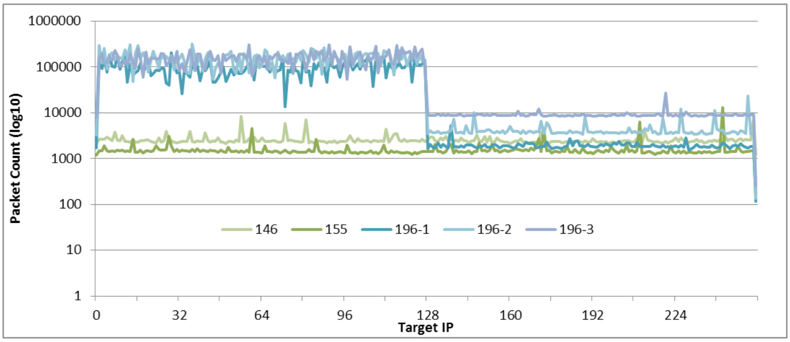

Figure 2. 445/tcp Packet count by target address

In both cases, two trends are notable. The first of these is that the three datasets in 196/8 exhibit fairly similar behaviour. Observations in the 146 and 155 datasets show just fewer than two orders of magnitude less traffic. The second trend is the substantial differentiation between the upper and lower /25 portions of each netblock for the 196/8 datasets. In all cases the highest monitored address x.x.x.255

received substantially less traffic. This reduced traffic volume is most likely due to it generally being considered as a broadcast address for a /24 subnet, and therefore unlikely to be utilised by a host. The sharp change in observed traffic levels occurs at x.x.x.128. This is due to a flaw [31], [32] in the propagation generation algorithm used by Conficker. The net effect of this is that the 2nd and 4th octets of the generated IPv4 address are limited to be in the range 0-127. The 146 and 155 datasets fall outside these ranges, whereas all of those in 196/8 are included. While this finding is not novel it is useful as a means of confirming the function of the data sources. While the majority of hosts scanning the lower /25 range could be regarded as being infected with the Conficker malware, this does not account for all traffic, as evidenced by the sustained scanning of addresses in the upper /25 of the monitored block. This is an important consideration when one considers that traffic targeting 445/tcp is the top destination by packet volume in all datasets with the exception of 155 where it is ranked 3rd.

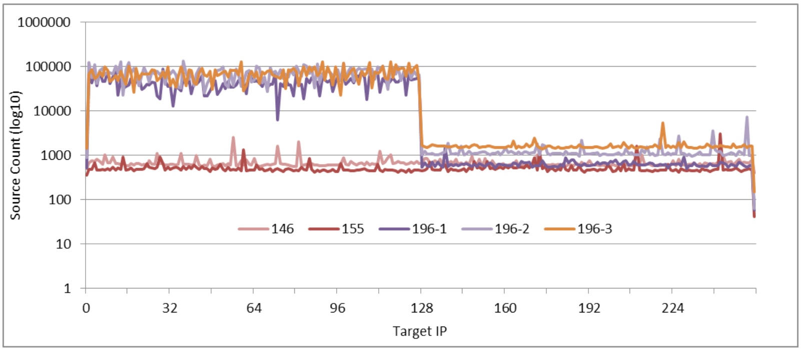

Figure 3. 445/tcp Distinct count by target address

Over the observation period 9 216 583 source IP addresses were observed as emitting datagrams targeting 445/tcp on the monitored networks. Of these 1.6 million (7.48%) were seen on two sensors and 3 432 were observed in four or more datasets. A summary of this can be seen in Table IV. When processing data, care needs to be taken to ensure that the volumes of traffic generated by hosts targeting 445/tcp don’t obscure other interesting datasets, with significantly smaller volumes. This is of particular concern in datasets such as those in 196/8 where it accounts for more than 68% of TCP traffic and greater than 62% of the total.

B. *RDP (3389/TCP)*

The Microsoft Remote Desktop Protocol (RDP) runs over port 3389/tcp. This case study was chosen for two reasons; the first being that there has traditionally been relatively little traffic observed targeting this port. In the first six months of the observation, the average number of packets observed was 324 per day, with 11 sources. This changed significantly from early August 2011, from which time traffic volumes increased substantially. The second reason is that the Microsoft Windows RDP service had a vulnerability disclosed, as detailed in MS12-020 [33]. The Morto worm also targeted this, gaining access to systems with this service exposed by guessing passwords. This was found in the wild and reported by Antivirus vendors on August 28th 2011 [12]. A detailed analysis of the worm and in particular its brute-force password technique can be found at [34].

An overview of the observed traffic can be seen in Figure 4, with Figure 5 containing an enlarged version of the area of interest from January to May 2012. The letters indicating areas of interest correspond in these two figures. The sharp spike in scanning activity can be seen starting in August 2011 (A), reaching a peak on August 24th 2011 (B), with 1 712 sources observed across all sensors. This trend of almost identical activity across all sensors continues until February 22nd 2012 (D) at which point there is a sudden divergence. The synchronised pattern observed in the source count during this period only varies by a few hosts, rarely differing by more than 120, and generally by less than 40 hosts between datasets. The cause for this departure from the trend is unknown. MS12-020 was published on March 13th 2012 and resulted in an almost instantaneous decrease (E) in the levels of scanning activity observed, reaching local minimums by March 20th 2012. The remaining period of observation (F) shows a steady increase in the volume of scanning activity observed. Points G and H are the result of connectivity outages for the two blocks being monitored at one physical site, although in the case of H, sensor 155 is also affected, possibly due to routing problems with the International peering link to the TENET network which were experienced around this time.

The traffic observed on the different network telescopes diverges from February 22nd 2012 (D). From this point, activity for 196-3 and 196-1 drops substantially, but these two remain at very similar levels for the remainder of the observed period, which is of interest given they are in different /16 netblocks, which are not adjacent. The data from 196-2 (which is in the same /16 network as 196-1) experiences an increase in traffic along with 146 and 155, attaining a local peak on February 28th. Traffic levels remain high through to March 18th when a substantial decrease is observed, possibly in response to the release of the Microsoft Security Advisory. From March 23rd, levels stabilise and start increasing again, with 146, 155 and 196-2 following similar paths. The dip in traffic (H) on April 30th 2012, marks a second split in the traffic similarity, with 196-2 experiencing relatively smooth continual growth, in contrast to 146 and 155. These two sensors then display a rapid growth, and maintain high levels of traffic until near the end of the observation period.

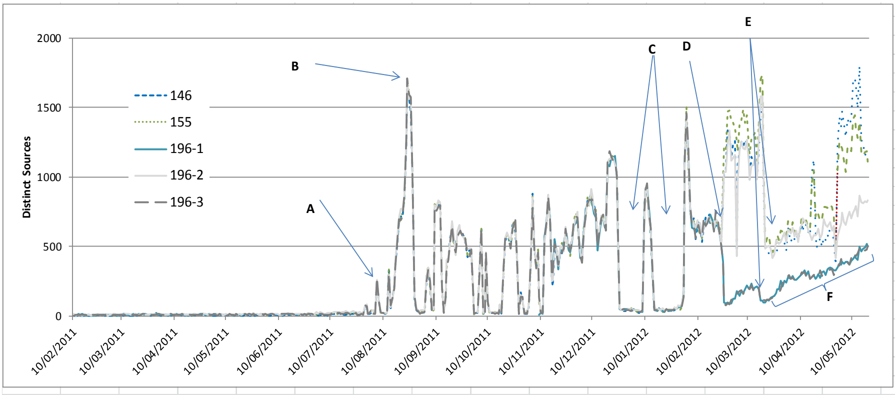

Figure 4. Distinct sources targetting 3389/tcp

While the reasons for the dips in activity as indicated by C are unknown, it is worth noting that the first trough started on December 25th 2011, through to January 8th 2012, followed by a week of traffic and then a lull for two weeks. A further exploration of the events around D and the changing composition of hosts at this point will serve as an area for further research, particularly comparing against other data sets.

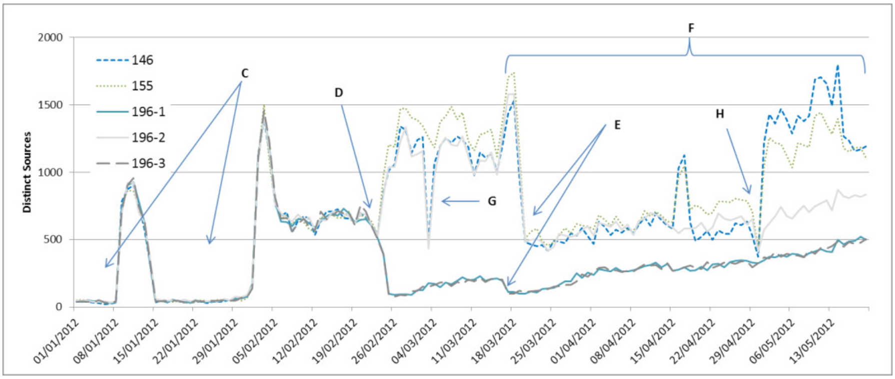

Figure 5. Distinct sources targetting 3389/tcp (1/1/2012 to 20/5/2012)

The most likely reason for the similarity of scanning across five distinct netblocks for much of the period of study is due to a series of co-ordinated scans being run.

The spike in traffic observed in August 2011, prior to the announcement of the Morto worm, can be seen as an early warning of unusual activity being observed on a wide scale, which may warrant further investigation. However as mentioned, without being able to capture traffic payloads, though tools such as honeypots, it is impossible to definitively state what was responsible for the scanning, although it is interesting that scanning almost ceased for a period following the announcements posted relating to the presence of Morto on August 28th 2011.

6. CONCLUSION

This paper has provided an introduction to the use of network telescopes in a coordinated manner across diverse IPv4 address space. Along with the general characterisation of the observed traffic presented in Section 4, and the highlighting of hosts being seen across multiple sensors, two specific case studies have been presented, illustrating two specific areas of interest within the dataset. These were chosen so as to be able to provide examples of the continued monitoring of an existing threat (in Conficker and similar malicious activity targeting 445/tcp) and the observation of two new and emerging threats targeting 3389/tcp.

A. CHALLENGES AND APPLICATION

The continued use of network telescopes faces a significant challenge. By their nature, they consume address space, which is becoming all the more valuable with IPv4. Work, such as this, demonstrates the effective use of smaller address blocks than have traditionally been used for conducting similar research. The introduction of IPv6 also brings challenges. In the researchers' experience, unsolicited traffic was not observed in the /48 IPv6 sensor that was previously operated. This may be a measure of the lack of general deployment of IPv6, combined with the general infeasibility of scanning such large swathes of address space.

An identified weakness of this collection technique, as previously identified, is the lack of payload for TCP connections, due to the lack of 3-way handshake. This could be mitigated to some extent by using honeypot systems interspersed with the addresses used by the telescope sensors.

The information produced by a network telescope can be used in conjunction with existing network security technologies to allow for a means of shunning or otherwise managing potentially hostile hosts, and protecting clients inside a network. This could be achieved through a variety of means, as appropriate for an organisation, ranging from route black holing to blacklist population. The observed issues, as exhibited by the problems with Conficker's propagation algorithm, highlight the importance in considering the diversity of placement of network telescope sensors in the future. Where possible ranges should be spread across a /16 blocks, and in both halves of a /24 – particularly for researchers with relatively small (≤ /25) address space being utilised. The viability of a range of address blocks has been demonstrated in terms of the diverse behaviour seen across them. Some of this may be due to numeric locality (malware tends have a preference for 'close' address space), rather than poor implementations as in the case of Conficker.

B. FUTURE WORK

The datasets used in the research can still be further analysed, particularly from the point of an extended geopolitical and topological analysis, such as that performed in [9]. Further exploration of these datasets and other subsequently collected datasets may provide better insight into the spread of malware and related malicious activity on a global scale, as well as how to better monitor and defend against these threats.

A goal of the researcher is to foster improved information sharing within the security research community. One challenge around this is the confidentiality of datasets (particularly relating to the ranges monitored), and the size of the raw captures. This can to some extent be mitigated though the publishing of metrics relating to observed data, rather than the data itself. A discussion of the specifics behind this, and some recommended metrics which can be used can be found in [9] and [35]. A significant step towards this would be the completion of an extended data processing framework what was prototyped for this analysis, which could publish reports on a regular basis.

REFERENCES

[1] R. Pang, V. Yegneswaran, P. Barford, V. Paxson and L. Peterson, «Characteristics of internet background radiation,» in *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement,* New York, NY, USA, 2004.

[2] E. Wustrow, M. Karir, M. Bailey, F. Jahanian and G. Huston, «Internet background radiation revisited,» in *Proceedings of the 10th annual conference on Internet measurement*, New York, NY, USA, 2010.

[3] D. S. Pemberton, «An Empirical Study of Internet Background Radiation Arrival Density and Network Telescope Sampling Strategies,» 2007.

[4] F. Baker, W. Harrop and G. Armitage, «RFC6018 IPv4 and IPv6 Greynets,» IETF, September 2010. [Online]. Available: http://www.ietf.org/rfc/rfc6018.txt.

[5] W. Harrop and G. Armitage, «Greynets: a definition and evaluation of sparsely populated darknets,» in *MineNet '05: Proceedings of the 2005 ACM SIGCOMM workshop on Mining network data*, New York, NY, USA, 2005.

[6] J. Goebel, T. Holz and C. Willems, «Measurement and Analysis of Autonomous Spreading Malware in a University Environment,» in *Detection of Intrusions and Malware, and Vulnerability Assessment*, Springer Berlin / Heidelberg, 2007, pp. 109-128.

[7] D. Moore, «Network Telescopes: Observing Small or Distant Security Events,» August 2002. [Online]. Available: http://www.caida.org/publications/presentations/2002/usenix_sec/.

[8] D. Moore, C. Shannon, G. . M. Voelker and S. Savage, «Network Telescopes: Technical Report,» 2004. [Online]. Available: http://www.caida.org/outreach/papers/2004/tr-2004-04/tr-2004-04.pdf.

[9] B. Irwin, «A framework for the application of network telescope sensors in a global IP network,» Grahamstown, 2011.

[10] Microsoft, «Virus alert about the Win32/Conficker worm (KB962007),» August 18 2008. [Online]. Available: http://support.microsoft.com/kb/826234.

[11] Microsoft, Win32/Conficker, Jan *8* 2009. Updated: Nov 10, 2010. [Online]. Available: http://www.microsoft.com/security/portal/Threat/Encyclopedia/Entry.aspx?Name=Win32/Conficker

[12] Microsoft, «Worm:Win32/Morto.A,» 28 August 2011. [Online]. Available: http://www.microsoft.com/security/portal/threat/encyclopedia/entry.aspx?name=Worm%3AWin32%2FMorto.A.

[13] D. Moore, G. Voelker and S. Savage, «Inferring Internet Denial-of-Service Activity,» in *In Proceedings of the 10th Usenix Security Symposium*, 2001.

[14] CERT, *CERT Advisory CA-2001-19 "Code Red" Worm Exploiting Buffer Overflow In IIS Indexing Service DLL,* 2001.

[15] D. Moore and C. Shannon, «The CAIDA Dataset on the Code-Red Worms - July and August 2001, (collection),» August 2001. [Online]. Available: http://www.caida.org/data/passive/codered_worms_dataset.xml.

[16] D. Moore, C. Shannon and K. Claffy, «Code-Red: a case study on the spread and victims of an internet worm,» in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, New York, NY, USA, 2002.

[17] C. Shannon and D. Moore, «The Spread of the Witty Worm,» *IEEE Security and Privacy,* vol. 2, pp. 46-50, July 2004.

[18] C. Shannon and D. Moore, «The CAIDA Dataset on the Witty Worm - March 19-24, 2004, (collection),» March 2004. [Online]. Available: http://www.caida.org/data/passive/witty_worm_dataset.xml.

[19] A. Kumar, V. Paxson and N. Weaver, «Exploiting underlying structure for detailed reconstruction of an internet-scale event,» in *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, Berkeley, CA, USA, 2005.

[20] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford and N. Weaver, «Inside the Slammer Worm,» *IEEE Security and Privacy,* vol. 1, no. 4, pp. 33-39, 2003.

[21] U. Harder, M. W. Johnson, J. T. Bradley and W. J. Knottenbelt, «Observing Internet Worm and Virus Attacks with a Small Network Telescope,» *Electronic Notes in Theoretical Computer Science,* vol. 151, no. 3, pp. 47-59, #jun# 2006.

[22] C. Zou, N. Duffield, D. Towsley and W. Gong, «Adaptive defense against various network attacks,» *IEEE Journal on Selected Areas in Communications,* vol. 24, no. 10, pp. 1877-1887, 2006.

[23] Microsoft, «MS03-026 : Buffer Overrun In RPC Interface Could Allow Code Execution (KB823980),» July 16 2003. [Online]. Available: http://www.microsoft.com/technet/security/Bulletin/MS03-026.mspx.

[24] Microsoft, «MS03-039 : Buffer Overrun In RPCSS Service Could Allow Code Execution (KB824146),» September 10 2003. [Online]. Available: http://www.microsoft.com/technet/security/Bulletin/MS03-039.mspx.

[25] Microsoft, «Virus alert about the Nachi worm (KB826234),» August 18 2003. [Online]. Available: http://support.microsoft.com/kb/826234.

[26] Microsoft, «MS04-011: Security Update for Microsoft Windows (KB835732),» April 13 2004. [Online]. Available: http://www.microsoft.com/technet/security/bulletin/MS04-011.mspx.

[27] Microsoft, «MS06-040 : Vulnerability in Server Service Could Allow Remote Code Execution (KB921883),» September 12 2006. [Online]. Available: http://www.microsoft.com/technet/security/bulletin/ms06-040.mspx.

[28] Microsoft, «MS08-067 : Vulnerability in Server Service Could Allow Remote Code Execution (KB958644),» Oct 23 2008. [Online]. Available: http://www.microsoft.com/technet/security/Bulletin/MS08-067.mspx.

[29] E. Aben, «Conficker/Conflicker/Downadup as seen from the UCSD Network Telescope,» February 2009. [Online]. Available: http://www.caida.org/research/security/ms08-067/conficker.xml.

[30] B. Irwin, «A network telescope perspective of the Conficker outbreak,» in *Proceedings of Information Security South Africa (ISSA)*, Sandton, South Africa, 2012.

[31] M. Richard and M. Ligh, «{Making fun of your malware},» August 2009. [Online]. Available: https://www.defcon.org/images/defcon-17/dc-17-presentations/defcon-17-michael_ligh-matt_richard-making_fun_of_malware.pdf.

[32] Carnivore.IT, «Conficker does not like me?,» 3 November 2009. [Online]. Available: http://carnivore.it/2009/11/03/conficker_does_not_like_me.

[33] Microsoft, «Microsoft Security Bulletin MS12-020 Vulnerabilities in Remote Desktop Could Allow Remote Code Execution (2671387),» 13 March 2012. [Online]. Available: http://technet.microsoft.com/en-us/security/bulletin/ms12-020.

[34] T. Bitton, «Morto Post Mortem: Dissecting a Worm,» September 2011. [Online]. Available: http://blog.imperva.com/2011/09/morto-post-mortem-a-worm-deep-dive.html.

[35] B. Irwin, «Network Telescope Metrics,» in *Southern African Telecommunications and Applications Conference (SATNAC)*, George, South Africa, 2012.