One-way Traffic Monitoring with `iatmon`

Nevil Brownlee

CAIDA, UC San Diego, and
The University of Auckland, New Zealand,
nevil@auckland.ac.nz

Abstract. During the last decade, unsolicited one-way Internet traffic has been used to study malicious activity on the Internet. Researchers usually observe such traffic using *network telescopes* deployed on *darkspace* (unused address space). When darkspace observations began ten years ago, one-way traffic was minimal. Over the last five years, however, traffic levels have risen so that they are now high enough to require more subtle differentiation – raw packet and byte or even port counts make it hard to discern and distinguish new activities.

To make changes in composition of one-way traffic aggregates more detectable, we have developed `iatmon` (Inter-Arrival Time Monitor), a freely available measurement and analysis tool that allows one to separate one-way traffic into clearly-defined subsets. Initially we have implemented two subsetting schemes; *source types,* based on the schema proposed in [12]; and *inter-arrival-time (IAT) groups* that summarise source behaviour over time.

We use 14 types and 10 groups, giving us a matrix of 140 *type + group subsets.* Each subset constitutes only a fraction of the total traffic, so changes within the subsets are easily observable when changes in total traffic levels might not even be noticeable.

We report on our experience with this tool to observe changes in one-way traffic at the UCSD network telescope over the first half of 2011. Daily average plots of source numbers and their traffic volumes show clear long-term changes in several of our types and groups.

1 Introduction

Since about 2002, observations of unsolicited one-way Internet traffic have yielded visibility into a wide range of security-related events, including misconfigurations (e.g., mistyping an IP address), scanning of address space by hackers looking for vulnerable targets, backscatter from denial-of-service attacks using random spoofed source addresses, and the automated spread of malware such as worms or viruses. Researchers have generally observed such traffic using *network telescopes*, deployed on *darkspace* (unused address space). When unsolicited traffic observations began, one-way traffic was minimal. Because of increased botnet-related activities over the last five years, e.g., [11], one-way traffic at the UCSD network telescope has risen – for example Aben [1] observed the increase in sources scanning TCP port 445 from almost none to 220,000 per hour over the

three months beginning on 21 November 2008. Now we see 6 GB/h of one-way traffic, i.e. high enough to require more subtle differentiation – raw packet and byte or even port counts make it hard to discern and distinguish new activities.

Seeking a better understanding of current and emerging one-way traffic behavior, we introduce and implement a methodology in a freely available measurement and analysis tool (`iatmon`) that provides an effective platform for separating one-way traffic into well-defined subsets, so that changes in a subset can be more easily recognised.[1] We implemented two subsetting schemes in `iatmon`, *source types* and *IAT groups*, described in section 3.3. This taxonomy facilitates analysis of not only the increasing unsolicited IPv4 traffic, but also pollution to IPv6 addresses, and comparison between the two. Tools such as `iatmon` can also enable consistent distributed monitoring of unused addresses across many different sites, effectively producing a wide-area view of unsolicited one-way traffic. The tool's utility is not limited to empty address space; one could equally deploy `iatmon` on partially populated address space, ignoring traffic to assigned hosts, or use it to monitor unsolicited traffic to all IP addresses on a stub network, since `iatmon` ignores bidirectional traffic.

We first review related work in darkspace traffic analysis, and then summarise research challenges and opportunities specific to darkspace measurement. In Section 3 we describe the data and our analysis methods. Section 4 presents results of our analyses. Section 5 summarises our contributions and future plans for one-way traffic analysis.

2 Related work

Data from the UCSD network telescope has supported significant research on DOS attacks [9], Internet worms and their victims, e.g., Code-Red, Slammer, Witty, and the Nyxem email virus. Data sets curated from telescope observations of these events became a foundation for modeling the top speed of flash worms, the worst-case scenario economic damages from such a worm, the pathways of their spread, and potential means of defense.

In 2004, Pang *et al.*[10] analysed one-way traffic destined for five different empty prefixes (a /8, two /19's, two sets of 10 contiguous /24 subnets) announced from two sites. They found that, relative to legitimate traffic, traffic to darkspace "is complex in structure, highly automated, frequently malicious, potentially adversarial, and mutates at a rapid pace." TCP packets dominated in all of their traffic samples, 99% of which were TCP/SYNs indicating either scanning or backscatter.[2] Also in 2004, Cooke *et al.* found diversity in incoming traffic to ten unused address blocks [6] ranging in size from a /25 to a /8, announced from service provider networks, a large enterprise, and academic networks. They passively recorded incoming packets and actively responded to TCP SYN requests to obtain more data from the sources. They found traffic

[1] The `iatmon` tool is available at `http://www.caida.org/tools/measurement/iatmon/`.

[2] At UCSD from Jan–Jun 2012 only about 30% of the packets were TCP.

diversity along three dimensions: across protocols and applications; for a specific protocol/application using TCP port 135, and for a particular worm signature (Blaster).

In 2006 Barford *et al.* analysed the source address distribution of malicious 'Internet background' traffic. They evaluated traces from network telescopes running active responders on portions of two /16s and one /8 network, in addition to a large set of intrusion detection system logs provided by Dshield.org [2]. They found a bursty distribution of source addresses, many from a small set of tightly concentrated network locations, which varied across segments of darkspace but were consistent over time for each separate segment.

More recently, in 2010 Wustro *et al.* [14] analysed background traffic using one-week traces from four /8 darkspaces. Traces of one-way traffic destined to 35/8 from 2006 to 2010 showed an eight-fold rise in traffic rate over the five years, with daily variations during a week. The 2010 data rate (about 20 Mb/s) was similar to that at the UCSD network telescope. One-week traces from three of the four darkspaces suggest that the overall daily variations in traffic volume were similar across all sites. Their TTL distributions were also similar, suggesting that all the sites see similar spatial traffic distributions.

In 2011, Treurniet [12] proposed a traffic *activity classification schema* to monitors state changes in TCP, UDP and ICMP flows in order to help detect low-rate network scanning activities. She used state changes per sending host to classify that host as normal, DoS, backscatter, scanning, etc. She tested the approach on bidirectional trace data from outside the border of a /14 network with some responding servers in it, although fewer than 1% of observed flows were bidirectional communications with these servers. The rest of the observed traffic was malicious, mostly slow scans of various types.

3 Methodology

The UCSD network telescope [5, 4] uses a /8 network prefix, most of which is dark. An upstream router filters out the legitimate traffic to the reachable IP addresses in this space, so we monitor only traffic destined to empty address space. Management of the UCSD network telescope requires continual navigation of the pervasive challenges in network traffic research methodology: collection and storage, efficient curation, and sharing large volumes of data. The large volume of data captured by the telescope incurs considerable expenses for data storage and limits the number of researchers who can realistically download data sets. The situation is worse during malicious activity outbreaks when the data volumes increase sharply, yet rapid analysis and response are necessary.

The UCSD network telescope remains a purely passive observer of unsolicited traffic. We do not rule out active response by the telescope in the future, but active responding requires resources and careful navigation of legal and ethical issues. We have found that much can be gleaned with non-intrusive methods and external knowledge of malware behavior. For example, we indirectly observed the rise of Conficker A and B because Conficker induced a conspicuous increase in

Description	**Type**
TCP	TCP probe
	TCP vertical scan
	TCP horizontal scan
	TCP other
UDP	UDP probe
	UDP vertical scan
	UDP horizontal scan
	UDP other
Other	ICMP only
	Backscatter
	TCP and UDP
	μTorrent
	Conficker C
	Untyped

(a) Source Types

IAT distribution	**Group**
Long-lived	Stealth &3 s mode
	Stealth & Spikes
	Stealth other
3 s mode	Left-skew
	Even
	Right-skew
Other	Short-lived
	High-rate
	DoS
	Ungrouped

(b) IAT Groups

Fig. 1: Subsetting Schemes for One-way Traffic Sources

the number of probe packets aimed at TCP port 445, using poorly randomized destination addresses [1].

3.1 Data set

The UCSD network telescope collects full-packet traces continuously. These traces are stored online for at least sixty days, allowing vetted researchers to analyse the data in various ways. (See CAIDA web page for access to data [3].) One of the goals of this research is to enable retention of efficient but rich summary statistics of historical raw data that is itself too expensive to archive indefinitely. [3]

We analysed each hourly trace from the UCSD network telescope for every hour from 3 Jan through 30 Jun 2011. During that time, 23% of the sources were TCP, contributing 30% of the packets and 69% of the volume (MB/h).

Note that since most traffic into darkspace represents failed attempts to initiate connections, about 99.9% of the TCP packets carry no payload. However, the number of UDP packets carrying payload has increased in recent years. As of January 2011, about 67% of the sources sent only UDP packets, accounting for about 55% of each hour's packets. In recent years the amount of UDP traffic has increased; in January 2011 the TCP traffic per average hour contributed only 66% of the bytes, 44% of the packets and 32% of the one-way sources – by June 2011, the hourly averages for TCP had changed to 70% of the bytes, but only 23% of the packets and 18% of the one-way sources.

3.2 Analysis strategy

In 2002, when CAIDA began analyzing telescope data, one-way traffic volumes were low, so that rapid increases caused by viruses and fast-spreading worms

[3] Current storage pricing for researchers at SDSC are $390/TB-year, which at 6GB/hour results in over $20,000/year of storage costs, which multiply for each year of data to be stored, and increase as unsolicited traffic rates grow.

were easy to discern. As of June 2011, we see 6 to 9 GB/h of one-way traffic, so that the early stages of a new rapidly spreading attack are much harder to observe in the continuous swamp of background traffic. For example, we observe that each hour 76% of the sources send UDP packets, but these packets account for only 23% of the hourly byte volumes, so that the UDP sources are swamped (in terms of byte volumes) by TCP sources. One goal of our `iatmon` methods and tool is to separate the one-way traffic into various subsets so that changes within subsets are easier to detect.

3.3 `iatmon` implementation

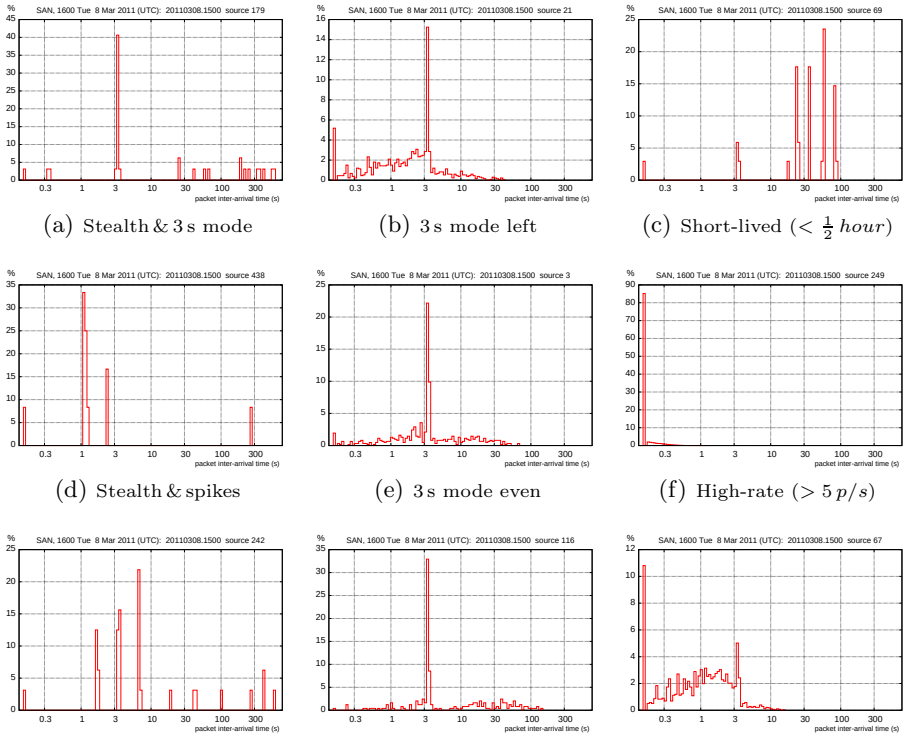
First, we consider the one-way traffic in terms of its *sources*. Specifically, we construct a *source table*, with entries that summarise the set of packets arriving from a source IP address. At the end of each hour's trace `iatmon` scans the source table, calling various classifying functions to separate the sources into clearly-defined subsets. We describe our initial implementation of these subsetting schemes below. `iatmon` has a Ruby outer block, which sets up `iatmon's` configuration and writes its hourly summary file. However, most of the processing occurs in a C extension module that reads trace packets, maintains the source hash table, classifies sources using the various subsetting schemes, and passes results back to the outer block. For normal hours, `iatmon` running on an 8-core machine with 32 GiB of memory takes 7 to 15 minutes of elapsed time to process each hour's trace file, easily fast enough to keep up with new data from the telescope.

`iatmon` handles its own storage allocation, requesting memory in 4 GB chunks as needed. Two such chunks are sufficient for most hour-long intervals of trace data, but for nine of the hours `iatmon` was unable to get enough memory to process the entire hour – even when using 28 GiB of real memory. We examined the first few thousand packets of each of those nine hours, and found that most of their sources sent just two TCP packets, from apparently random source addresses. To allow `iatmon` to handle traces when such DoS attack conditions occur, we discarded sources that only sent two packets and were then idle for at least 120 s. One of these nine traces had 122.5 *million* source addresses that we discarded (97% of the hour's sources), but most packets in the flood traces had the same TTL, suggesting a high-rate DoS attack using spoofed source addresses.

Source *types* One obvious approach to classifying one-way sources is to examine their use of IP addresses and ports. For example, a *probe* source sends all its packets to a single port on a single host. A *vertical scan* source sends packets to various ports on a single host. A *horizontal scan* source sends packets to the same port on many hosts. We used this approach in early development work on `iatmon`. Our current *source types* scheme is now based on Treurinet's [12] recent classification scheme, using the source types listed in Table 1(a).

In 2008 we saw the rise of the Conficker worm/botnet. Conficker C may send both TCP and UDP packets as part of its p2p network establishment; SRI's

algorithm [11] allows us to estimate which incoming packets are likely Conficker C p2p packets. We were curious as to whether BitTorrent has contributed to the rise in UDP one-way traffic observed in the last two years. The BitTorrent protocols are well-documented [8, 13], allowing us to identify μTorrent sources from UDP packet payloads. Table 1(a) includes the two source types we added to represent these application-specific sources.



(a) Stealth & 3 s mode (b) 3 s mode left (c) Short-lived ($< \frac{1}{2}\, hour$)

(d) Stealth & spikes (e) 3 s mode even (f) High-rate ($> 5\, p/s$)

(g) Stealth other (h) 3 s mode Right (i) DoS

Fig. 2: IAT groups: Packet Inter-Arrival Time percentage distributions for a typical hour, ending at 1600 on 8 Mar 2011. We determine each source's IAT group using a set of distribution metrics, such as *% IATs* < 150 *ms, mode %, skew, %* > 120 *s;* only 1% of the hour's sources remain unclassified.

Source *IAT groups* About 30% of the packets that reached the UCSD telescope in the first half of 2011 were TCP SYNs, carrying no payload. In order to further characterise the source behaviors, especially those from TCP sources, we investigated the inter-arrival time (IAT) distributions from sources active in a typical hour. To search for recognisable IAT patterns we plotted many sheets

of 'postage-stamp size' IAT distribution plots, with each distribution's parameters shown on its plot. We examined these plots manually to find common patterns, then developed algorithms that captured these recognisable subsets of the sources, based on statistical properties of their IAT distributions. For example, many sources exhibit a strong inter-arrival time mode at 3 s, the standard time for TCP retries, aggregated with an underlying wide range of IATs. A Poisson process that sends TCP SYN packets, and resends each packet after 3 s intervals will produce this kind of distribution, with skew that depends on the process's average time between sending new packets.

Once we identified a clearly distinguishable IAT pattern, we developed an algorithm to capture it, and assigned it an *IAT group* label. After several cycles of that process, we settled on a scheme with nine different groups (listed in Table 1(b)) that meaningfully distinguished about 99% of the sources in each hour.

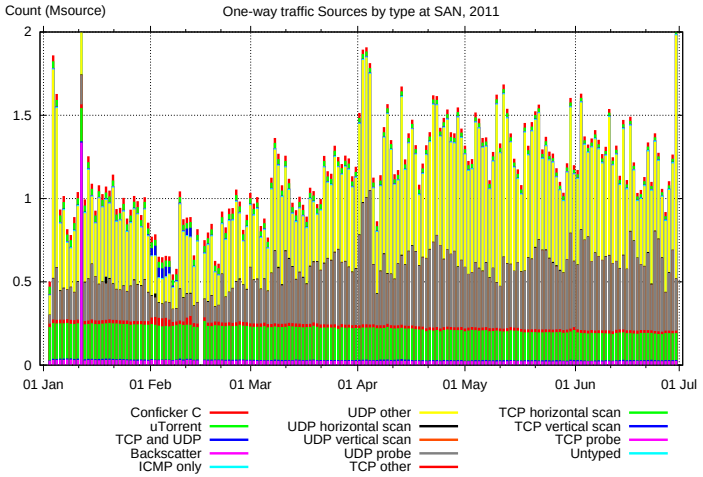
Figure 2 shows example distributions for each group. The centre column shows three groups that have a packet inter-arrival time mode at 3 s, the standard time for TCP retries. We also found that some UDP sources had IAT distributions suggesting an application with a 3 s retry time. These three group names end in *left, even* and *right* to show which side of their distribution's mode has more counts. The left-hand column shows *stealth* sources – those sending fewer than 120 packets, remaining active for more than 30 minutes, and having long ($>= 5\,minute$) quiet intervals. (IAT distributions are an effective way to detect stealth sources, in spite of their low average packet rates.) The right-hand column shows IAT distributions for three other groups: at the top, short-lived sources, active less than 30 minutes and sending less than 120 packets; in the middle, high-rate sources, sending packets almost back-to-back; and at the bottom, 'DoS' sources sending more than 10% of their of their packets in brief ($<= 15\,ms$) bursts.

4 Observations

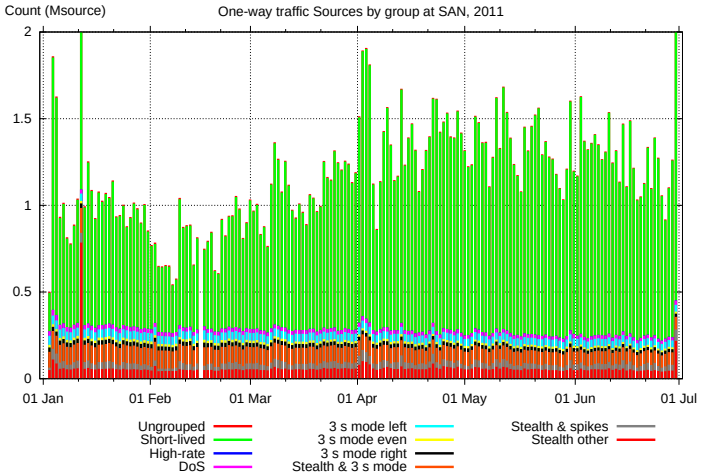
4.1 Long-term (six-month plots)

Figure 3 shows daily average total counts and volumes per hour as stacked bars for source types and IAT groups observed at the UCSD network telescope from 3 Jan 11 to 30 Jun 11. Over those six months, typically between 0.5 and 2.0 million sources were seen on average during a typical 6-10 GB/hour of one-way traffic. The number of unique source IP addresses (in millions) declined from 1.08 in mid-January to 0.65 at the beginning of February, then rose again to 1.37 by mid-April, after which there was no obvious trend in total source counts or MB volumes.

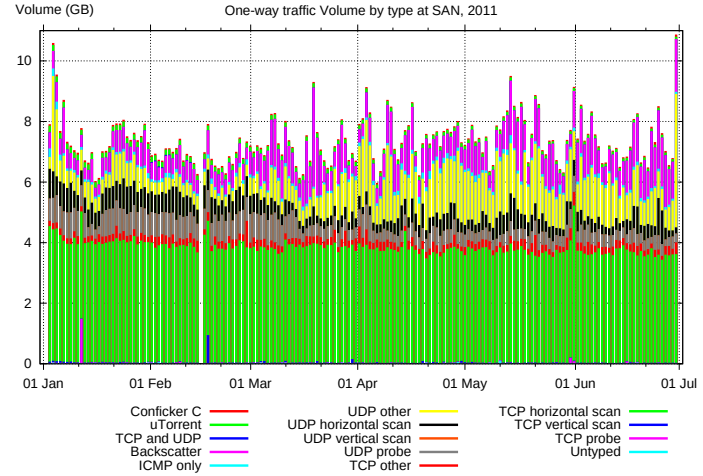
For source types (Figure 3(a)), the number of *TCP horizontal scan* sources in the daily average hour showed a steady decline while *TCP probe, TCP vertical scan* and *untyped* source numbers remained at steady, much lower values. *UDP probe* and *unknown* source numbers increased gradually from about 1 February.



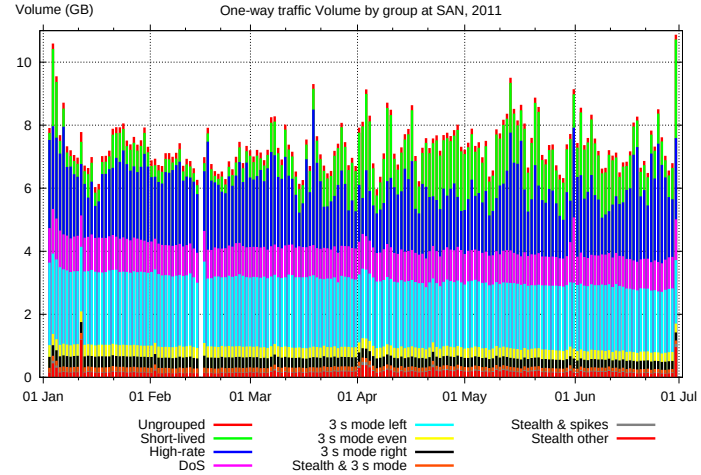
(a) Type counts (Millions of sources)



(b) Group counts (Millions of sources)



(c) Type volumes (GB)



(d) Group volumes (GB)

Fig. 3: Daily average total counts and volumes per hour for source types (left) and source groups (right); 3 Jan – 30 Jun 2011, UCSD Network Telescope

Apart from the gradual long-term changes, Figure 3(a) also shows short-lived activity, for example the huge spike in *TCP probes* on 12 January, the increase in *TCP* and *UDP* during the first week of February, and the doubling of *UDP probes* in the first week of April.

Figure 3(c) shows the daily average hour's traffic volume (in GB) for the source types. *TCP horizontal scan* traffic accounted for about half the total volume, around 4 GB/h. *UDP probe* traffic decreased to about 0.2 GB/h, which seems surprising since the number of *UDP probe* sources increased. Also, there was about 0.1 GB of *UDP horizontal scan* traffic each hour, even though the number of *UDP horizontal scan* sources was too small to be visible in Figure 3(c).

Considering source group activity, Figure 3(b) shows that short-lived sources dominate, accounting for about 3/4 of the sources observed. The number of *stealth 3 s mode* sources (TCP and UDP) declined steadily from 100 M to 30 M sources during these six months, while the number of sources in the other groups remained steady. On 12 January we saw 78,000 *stealth other* sources, these correspond with the spike in *TCP probes* in Figure 3(a). Similarly, in the first week of April we saw a rise in *stealth other* sources, corresponding with a similar rise in *UDP probes*. Traffic volumes for the source groups are shown in Figure 3(d). Although the lower four bands on the plot correspond well with their group counts (Figure 3(b)), the next three groups – *3 s mode left*, DoS and high-rate sources – account for up to 80% of the total traffic volume. Again, the *3 s mode left* volumes declined from 2.5 to 1.8 GB/h over the six months, corresponding with our observation of the fall in *TCP horizontal scans*.

5 Conclusions and future work

Building on the demonstrated utility of the activity classification scheme in Treurniet [12], we have developed a taxonomy for one-way traffic sources using two independent classifying schemes: 14 source *types* and 10 *IAT groups*. These schemes separate the one-way traffic into 140 subsets, allowing us to determine which source subsets were active during any hour, and to track subset behaviour over weeks or months as the characteristics of one-way traffic evolve. Using these subsets, we found that:

– Long-term plots of *type* and *group* subsets indicate distinguishable changes in the proportions each *type* or *group* contributes to total traffic.

– For the first six months of 2011, although total daily average one-way traffic into the UCSD network telescope did not increase significantly, the composition of per-source traffic behavior has changed, with an increase in *TCP horizontal scan* and corresponding decrease in *stealth 3 s mode* sources. Although we see many stealthy (long-lived low-rate) sources, most of the telescope traffic comes from short-lived ($< \frac{1}{2}$ hour) sources.

– We have used our `iatmon` tool to apply this taxonomy to hourly trace files, its implementation runs at least fast enough to be used on a live 1 Gb/s network. `iatmon` also includes tools to extract trace files for sources in *type*+*group* subsets that show interesting behaviour. Such 'source trace' files can then be examined

in detail, for example to determine whether they represent traffic from known malware.

In the future, we plan to develop `iatmon` so that it can detect significant changes in source subset counts or volumes, experiment with the thresholds in our IAT group classification scheme, investigate other possible classification schemes (perhaps as 'plug-in modules' for `iatmon`), and most importantly, explore a cooperative global effort to compare unsolicited traffic across a wider diversity of address space. We are currently operating `iatmon` in real-time mode on the University of Auckland's production Internet gateway, a 1 Gb/s link carrying about 70,000 packet/s; we find that 3% of the total traffic inbound to the University each hour is one-way, consistent with another recent study of this link [7].

Acknowledgment: Thank-you to the anonymous reviewers, and to my colleagues at CAIDA for their helpful suggestions for improving this paper.

This material is based upon work supported by the National Science Foundation under Grant No. 1059439.

References

1. E. Aben. Conficker as seen from UCSD Network Telescope, Feb 2009. `http://www.caida.org/research/security/ms08-067/conficker.xml`.
2. P. Barford, R. Nowak, R. Willett, and V. Yegneswaran. Toward a Model for Source Address of Internet Background Radiation. In *Proc. Passive and Active Measurement Conference, PAM '06*, Adelaide, Australia, 2006.
3. CAIDA. Ucsd network telescope data use policy and request form. `http://www.caida.org/data/passive/telescope_dataset_request.xml`.
4. CAIDA. UCSD Network Telescope global attack traffic. `http://www.caida.org/data/realtime/telescope/`.
5. CAIDA. UCSD Network Telescope Research. `http://www.caida.org/data/passive/network_telescope.xml`.
6. E. Cooke, M. Bailey, Z. Mao, D. Watson, F. Jahanian, and D. McPherson. Toward understanding distributed blackhole placement. In *Proc. ACM workshop on Rapid malcode, WORM '04*, pages 54–64, Washington DC, USA, 2004.
7. D. Lee and N. Brownlee. Passive Measurement of One-way and Two-way Flow Lifetimes. In *ACM SIGCOMM Computer Communication Review*, 2007.
8. A. Loewenstern. DHT protocol, 2008. `http://www.bittorrent.org/beps/bep_0003.html`.
9. D. Moore, C. Shannon, D. Brown, G. Voelker, and S. Savage. Inferring Internet Denial-of-Service Activity. *ACM Transactions on Computer Systems*, May. 2006.
10. R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of Internet Background Radiation. In *Proc. of the 4th ACM SIGCOMM conference on Internet Measurement, IMC '04*, pages 27–40, Sicily, Italy, 2004.
11. P. Porras, H. Saidi, and V. Yegneswaran. Conficker C P2P Protocol and Implementation. In *SRI International Technical Report*, 21 Sep 2009. `http://mtc.sri.com/Conficker/P2P/`.
12. J. Treurniet. A network activity classification schema and its application to scan detection. *IEEE/ACM Transactions on Networking*, 19(5), 2011.
13. wiki.theory.org. Bittorrent protocol specification v1.0, 2006. `http://wiki.theory.org/BitTorrentSpecification`.

14. 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*, IMC '10. ACM, 2010.