The Use of Packet Inter-Arrival Times for Investigating Unsolicited Internet Traffic

Jacob Zimmermann, Andrew Clark and George Mohay*
Fabien Pouget and Marc Dacier†

Abstract

Monitoring the Internet reveals incessant activity, that has been referred to as background radiation. In this paper, we propose an original approach that makes use of packet Inter-Arrival Times, or IATs, to analyse and identify such abnormal or unexpected network activity. Our study exploits a large set of data collected on a distributed network of honeypots during more than six months. Our main contribution in this paper is to demonstrate the usefulness of IAT analysis for network forensic purposes, and we illustrate this with examples in which we analyse particular IAT peak values. In addition, we pinpoint some network anomalies that we have been able to determine through such analysis.

1: Introduction

Unsolicited traffic on the Internet includes malicious traffic which may be caused by a variety of malicious software (malware) as typified by worms [1] and botnets [2] and it also includes benign traffic which may be due to a number of different circumstances, including for example software misconfiguration. This unsolicited traffic has been referred to collectively as background radiation by Pang Ruoming et al in [3]. There are broadly speaking two approaches that have been used for studying this traffic, the first approach using honeypots, the second using so-called telescopes and darknets. Both approaches may analyse traffic at either the packet header level or at the payload level or both. It is noted that there are advantages in using traffic behaviour (relying upon packet header information only) rather than code signatures (relying upon packet payloads), one reason being robustness in the face of obfuscation [2], another being efficiency. In this paper we focus on the analysis of packet inter-arrival times (IATs) and the use of a honeypot network to investigate and identify unsolicited Internet traffic.

The hypothesis we address is that IATs provide a characteristic fingerprint of the processes and tools being used to target the honeypot network and are thus of value in network forensics. At the very least we wish to determine if the information provided by an analysis of IATs may provide additional insights into the nature and identification of the attack tools being used.

* Information Security Institute, Queensland University of Technology, GPO Box 2434, Brisbane Queensland, Australia 4001
{a.clark, g.mohay, j.zimmerm}@qut.edu.au

† Institut Eurécom, BP 193, F-06904 Sophia Antipolis cedex, France
{marc.dacier, fabien.pouget}@eurecom.fr

The remainder of this paper is organized as follows. Section 2 discusses previous work relating to the investigation of unsolicited Internet traffic, while Section 3 describes the experimental environment used in this work and previous work by two of the authors on the use of that environment for this purpose. Section 4 discusses packet inter-arrival times (IATs) and presents an overview of our results in analysing IATs across the honeypot network. Section 5 then examines in detail two IAT peaks that we have found in order to demonstrate the value of IAT analysis for network forensics. Section 6 concludes the paper and identifies continuing and further work relating to this research.

2: Related Work

2.1: Honeypots and Telescopes

As mentioned above, there are broadly speaking two approaches that have been used for monitoring unsolicited Internet traffic.

The first approach uses honeypots. A honeypot is a node whose existence is not publicised in any way to the wider Internet community and which does not solicit any traffic: it should therefore experience no traffic of any kind. In practice, however, a honeypot will observe many different types of traffic which we describe in further detail below. The use of honeypots was advanced many years ago [4] both to serve as decoys and also for the purpose of facilitating the analysis of Internet attack traffic by confining the traffic to be analysed to that which is solely illegitimate. Studying the traffic on a bona fide production system is vexed by the extent of legitimate traffic, which tends to obfuscate the presence of any attack traffic. The term *honeynet* as originally defined [5] signifies a network of standard production systems so-called full honeypots - intended to be attacked and whose value lies in the limitless interactivity of a range of real production services. A honeynet may be both logically and geographically dispersed [1]. With time, the concept has become more widely used and is now taken to signify any group or collection of honeypots. A honeynet may be a local network artefact or it may be a distributed collection of honeypots whose traffic is correlated in some way in order to identify trends across the Internet. The original work in this area is that of Spitzner [5] while Pouget and Dacier [6] have more recently developed a network of honeypots, the *Leurre.com* project, a world wide honeypot network for attack detection involving over 30 honeypots in 20 countries.

The second approach is to use an Internet telescope and to employ unused subnets within a locally allocated address space, so-called darknets, in order to monitor all traffic directed to these unused addresses. The traffic can then be processed and analysed as suits. The larger the locally allocated address space the greater is the likelihood of unused subnets of significant size. Pang Ruoming et al [3] for instance report using parts of a Class A network, two /16 nets and ten /24 nets in this way. A variation on monitoring darknet traffic on entry is to do egress monitoring viz., in the case of a darknet which is not 100% passive and which emulates some services, in addition to ingress monitoring there can be egress monitoring of both packet headers and packet payloads and this can assist in worm detection and identification.

There are other approaches too (such as *DShield*) that likewise analyse Internet traffic and which do not fall naturally into either of the above two categories, these are discussed in Section 2.2 below.

2.2: Unsolicited Internet Traffic

There have been a number of hallmark studies of Internet background radiation in the past few years. A well known phenomenon and yet one not entirely well understood are the diurnal patterns of activity, see for instance [7].

An important early work in 2001 was that by Moore et al. [8] on the study of Internet DoS activity focusing upon the analysis of what was termed backscatter which is traffic generated by a real target in unknowing response to a source-spoofed communication. Moore et al. [8] used a lightly utilised /8 network for their study.

In 2004, Cooke et al. [9] report on the non-uniformity exhibited by background radiation across the global Internet and in the same year Moore et al. [10] reported a mathematical analysis of the relationship between the size of a network telescope and the network events that can be detected and with what precision. Also in 2004, Pang Ruoming et al. [3] and Vanderavero et al. [11] reported on the need for traffic filtering and the use of stateless emulated services to reduce the load on Internet monitors or measurement systems.

Pang Ruoming et al. [3] describe use of the *iSink* system and the similar LBL Sink system for traffic capture and analysis, with a particular focus on the filtering techniques used to cope with the volume of traffic and the nature of the stateless active responders used by the two systems.

Pouget and Dacier have recently developed a network of honeypots [6], the *Leurre.com* project, a world wide honeypot network for attack detection involving over 30 honeypots in 20 countries. This approach solves the problem posed by the volume of traffic to be captured and analysed by Internet telescopes and the like. The volume of traffic to be dealt with is orders of magnitude smaller but still provides rich data which allows for in depth analysis of malicious traffic patterns [6, 12].

DShield is a post-hoc traffic analysis and attack detection system funded by *SANS* and forming part of its Internet Storm Center [13, 14]. It collects data about malicious activity from across the Internet and its aim is to identify widespread attacks. Today it gathers millions of intrusion detection (and firewall) log entries every day, from sensors covering over 500,000 IP addresses in over 50 countries. Symantec publishes a six-monthly Internet Security Threat Report in which they report amongst other things on attack trends. The latest report [15] notes the growing threat posed by remotely controlled bot networks. The analysis in that report utilises data gathered by the DeepSight Threat Management System [16], an early warning system which collects IDS and firewall events from over 20,000 sensors and 180 countries.

Yegneswaran et al. [17] describe an analysis of four months of DShield data (August 2001 and May - July 2002) from over 1600 sources comprising 5 Class B networks, over 45 Class C networks and other smaller subnets.

Previous work by two of the current authors [6] used a clustering technique to analyse network traffic on three co-resident honeypots in order to characterize the root causes of attacks on those honeypots. That work introduced the concept of port sequences - the sequence of ports accessed by traffic within a cluster was found to exhibit uniformity although they found that the same port sequences would occur across separate clusters (clustering features included source IP address, number of honeypots targeted, number of packets etc).

In this paper, we focus on investigating and characterizing Internet traffic using Inter-Arrival Times. Indeed, IATs are very easy to monitor and, compared with payload-based techniques, their processing is lightweight and requires few CPU and memory resources.

They may thus be useful as a first-stage anomaly detection technique.

There has been little work reported on this subject. Debar and Lefranc [18] report on the use of information from web server logs to characterize the activity of the *Nimda* worm by examining the burst characteristics of the associated *tftp* requests. Zhou and Lang [19] have used discrete Fourier transforms to analyse time series data from the synthetic network intrusion data of the DARPA datasets and have identified a number of attacks in that dataset on the basis of packet IATs. Our work differs from both of the above in that we deal with a large volume of real Internet traffic across a large network of honeypots and process all of that traffic at the packet level in order to characterize Internet background activity as a whole.

3: The Experimental Environment and Dataset

The work described in this paper builds upon previous work by two of the authors in the use of low interactivity honeypots for worm detection and identification using a dataset obtained from a set of honeypot platforms (see [20] for details). This distributed honeypot environment currently consists of 30 platforms located in 20 different countries, covering the 5 continents. Some of the platforms have been active for more than 2 years. All in all, some 902600 different IP addresses have been observed over the whole period. Those addresses originate from 185 different countries. Note also that only a small fraction of those addresses have been observed twice, that is on two different days.

Each platform emulates three virtual machines running different operating systems (Windows 98, Windows NT Server, Linux RedHat 7.3) with various services. Traffic data from the platforms are centralized in a database that contains a large variety of information, such as:

- Raw packets (entire frames including the payloads are captured with tcpdump);
- IP geographical localization obtained with MaxMind;
- Passive Operating System fingerprinting obtained with Disco, p0f and ettercap;
- TCP level statistics using TCPstat;
- DNS reverse lookup, whois queries, etc...

It is this database which serves as the dataset for the work described in this paper.

Previous work by two of the authors [6, 21] focused on a clustering technique to identify the tools behind the observed attacks using a clustering algorithm detailed in [6]. The work described in this paper in contrast focuses on the extent to which packet inter-arrival times (*IATs*) are characteristic of attack tools and background traffic and correspondingly the forensic value of packet inter-arrival times for investigating Internet traffic. Two aspects of the cluster-based analysis described in [6] are relevant to the current IAT-based analysis, and we therefore mention them here:

- *Attack Source*: One IP address that targets a honeypot platform on a given day. Thus, the same IP address seen in two different days corresponds to two different *attack sources* (see [6] for more details).
- *Ports Sequence*: An ordered list of ports targeted by an *attack source* on a virtual machine. For instance, if source A sends requests to port 80 (HTTP), and then to

ports 8080 (HTTP Alternate) and 1080 (SOCKS), the associated *ports sequence* will be {80;8080;1080}.

4: Packet Inter-Arrival Times (IATs)

4.1: Packets IATs in the Honeypot Environment

The clustering technique referred to above is founded on the concept of a sequence of one or more (usually more) packets from one source over a 24-hour period (= 86400 seconds) arriving at one of the virtual environments of a platform. For historical reasons we refer to such a sequence as a *tiny session* (as opposed to a *large session* which refers to all the tiny sessions there can up to three of these - targeting one platform over the same period). In this paper we will henceforth refer to a *tiny session* simply as a *session*. (There are good reasons for limiting the scope of such sessions to 24-hours e.g., the dynamic allocation of IP addresses, use of NAT, many ISPs re-allocate IP addresses every 24 hours, etc. In any case, as previously mentioned, only a negligible number of addresses have been observed to be persistent over more than 24 hours.)

While the cluster-based analysis above includes the average packet inter-arrival time as one feature amongst many in order to identify attack tools, here in this paper we address the more general question of the forensic value of packet inter-arrival times alone for investigating Internet traffic. Intuitively one expects that the IATs of packets from a particular source arriving at a honeypot will be determined by network latency, by the software at the source host which sends the packets, and also by possible non-deterministic effects at both the source and destination. Non-deterministic effects may affect IATs as follows:

1. non-deterministic effects will occur at the source if the sending program is interactive i.e., driven by a human, and
2. non-deterministic effects at the destination may affect the dispatching of a response packet by the destination and thereby affect the sending of any further source packets which rely upon prior receipt of those responses.

In our case, in the case of a non-interactive honeypot and in the absence of such non-determinism, a honeypot will see a sequence of n packets per session and the sequence of n-1 IATs in the session will then be determined by the sending program and supporting network software, and by network latency. Two important research questions are therefore as follows:

1. to what extent are such sequences of IATs characteristic of the sending program and system, and
2. how much use are such IATs in the forensic investigation of Internet traffic.

In the research described in this paper we address the latter point and present an analysis of *IATs and their distribution*. We have also commenced the IAT sequence analysis and results from that will be reported upon at a later time. As discussed below, we have observed repeated and apparently characteristic sequences of IATs which motivates this further work.

4.2: IAT Analysis

The dataset of 2195483 sessions captured at all the honeypot platforms at our disposal worldwide has been analysed as follows. The packets for the sessions are captured using tcpdump which also captures the local arrival time for each packet, and computing the IAT sequence for each session is then straightforward. IATs are computed with a resolution of one second.

We have carried out a simple frequency analysis of the computed IATs which has resulted in Figures 1 and 2. Figure 1 shows the distribution of sessions by IAT, for IATs in the range 0 to 86399 seconds. Figure 2 zooms in on the 0 to 10000 second range and shows the distribution of sessions by IAT but with IATs in that smaller range in order to allow us to examine the low end of the IAT distribution in more detail. Table 1 lists the ten most prominent peaks and their frequencies excluding peaks in the 0 60 seconds range. We have decided to postpone the study of these short sessions for further work and focus, in a first step, on those sessions where long IATs exist. The numbers shown in this table are calculated using a tolerance of +/- 1 seconds. Note that the peaks are quite narrow, typically with a width of only a couple of seconds, consistent with variations in network latencies of that order.

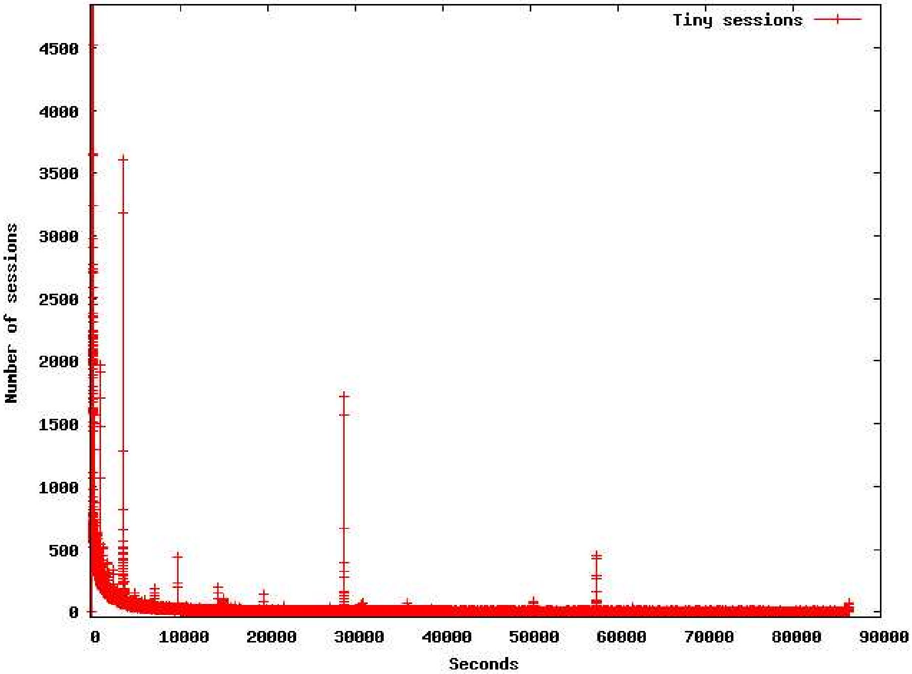

Figure 1. The distribution of IATs by Session

Nine of the peaks occur at various multiples of 300, and in addition we also see a peak at 9754 seconds. We expect that all these peaks are in some way characteristic of the source software. We examine two of the peaks in detail in Section 5 below, the one at 28800 seconds (8 hours) and the one at 9754 seconds (2h 42m 34s), for the reasons described there, and provide possible explanations for the phenomena which are their root cause to demonstrate the value of IAT analysis in network forensics.

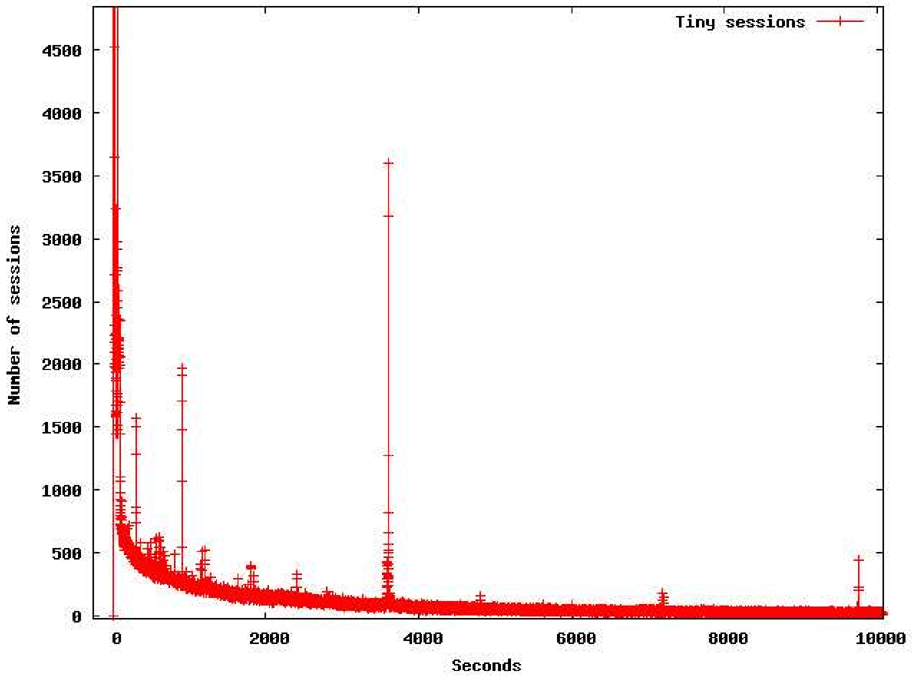

Figure 2. The distribution of IATs by Session in the 0-10000 seconds range.

4.3: IAT Interdependencies

In analysing the overall distribution of IATs, we undertook also a preliminary examination of the extent to which sessions contained only one, or conversely more than one, of the ten most commonly occurring IAT values. We exclude consideration at this stage of other, less frequently occurring, IAT values, and as noted earlier we also exclude IAT values in the 0-60 seconds range.

Table 2 shows the numbers of sessions which include specific IAT values. Diagonal elements indicate the number of sessions which include one or more occurrences of the particular IAT value, while off-diagonal elements indicate the number of sessions which include both IAT values. We observe that the off-diagonal values in Table 2 are generally much smaller than the diagonal elements, and that the vast majority of sessions (of the order of 90%) are characterized by only the one IAT value (of the ten most frequently occurring IAT values). In a few cases, we have quite large off-diagonal values (e.g., 183, 199, 215 - the 215 is especially noteworthy, given that it is relatively large compared to both of the corresponding diagonal elements) and this implies potentially characteristic IAT sequences or patterns, something we referred to earlier in the paper and which we have flagged for further work.

5: Investigations of Some Peak IAT Values

In this section we provide details of investigations into two of the peak IAT values observed in the global IAT distribution presented above. These two particular peak values are interesting for different reasons. The first peak value we investigate, the IAT of 28800

IAT value (+/- 1 second)	Number of tiny sessions
3600	5649
28800	2686
300	2676
900	2530
600	1408
1200	1048
57600	767
1800	707
2400	610
9754	490

Table 1. The ten highest IAT peaks

IAT (+/- 1 sec.)	57600	28800	9754	3600	2400	1800	1200	900	600	300
300	6	5	3	32	15	9	53	46	47	2676
600	0	8	1	33	9	12	16	54	1408	
900	0	0	0	17	14	3	22	2530		
1200	3	3	1	183	215	3	1048			
1800	0	0	1	13	7	707				
2400	3	2	0	199	610					
3600	3	9	0	5649						
9754	0	2	490							
28800	13	2686								
57600	767									

Table 2. IAT interdependence - numbers of sessions with specific IAT values

seconds, is equal to exactly eight hours and is longer than anticipated typical networking time-out or retransmission values and therefore of interest. The second peak value we investigate is the IAT of 9754 seconds (2 hours, 42 minutes and 34 seconds). Once again, this is not readily attributable to known timeout-triggered retransmission values and is hence of interest.

During the analysis of the global IAT distribution we observed that the peak values are not single-valued spikes, but rather are attributed to a set of IAT values in the vicinity of the peak. For this reason we group the data pertaining to the two peaks we investigate by including the sessions containing the three consecutive IAT values close to the peak. For example, for the peak value of 28800 seconds which we investigate below, we actually include occurrences of the IATs of 28800 ± 1, i.e. occurrences of 28799, 28800, and 28801. We attribute the variance in IAT values to variations in network latency across the network between the packet source and the honeypot and also to the precision of the timestamps associated with the time of arrival at the honeypot (timestamps are accurate to the second).

5.1: The IAT Peak at 28800 Seconds

We grouped the data corresponding to the IAT values 28799, 28800, and 28801 which occurred 772, 2093, and 2212 times, respectively, in the data set (see Figure 3). These 5077 occurrences of the IATs can be attributed to 2686 distinct sessions. The actual number of sessions containing these three IATs are 662, 1574, and 1718, respectively. Note that a number of sessions contain multiple IAT values from the set of three being investigated in this case.

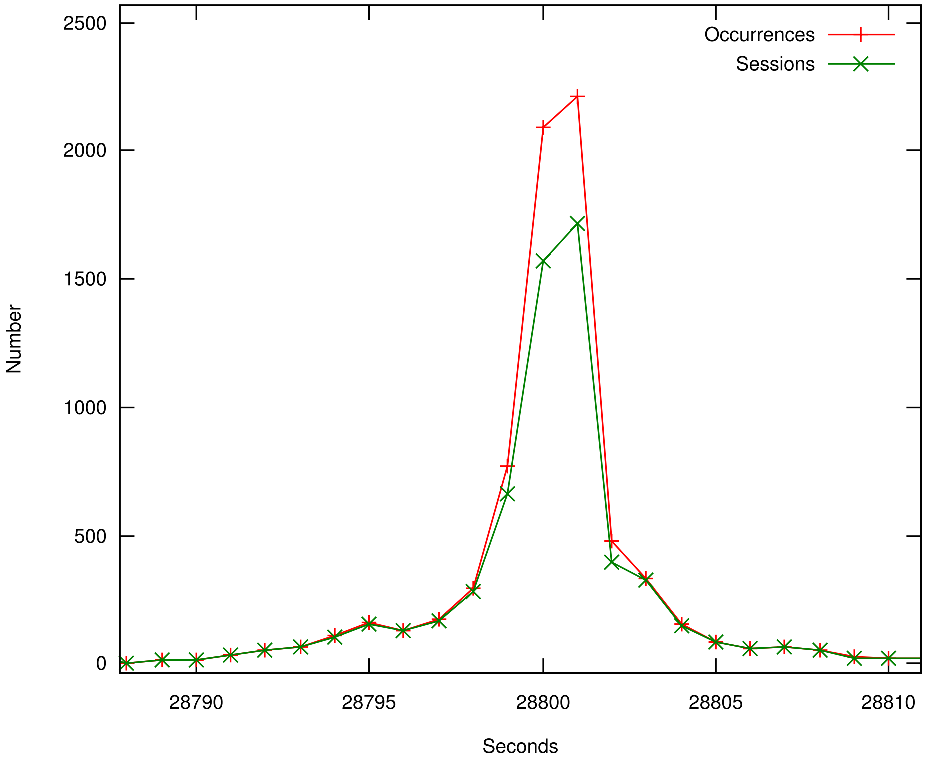

Figure 3. The peak at 28800 seconds.

The next step in the investigation was to determine if these sessions are similar in the ports they target on the honeypot. In this case we find that almost all of the sessions in the set of 2686 contain packets targeting UDP port 38293 (in fact 2674 out of 2686, or 99.6%, contain such packets). The total number of packets belonging to these 2686 sessions is 56975. Of these, 56731 (or, coincidentally, 99.6%) target the port UDP/38293. We also observe that many of the sessions contain exactly 18 packets (1890 sessions) or 36 packets (363 sessions). Thus, the vast majority of the sessions containing packets with an IAT of 28800 (or thereabouts) also contain traffic which appears to display some highly regular behaviour.

A search on the Internet reveals that the port UDP/38293 is used by the Norton Antivirus automatic update server [22]. The server, which is usually used within an enterprise environment, periodically scans the network looking for hosts running the Norton Antivirus update client. So why would we expect to see these servers scanning outside the enterprise boundary, and onto the Internet at large? We suspect that this traffic may be attributable to misconfigured Norton Antivirus automatic update servers (or at least to a network misconfiguration close to the location of the server). Interestingly, the vast majority of the sessions in which we found this particular IAT value were observed at a single honeypot environment (although a small number were observed at other environments).

At this point we investigated the source addresses to which this traffic could be attributed. We found that the 2686 sessions are caused by only 162 source IP addresses, and, in fact, 2189 (or 81.5%) are caused by 11 of those addresses. Of those 11, seven are located in the US, two in Great Britain, and one each in Japan and France.

Inspection of the payload of the 56731 packets targeting UDP port 38293 reveals that there are exactly two distinct payloads, and each payload appears the same number of times (actually, the difference is one, probably attributable to the loss of a single packet). In order to further understand this phenomenon the 1890 sessions which contain exactly 18 packets were examined in more detail. It was found that each of these sessions consisted of three bursts of 6 packets, with each burst separated by approximately 28800 seconds. Upon examination of the payloads of these packets it was found that in each burst of six packets 3 contained one of the observed payloads, and three the other. Given the homogeneity of the payloads, and the other regularities observed for these sessions, it is almost certain that each of these packets is attributable to the same software.

In this case while not identifying an attack, the analysis of IATs has provided us with a technique to differentiate a subset of the traffic directed at a particular port from the remainder of the traffic targeting that port. We can also note that honeypot usefulness seems thus to extend beyond the capture of malicious traffic and can be useful in identifying misconfigured or erroneous software.

To further highlight this we present some results from our investigation into all traffic targeting port UDP/38293. The total number of sessions containing packets targeting this port is 9663. The total number of packets targeting this port is 138601. Thus, we can see that a little under one half of all packets targeting this port can be found in sessions with an IAT close to 28800. Figure 4 shows the distribution of IAT values for the 9663 sessions containing packets targeting UDP port 38293.

It is interesting to note that traffic targeting this port is also responsible for the IAT peak at 57600 seconds (see Figure 5). Observe that this value is equal to two times 28800. A further analysis reveals that 100% of the traffic which features this IAT, that is 767 sessions, also target UDP/38293. Moreover, all these 767 sessions originate at 50 distinct source IP addresses, the most frequent being located in the US (226 sessions), France (206 sessions), Japan (166 sessions) and Australia (95 sessions). Among those source IP addresses, 8 are also used by the 28800 IAT traffic; these addresses correspond to 46 sessions featuring the 57600s IAT (i.e. 8%) and 872 sessions featuring the 28800s IAT (i.e. 32%). There are only 2 environments targeted by this traffic, located in the same geographical and economical area. Both are also targeted by the IAT 28800 traffic.

As with sessions observed with an IAT of 28800, the sessions with an IAT of 57600 predominantly consist of 18 packets (559 sessions in total). Moreover, 161 sessions have 12 packets, all other numbers of packets occur less than 10 times. An investigation of the sessions containing exactly 18 packets indicates that these sessions exhibit exactly the same behaviour as the ones with an IAT of 28800, namely that they consist of three bursts of 6 packets (three of each of the two payloads commonly observed) each separated by an IAT of around 57600 seconds.

Finally, these sessions feature 11 different packet payloads, 9 of which appear only once

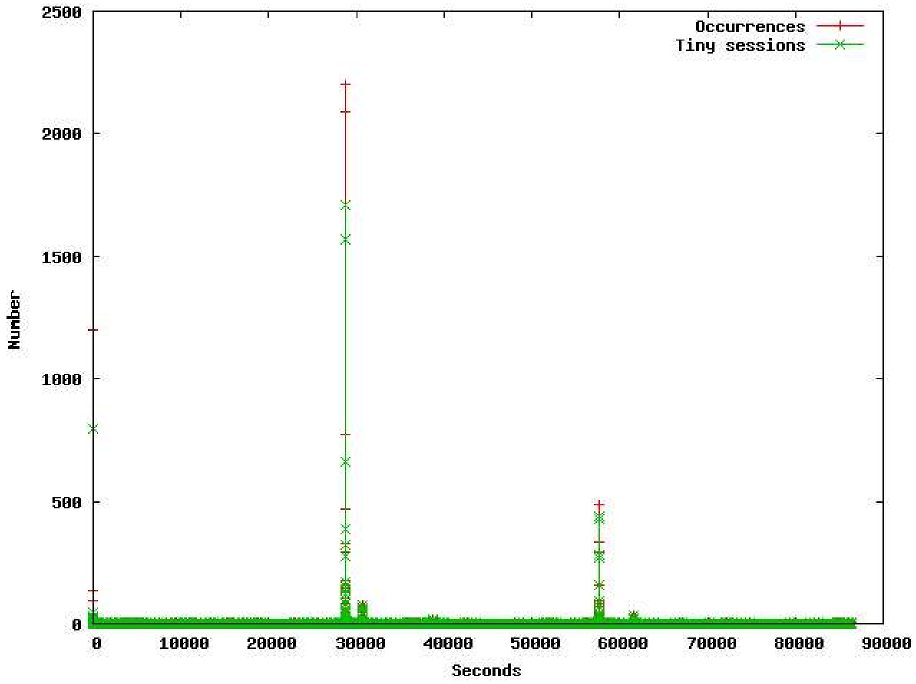

Figure 4. The IAT distribution of traffic targeting port UDP/38293.

(including 1 null payload.) The 2 remaining payloads, which appear respectively in 9971 and 9968 packets, are the only two payloads observed for peak at 28800 seconds. These two payloads are identical to the ones observed in the traffic with the IAT of 28800.

In summary, 49% of the traffic directed to UDP/38293 is characterized by these two IATs; the remaining 51% feature IATs more or less uniformly spread between 0 seconds and 24h. Even if the number of sessions featuring both 28800 seconds and 57600 seconds IATs is low, there is thus strong connection between these two peaks. Given all the observed similarities, it appears that these two peaks provide an appropriate signature for the observed traffic.

The above analysis has demonstrated the usefulness in general of global IAT analysis for characterizing network traffic, and indeed identifying the software which is the source of the traffic. IAT analysis provides a low overhead technique for identifying network traffic, a technique which stands alone without the overhead of painstaking payload analysis. In this particular case, it has provided the means for remotely identifying the suspected presence and location of misconfigured software. It has in particular in this case also identified a relationship between sessions characterized by IATs at 28800 seconds and sessions characterized by a harmonic at 57600 seconds. The precise nature and cause of this relationship is not clear and requires further investigation.

5.2: The IAT Peak at 9754 Seconds

The IAT peak that we observe around 9754 seconds (see Figure 6) is unusual. This is the most frequent IAT which does not correspond to a regular value (for example, a multiple of five minutes). This IAT corresponds to 2 hours, 42 minutes and 34 seconds.

We grouped the data corresponding to the three most frequent IATs in the vicinity of 9754, namely 9753, 9754 and 9755. They can be attributed to only 490 distinct sessions,

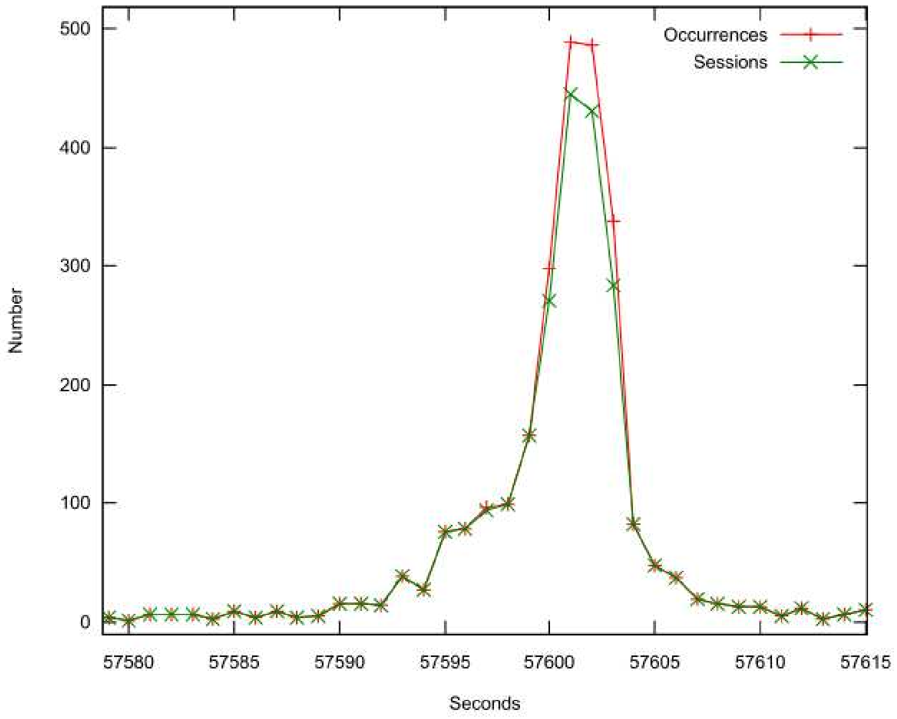

Figure 5. The peak at 57600 seconds.

indicating once again that this IAT is generated repeatedly by the same source against the same target honeypot. The actual number of sessions containing these three IATs are 226, 442, and 199, respectively. As before, a number of sessions contain multiple IAT values from the set of three being investigated. The first session containing this IAT was observed in November 2003, with the most recent in June 2005.

Similarly to the previous example, we next investigated the ports targeted by the packets in these 490 sessions. In this instance the majority of the sessions which feature this IAT include packets targeting UDP port 1026 (a total of 437 out of 490 sessions, or 89.2%). This port is used by Microsoft Windows operating systems for the Windows Messenger service, among other things. At least one worm is known to propagate via a vulnerability with this service [23]. This port is also known to be utilised for the distribution of spam over the Windows Messenger service [24]. The sessions that include this IAT are made up of a total of 6141 packets. The distribution of the target ports of these packets is given in Table 3 below.

In this case the percentage of packets targeting UDP/1026 is somewhat lower (only 60.6% compared to the 99.6% of packets targeting UDP/38293 in the previous example) but nevertheless it is still a significant proportion of the packets. A large number of these sessions are relatively small, containing between 5 and 9 (inclusive) packets (these sessions make up 66.1% of the 490 observed).

Once again, we next investigated the sources of these sessions. In this case almost all (427 out of 490, or 87%) occur from the same IP address in China. The remaining 52 source addresses of these sessions occur usually once, with a few exceptions (interestingly 35 of these 52 originate from the same class B subnet in Taiwan). Thus it appears that this

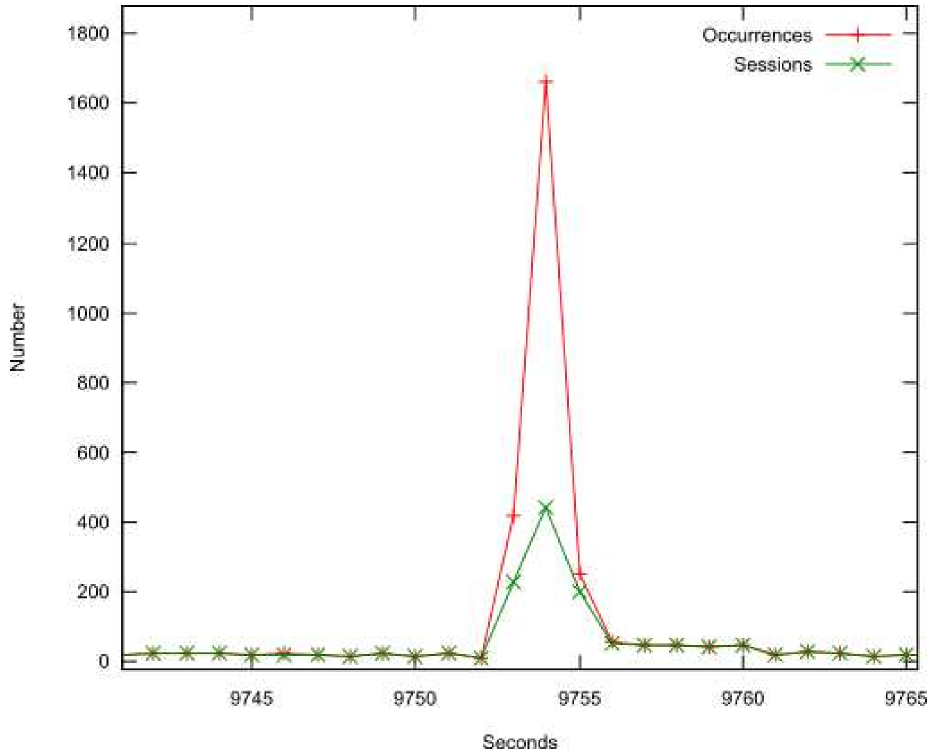

Figure 6. The peak at 9754 seconds

particular IAT can be attributed mainly to a single computer in China. Activity from this computer which exhibits the IAT of 9754 commenced in December 2004 and has continued until June 2005.

Of the 3724 packets from the sessions which target port UDP/1026, 3691 (or 99.1%) of them contain the same payload. All of the packets from the dominant source in China (a total of 3607) contain this particular payload. The remaining 84 packets containing this payload were generated from two distinct IP addresses, also located in China.

The particular traffic we have investigated represents a very small proportion of all the traffic captured on UDP port 1026 (only 3724 of a total of 186541 packets). However, as Figure 7 shows, in the IAT distribution for all sessions containing UDP packets targeting port 1026, there is a clear peak at the value 9754. Interestingly, 155212 of the packets targeting UDP port 1026 contain the same payload as that observed from the dominant source in China. In this instance the IAT analysis has identified a very particular subset of the traffic targeting this port.

6: Conclusions

There is an important volume of non-productive Internet traffic that can be passively collected by means of honeypots. Such traffic can reflect either malicious activities (due to worms, scans,etc) or benign ones (due to mis-configurations). While there has previously been a focus on capturing such traffic, there has been relatively little attempt to analyse it in toto apart from some use of classical statistics to do so.

In this paper, we propose the use of packet Inter-Arrival Times (IATs) to carry out such

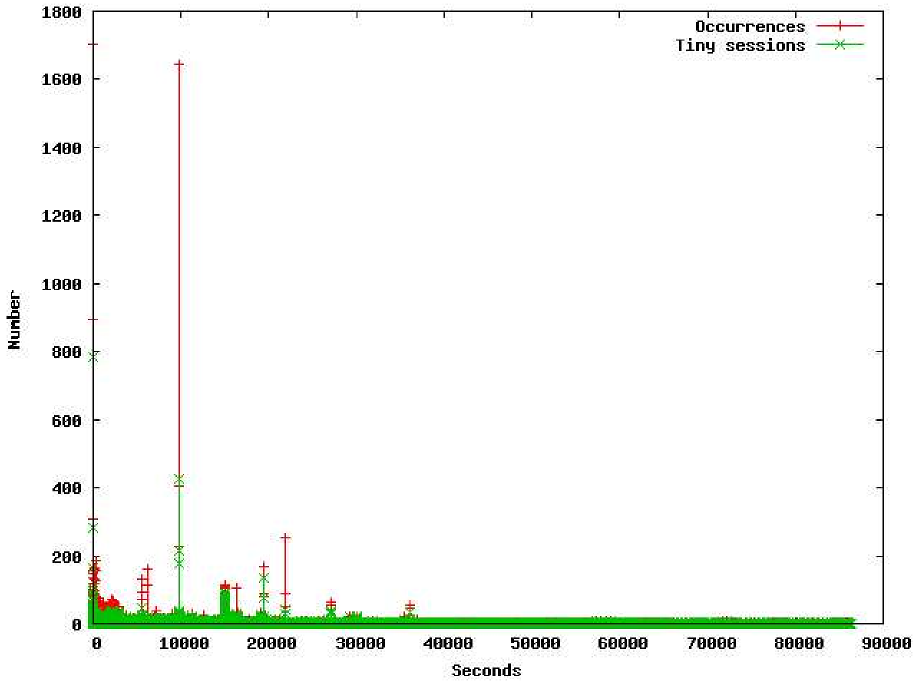

Figure 7. The IAT distribution of traffic targeting port UDP/1026.

analysis. Based on a significant dataset of real Internet traffic, we have demonstrated that IATs have the potential to provide a valuable contribution to network forensics. Specifically, we have analysed a large set of data collected from a distributed honeypot environment over more than 6 months. Our analysis has enabled us to identify some suspected anomalies as a result of the unexpected and very prominent IAT peaks that have been exposed. Two of these anomalies are detailed in the paper.

One novelty of this approach resides in the fact that honeypots, which are supposed to get very specific types of traffic, namely only attacks, also offer the kind of data that can help in understanding other phenomena observed on the Internet. In other words, the value of honeypots is not limited to the analysis of attack traces.

This preliminary study lets us also to foresee other IAT usages in order to enrich network forensics. Future work will consist in investigating some of them. We have observed that IATs tend to form repeated patterns and we are particularly interested in analysing such sequences. In particular, we intend to investigate to what extent IAT patterns can be used to identify attack tools or network phenomena.

Finally, we hope that the usage of the data will be a good incentive for some partners to share data and bring their expertise to the distributed honeypot project.

Acknowledgements

The above work was supported in part by a joint French-Australian Science and Technology (FAST) Programme grant and an ARC International Linkage grant. In addition to the authors there have been a number of other people involved at various times with aspects of the work described in this paper and their contributions are gratefully acknowledged: Saleh

Target Port	Number of Packets
ICMP	150
TCP/80	12
TCP/135	1677
TCP/139	62
TCP/448	354
TCP/1025	12
TCP/1433	6
TCP/2745	12
UDP/137	44
UDP/1026	3724
UDP/1027	88

Table 3. Distribution of target ports for packets in sessions containing an IAT of 9754 seconds.

Almotairi, Reda Tber and Van-Hau Pham.

References

[1] Jose Nazario. *Defense and Detection Strategies against Internet Worms.* Artech House, 2004.

[2] D. Geer. Malicious bots threaten network security. *IEEE Computer*, 38(1):18–20, 2005.

[3] Pang Ruoming, Vinod Yenheswaran, Paul Barford, Vern Paxson, and Larry Peterson. Characteristics of internet background radiation. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, Taormina, Sicily, Italy*, 2004.

[4] Lance Spitzner. *Honeypots - Tracking Hackers.* Addison - Wesley, 2003.

[5] Lance Spitzner. The Honeypot Project: Trapping the Hackers. *IEEE Security and Privacy*, 1:15, 2003.

[6] Fabien Pouget and Marc Dacier. Honeypot-based forensics. In *Proceedings of the Asia Pacific Information Technology Security Conference (AusCERT)*, 2004.

[7] D. Moore, C. Shannon, and K. Claffy. CodeRed: A Case Study on the Spread and Victims of an Internet Worm. In *Proceedings of ACM SIGCOMM Internet Measurement Workshop, Marseilles, France*, 2002.

[8] D. Moore, G. Voelker, and S. Savage. Inferring internet denial of service activity. In *Proceedings of 2001 USENIX Security Symposium, Washington D.C.*, 2001.

[9] Evan Cooke, Michael Bailey, Z. Morley May, Danny McPherson, David Watson, and Farnam Jahanian. Towards understanding distributed blackhole placement. In *WORM'04, Washington D.C.*, 2004.

[10] David Moore, Colleen Shannon, Geoffrey M. Voelker, and Stefan Savage. Network Telescopes: Technical Report. http://www.caida.org/outreach/papers/2004/tr-2004-04/tr-2004-04.pdf, 2004.

[11] Nicolas Vanderavero, Xavier Brouckaert, Olivier Bonaventure, and Baudouin Le Charlier. The HoneyTank: a scalable approach to collect malicious Internet traffic. In *Proceedings of the International Infrastructure Survivability Workshop 2004 (IISW'04), Lisbon, Portugal*, 2004.

[12] F. Pouget, M. Dacier, and V. H. Pham. Towards a better understanding of internet threats to enhance survivability. In *Proc. of the International Infrastructure Survivability Workshop (IISW'04), Lisbon, Portugal*, 2004.

[13] DShield: Distributed Intrusion Detection System. http://www.dshield.org.

[14] Internet Storm Center. http://isc.sans.org/about.php.

[15] Symantec. Internet threat reports. http://ses.symantec.com/content.cfm?articleid=1539, 2004.

[16] Symantec. DeepSight Threat Management System. http://enterprisesecurity.symantec.com/content/displaypdf.cfm?pdfid=301.

[17] V. Yengeswaran, P. Barford, and J. Ullrich. Internet intrusions: Global characteristics and prevalence. In *Proceedings of ACM SIGMETRICS 2003, San Diego, CA.*, 2003.

[18] Herve Debar and David Lefranc. Observations on the internet traffic reaching broadband-connected users. In *Proceedings of EICAR Conference, Copenhagen, Denmark*, May 2003.

[19] M. Zhou and S. D. Lang. Mining frequency content of network traffic for intrusion detection. In *Proceedings of IASTED International Conference on Communication, Network and Information Security (CNIS 2003)*, December 2003.

[20] The Leurre.com Home Page. http://www.leurrecom.org.

[21] F. Pouget, M. Dacier, and V. H. Pham. Leurre.com: On the advantages of deploying a large scale distributed honeypot platform. In *Proceedings of E-Crime and Computer Conference (ECCE'05), Monaco*, March 2005.

[22] Symantec Service Support. Ports used for communication in Norton AntiVirus Corporate Edition. http://service1.symantec.com/SUPPORT/ent-security.nsf/docid/2000101210181048.

[23] LURHQ and Corporation. Critical Microsoft Messenger Patch Released, October 2003.

[24] LURHQ and Corporation. Windows Messenger Popup Spam on UDP Port 1026, June 2003.