Parametric Differences Between a Real-world Distributed Denial-of-Service Attack and a Flash Event

Sajal Bhatia, George Mohay, Alan Tickle, Ejaz Ahmed
Information Security Institute, Queensland University of Technology
GPO Box 2434, Brisbane 4001, Queensland, Australia
{s.bhatia, g.mohay, ab.tickle, e.ahmed}@qut.edu.au

***Abstract*—Distributed Denial-of-Service (DDoS) attacks continue to be one of the most pernicious threats to the delivery of services over the Internet. Not only are DDoS attacks present in many guises, they are also continuously evolving as new vulnerabilities are exploited. Hence accurate detection of these attacks still remains a challenging problem and a necessity for ensuring high-end network security. An intrinsic challenge in addressing this problem is to effectively distinguish these Denial-of-Service attacks from similar looking Flash Events (FEs) created by legitimate clients. A considerable overlap between the general characteristics of FEs and DDoS attacks makes it difficult to precisely separate these two classes of Internet activity. In this paper we propose parameters which can be used to explicitly distinguish FEs from DDoS attacks and analyse two real-world publicly available datasets to validate our proposal. Our analysis shows that even though FEs appear very similar to DDoS attacks, there are several subtle dissimilarities which can be exploited to separate these two classes of events.**

***Keywords*—Distributed Denial-of-service (DDoS); Network Security; Flash Event; Botnet**

I. Introduction

Even decades after the first full-scale DoS attack in the form of high-rate flooding was unleashed [1], these attacks in various modes still continue to constitute a pernicious threat within the Internet domain [2]. A DoS attack is a malicious attempt by the attacker to make computing resources or services unavailable to intended clients. One form of this attack in which an array of geographically spread compromised machines (aka bots, zombies, agents, etc.) is controlled by the bot-master and used against some target host(s) to cause DoS is called Distributed Denial-of-Service (DDoS) attack [3]. A set of these compromised machines or bots is called a Botnet.

A majority of DDoS attacks attempt to bring down the victim host or server by consuming most of its available bandwidth or other computing resources like memory, CPU etc. Some of these attacks use IP spoofing to conceal the attacker's identity while some use genuine IP addresses of the compromised machines. The former attack methods are relatively easier to detect as compared to the latter in which the request or attack packets look very similar (or may even be crafted) to the ones send by legitimate users.

An early 1970's short science fiction story defined *Flash Crowd* as the situation when thousands of people went back in time to see historical events again [4]. In today's online world, the term *Flash Event (FE)* is used to describe a similar situation when tens and thousands of valid users concurrently access a computing resource. This sudden burst of legitimate traffic is often due to a newsworthy sports event like the Olympics or new product release by companies like Apple, Microsoft etc. Victoria's Secret [5] webcast is one of the popular examples of a FE.

Sometimes, the occurrence of such events is known well in advance like a product release by Apple. These events can be termed as *predictable flash events*. However, there can be situations when the web servers do not have any prior knowledge and are not expecting a traffic surge. Such events can be labeled as *unpredictable flash events*. Some examples can be news and medical web servers experiencing high traffic after a natural disaster (floods, earthquakes) or a terrorist attack (9/11), or the onset of an epidemic (e.g. swine flu) respectively. The *Slashdot Effect* or *Slashdotting* is another type of FE when popular websites like Slashdot, Digg etc. comprise references or links to other websites, thereby causing a massive increase to their incoming traffic. If the referenced links are less well provisioned, the traffic could easily exceed the available bandwidth and render the site temporarily unreachable.

Notwithstanding the amount of research done in the past decade, accurate detection of Denial-of-Service attacks continues to remain a difficult problem. FEs make this problem even more difficult to solve by sharing many similar characteristics [6]. Both DDoS attacks and FEs represent *anomalies* in the Internet traffic causing overloading of servers and rendering them less responsive. From the network administrators point of view, a precise distinction between DDoS attacks and FEs is very important because these events require an entirely different set of actions to be undertaken after they have been sucessfully identified. Detection of a FE requires an increase in the number of CDN's (Content Distribution Networks) and various load sharing mechanisms to accommodate more legitimate users [6]. Whereas, in case of DDoS attack, the network administrator would like to enable various attack mitigation mechanisms to filter out the maliciuos traffic and continue uninterrupted access to the legitimate clients.

In this paper, we seek to identify a set of parameters

which can differentiate a real-world DDoS attack from an FE. Subsequently, we perform a detailed analysis of two publically available datasets representing a real-world DDoS attack and a FE to verify our approach. The datasets used for this purpose are the CAIDA "DDoS Attack 2007" Dataset [7] and the "1998 FIFA World Cup" Dataset [8], a DDoS attack and an FE respectively. The CAIDA "DDoS Attack 2007" Dataset (referred to as CAIDA in the rest of the paper) contains one hour of anonymised traffic from a DDoS attack that occured on August 4, 2007. The "1998 FIFA World Cup" Dataset (referred to as World Cup in the rest of the paper) contains one way traffic from anonymized sources to 33 different web servers for a 92 day period from April 30, 1998 to July 26, 1998.

The rest of the paper is organized as follows: Section 2 gives an overview of recent relevant work in the field of DDoS attack detection and its differentiation from FEs. Section 3 describes the two datasets used to validate the proposed distinguishing parameters. Section 4 describes our proposed parameters which can separate the aforementioned events. Section 5 presents the experimentation and evaluation results. Finally, Section 6 concludes and summarizes the research and presents future directions of research in this area.

II. Related Work

Accurate detection of DDoS attacks remains a tough challenge. Attackers mimicing normal Internet traffic further exacerbates the problem. DDoS detection has predominantly followed two broad directions: network traffic analysis and MIB (Management Information Base) data analysis.

The network traffic analysis method uses various theshold based techniques like observing the total traffic coming in and going out of an edge router [9], and counting the number of packets in a given time frame [10]. TCP/IP header analysis like ratio of incoming and outgoing packets [11], IP flow analysis [12], entropy detection [13], activity profiling [14] are some of the other techniques used. Peng et al. [15] used history based source IP address filtering at an edge router to detect DDoS attack. The proposed mechanism maintains a historical database of all the valid IP addresses i.e. those completing the three-way TCP handshake. The database is updated using a sliding window in order to store the most recent IP addresses. Whenever the edge router gets overloaded, the IP Address Database (IAD) is used to decide whether to accept the incoming packets. During an attack, only the packets originating from source IP addresses present in the database are allowed access. However, the IAD can be corrupted by those source IP addresses which first complete a three-way handshake and later on participate in the attack. A similar work done by Ejaz et al. [16] uses rate of arrival of new source IP addresses in combination with the change point analysis technique to detect the occurrence of a DDoS attack.

The second area for DDoS detection that has evolved over past few years has been statistical analysis of MIB (Management Information Base) data collected via SNMP agents. This integrates existing detection systems (like an Intrusion Detection Systems (IDSs)) with SNMP-based Network Management Systems (NMSs) to detect the onset of a DDoS attack [17] based on server load or memory utilisation. Bao [18] and Yu [19] proposed a fast, lightweight, hierarchical SVM based DDoS attack detection mechanism. The proposed mechanism could efficiently detect both known and novel attacks. It is based on analyzing the security status of the network by using the MIB data gathered by the SNMP agents. The proposed mechanism can distinguish attack traffic from normal traffic and characterize the attack data as TCP-SYN flood, UDP flood and ICMP flood. The results obtained from analyzing MIB data in conjunction with IDSs were comparable to results achieved by analyzing the traffic data.

There are potentially a few techniques for differentiating bots and humans which could be applied for separating the two class of events in which we are interested viz. DDoS attacks and FEs. Amongst all available techniques, CAPTCHAs [20] have been heavily used. CAPTCHAs are graphical puzzles designed on the premise that humans can solve them but machines or bots cannot. Kandula et. al [21] used these puzzles to build a system based on probabilistic authentication to protect a web server. However, the use of such puzzles introduces additional delays for legitimate clients. These grahical puzzles were effective until puzzle-breaking mechanisms were developed which made it possible for compromised machines or bots to break the visual CAPTCHAs using various automated methods [22].

In some recent work on distinguishing DDoS from FE, Hyund et al. have reported the use of randomness checking [23]. In both [23] and [6], the number of incoming requests increases dramatically both in case of a DDoS attack and an FE. Both [23] and [6] are based on datasets of predictable FEs not available in the public domain. The World Cup dataset used in our work described in this paper shows that even in case of a predictable flash event it can take some time for the FE traffic to build up and the intensification in the incoming network traffic to the victim can be gradual rather than abrupt. The author conjectures that in case of an unpredictable flash event this parameter might behave differently still. The lack of datasets representing FEs, both predictable and unpredicted, contributes to such speculation. One of the future works of the conducted research is to attempt a classification of various types of flash events.

Yi et al. [24] suggested a novel hidden semi-Markov based anomaly detector for detecting shrew HTTP flood attacks from Flash Crowds. A *Shrew HTTP flood attack* is a type of application level DDoS attack which is a combination of a shrew attack and HTTP flood. A *shrew attack* is a low-rate DoS attack which sends "legitimate appearing" requests at a sufficiently low rate to elude detection by counter DoS techniques [25]. Their proposed technique used Principal Component Analysis (PCA) and Independent Component Analysis (ICA) to abstract the multivariate observation vector. However, their proposed scheme was unable to differentiate and separate the malicious sources from legitimate. Thus it only served as an alert function to more complex monitoring mechanisms.

The mechanism proposed by Le et al. [26] can only be used to distinguish FE from spoofed DoS attacks. Yu et al. [27] proposed a technique to distinguish between a DDoS attack and Flash Crowd based on flow similarity. They define *flow* as the packets which are passing through the same router and have a common destination address. Their technique is based on the premise that there is a stronger similarity between the flows of a DDoS attack as compared to those in flash crowds. They used three abstract distance metrics (Abstract distance, Jeffrey distance and Sibon distance) to measure and compare the flow similarity among DDoS and flash crowd flows based on the parameter viz. number of packets (presumably packets per flow). Their research shows that Sibon distance is the most suitable metric to measure the flow similarity. When tested on real-datasets, the proposed algorithm produced a differentiating accuracy of around 65%. In this paper a detailed analysis of two real world datasets has resulted in identification of additional parameters (see IV) that can be used to effectively differentiate between FE and DDoS.

III. Datasets

The research conducted is validated using two real-world network traffic traces. Each of the traces used in this research is available in the public domain and contains information relating to time and pseudonymized source and destination IP addresses.

For a real-world DDoS attack, we used the CAIDA "DDoS Attack 2007" Dataset [7]. This dataset contains pseudonymized traces from a DDoS attack that occurred on August 4, 2007 for approximately an hour (20:50:08 UTC to 21:56:16). The attack represents a DoS attack where the attacking sources try to consume all of the available networking and the computing resources of a target host by sending huge amount of access requests. The dataset contains the attack traces to the victim and its responses. The traffic traces have been pseudonymized using CryptoPAn prefix-preserving pseudonymization using single key and the payloads have been zeroed.

The "1998 FIFA World Cup" Dataset [8], provided by the Internet Traffic Archive, is the FE dataset used in our work. This dataset contains all the requests made to the 1998 FIFA World Cup websites during a period of 92 days (April 30, 1998 to July 26, 1998). During this period, there were 33 different web servers hosting information related to the World Cup. These servers were located at four different geographical locations: Paris (France), Plano (Texas), Herndon (Virginia) and Santa Clara (California). The time on each server was synchronized with the local time in France (GMT + 0200), which was the host country. The source IP addresses in the traffic traces have been replaced by unique integer identifiers which are preserved throughout the dataset. To ensure privacy, the mapping file for the source IP address pseudonymization is not publicly available. Table I summarizes the macro-level statistics of the two datasets used in this research.

Table I
Macro-level statistics of CAIDA and World Cup datasets

Parameters	CAIDA	World Cup
File format	pcap	common log format
Packet/Request type	ICMP ECHO	HTTP GET
Number of target(s)	1	33
Activity type	DDoS attack	FE
Total number of packets sent	359,655,826	1,352,804,107
Total duration	66 min	92 days
Total capture size (uncompressed)	5.3 GB	8.1 GB

IV. Paramterizing DDoS Attacks and Flash Events

The term DDoS is often used interchangeably with high-rate flooding attacks. Common examples of DDoS or high-rate flooding attacks are TCP SYN flooding and HTTP GET request flooding. A Flash Event (FE) is a sudden surge in incoming traffic to a web site, thereby rendering it over-loaded and congesting the network links leading to it. These FEs are caused by legitimate clients as compared to DDoS attacks which are carried out by compromised machines. The DDoS attacks can be crafted to look very similar to FEs, only *intent* separates them and not the *content*. However, this criterion does not help in differentiating between these two events. In order to exploit the behavioral differences between these two class of events, we propose the following parameters. The parameters while not completely orthogonal capture different aspects of the traffic.

A. Change in Rate of Incoming Traffic

In DDoS attacks, the attacking machines or bots are often infected by malware which is programmed to send packets at a pre-defined rate. In order to cause the maximum damage, the bot-master triggers the compromised machines simultaneously. On receiving commands from the bot-master, the bots start sending huge amounts of data. Hence, the victim server often experiences a sudden burst in incoming traffic over a relatively short period of time. This unexpected surge causes the server to exceed its maximum pre-defined sustainable limits, thereby slowing it down considerably and in some cases forcing it to shut down. In contrast to this, whenever a newsworthy event occurs triggering a FE, it is highly unlikely that the entire web user community is simultaneously informed. Instead, it takes times for the news to spread among people. Therefore, the incoming traffic to the web server gradually increases over time before hitting a maximum. Hence, it should be possible to use the difference in the rates of incoming traffic can be used to differentiate DDoS and FE.

B. Change in Rate of New Source IP Addresses

DDoS attacks are usually carried out by a finite set of compromised machines also known as a Botnet. Recent studies conducted on Botnets defines two terms related to their size: the Botnet's *footprint* and its *live population* [28]. A Botnet's *footprint* is the total number of infected machines at any point in the Botnet's lifetime. A Botnet's *live population* is the

number of bots which are live or active at the same time in a Command and Control (C&C) channel[1]. Rajab et al. [28] estimated that even though the *footprints* of the Botnets tracked by them went as high as several tens of thousands, their effective size i.e. the *live population* at any given point in their lifetime was usually limited to a few thousands. On the other hand, the number of web users accessing a popular website to get the information related to a special event is substantially greater than the number of distinct bots. This forms the basis of another proposed parameter.

In case of a DDoS attack, as the bots are triggered simultaneously, the victim host experiences a large number of new (not seen previously) source IP addresses at the start of the attack. Due to a relatively smaller set of compromised machines, the attacker uses nearly the same set of sources during the course of the attack. In other words, after a DDoS attack starts, the victim host sees very few new source IP addresses sending packets to it. But in case of FEs, as the information travels through the web community, more and more new users come online to access it. Hence, the web server constantly observes new source IP addresses. Therefore, the rate of new source IP addresses as seen by the target can be used as one of the parametric differences between DDoS and FE.

C. Distribution of Requests Among Source IP Addresses

In case of a DDoS attack, as the attacker has control over a limited number of compromised machines it tries to achieve the desired maximum load by forcing each bot to send large number of requests. Hence, the sudden surge in incoming traffic to the victim is mainly attributed to a high number of requests per source IP address.

In contrast to this, during a FE, the majority of the clients are interested only in some specific information, hence a relatively small number of requests originate from each accessing client. Therefore, the excess load that the server experiences in a FE is mainly due to clients with a considerably smaller request per client rate as compared to a DDoS attack.

Summary

These three parameters viz. Rate of Incoming Traffic, Rate of New Source IP Addresses, and Distribution of Requests Among Source IP Addresses indicate the possibility of distinguishing a real-world DDoS attack from a FE. The following section presents the experiments conducted on the two publicly available datasets. It also provides an in-depth evaluation of our results.

V. Experimental Results and Evaluation

The parameters proposed for distinguishing between DDoS and FE in the above section are investigated and verified by the analysis of two datasets viz. CAIDA "DDoS Attack 2007" and "1998 FIFA World Cup". In addition to the statistics provided in Table I, the overall traffic profile for the two datasets is shown in Fig. 1 and Fig. 2.

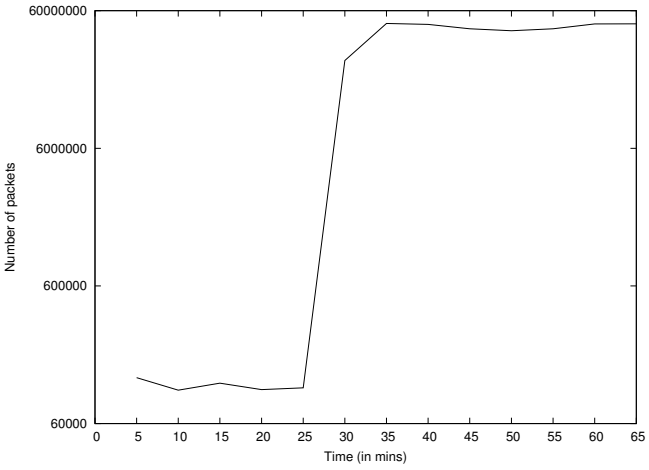

Figure 1. CAIDA traffic profile.

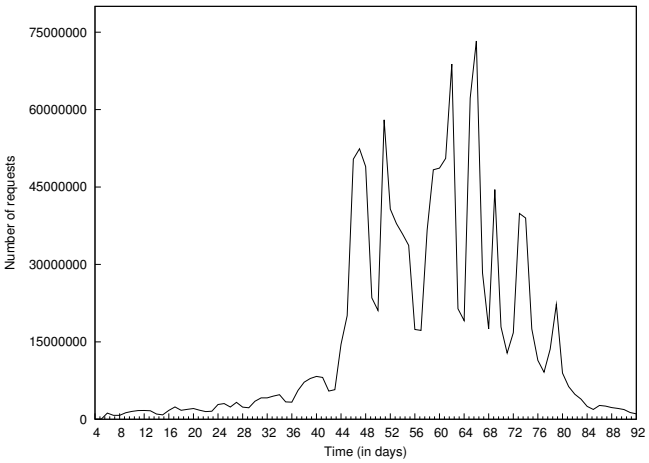

Figure 2. World Cup traffic profile.

Fig. 1 shows the CAIDA DDoS attack where after the initial 25 minutes, the packet rate abruptly jumps and attains the maximum. This maximum is held for the remainder of the attack period.

Fig. 2 shows the access request pattern during the World Cup, including a few days before and after the completion of the event. The presence of various peaks in the traffic profile indicate the occurrence of separate FEs during the course of this 92 day period.

A. Change in Rate of Incoming Traffic

This section discusses the rate at which the incoming traffic appears at the server or victim host in the two cases i.e. FE and DDoS attack. The number of requests or packets in both cases is considerably greater than in the normal situation i.e. non FE and non DDoS duration. Our focus is on analyzing the manner in which these incoming requests or packets increase in both

[1]The botmaster takes control over the bots and issues commands often using an Internet Relay Chat (IRC) server or a particular channel on a public IRC server. This server is also known as C&C server and the channel is known as C&C channel.

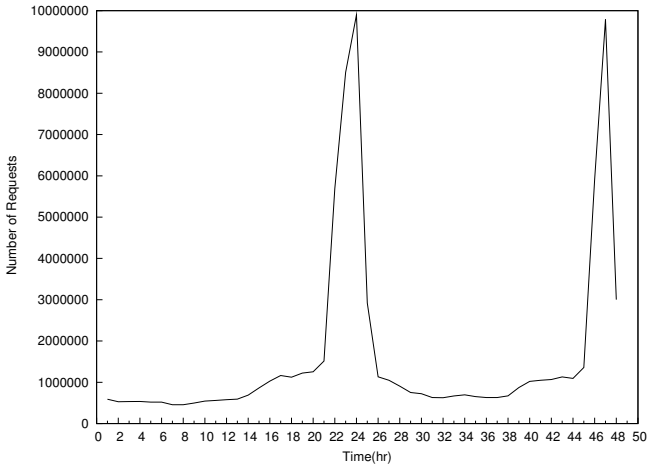

Figure 3. World Cup traffic profile during semi-final matches.

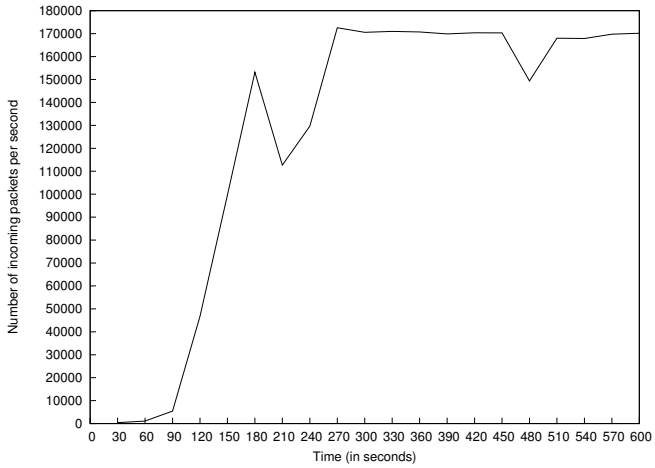

Figure 4. DDoS attack profile (10 minute period).

cases. In order to achieve this we compare the increasing part of the peaks for the two events. For DDoS this represented the ten minute period from 25 - 35 mins (Fig. 1). And in case of FE the selected interval was a two hour interval ($46^{th} and$ 47^{th} hours) during the semi-final matches of the World Cup i.e. $73^{rd} and$ 74^{th} day. Fig. 3 shows the traffic pattern during these two days.

In a DDoS attack, the compromised machines or bots are infected using malware which is pre-programmed to make these machines send packets at a fixed rate. Upon receiving the commands from the bot-master, all of these compromised machines simultaneously start sending large number of packets to the intended target. This causes the incoming traffic at the target machine to abruptly increase in a very short time interval and hence causes the rate of incoming traffic to increase drastically. Fig. 4 shows this trend in the DDoS attack. In Fig. 4 the sudden decrease in the number of incoming packets at time 210 seconds and 480 seconds can be due to the measurement frequency (30 seconds) used in this analysis. The nature of the data and lack of additional information makes it difficult to identify the exact cause of such behavior.

In contrast to the DDoS attack, in case of a FE, it is very unlikely that the entire web user community starts accessing the web server to get the information at precisely the same time. Consequently, the incoming packet rate to the server host will be more steady. Fig. 5 shows the rate of incoming traffic in case of FE.

The clear difference in the slopes for two traffic profiles suggests that this can be used as a distinguishable parameter to separate a real-world DDoS attack from a FE.

B. Change in Rate of New Source IP Addresses

This section presents the analysis of the rate of increase of new source IP addresses as a characteristic feature which can differentiate between a DDoS and FE. As discussed earlier, in cases of DDoS attacks, the attacker has control over a finite set of compromised machines. Activation of the bots produces a gigantic traffic load on the target host coming from a set

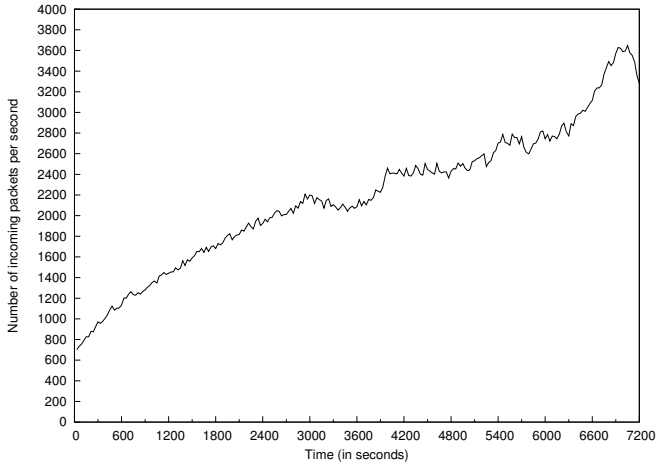

Figure 5. FE profile (120 minute period).

of sources which have not been previously seen. Therefore, it produces a sharp increase in the new source IP addresses synchronized with the onset of the DDoS attack. This limited set of compromised machines are iteratively used during the attack to produce the desired effect. In other words, this means that once the attack starts, the target host rarely sees any further increment in the new source IP addresses. Rather it experiences more packets from the same set of IPs.

Fig. 6 and Fig. 7 supports this conjecture with the new IP addresses suddenly increasing with the onset of the DDoS attack and remaining fairly low and constant for the rest of the attack period. For analyzing the rate of new source IP addresses we performed experiments using different window sizes as the history period including a five minute and one minute window size. For each window size we have calculated the number of new source IP addresses and compared it with the history period i.e. previous five minute and one minute interval respectively.

On the contrary, during a FE the web server sees a relatively constant or slowly increasing rate of new IP addresses,

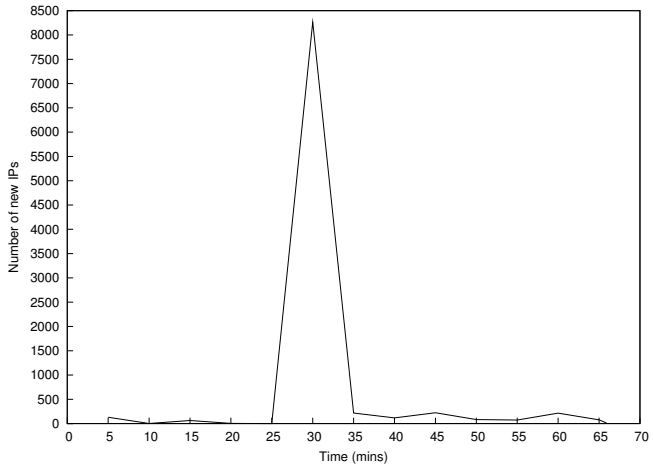

Figure 6. New source IP addresses with five minute history period (DDoS).

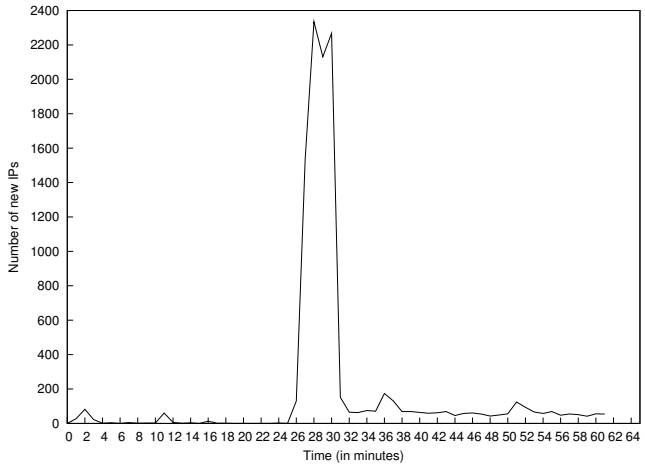

Figure 7. New source IP addresses with one minute history period (DDoS).

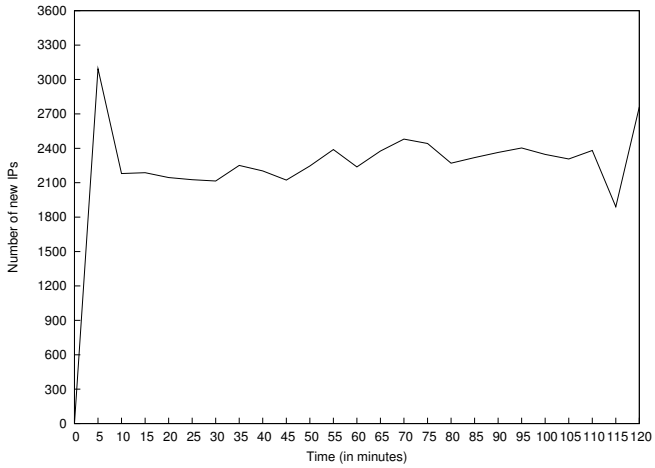

Figure 8. New source IP addresses with five minute history period (FE).

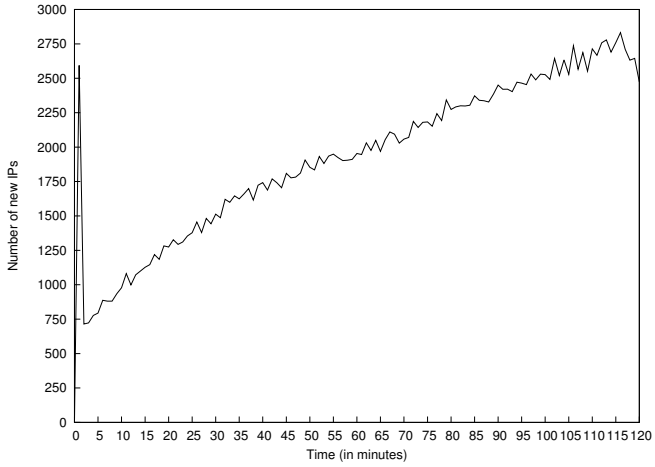

Figure 9. New source IP addresses with one minute history period (FE).

considerably different from the DDoS attack. Fig. 8 shows that apart from the initial five minutes, the number of new sources accessing the server remains fairly constant. For a one minute history period (Fig. 9), the number of new IP addresses increases gradually but it is still different from the DDoS attack where it is more abrupt. Hence, this feature can be used to distinguish between these two class of events.

Rate of arrival of new source IP addresses using the change-point analysis has been used in our previous work to detect the onset of a DDoS attack [16]. There was no attempt in that work to distinguish DDoS attacks from FEs. Using a proof of concept implementation, our research demonstrated how a simple network traffic parameter like new source IP address is sufficient to effectively detect flooding attacks.

C. Distribution of Requests Among Source IP Addresses

For the observation period of these two events i.e. ten minutes for DDoS and two hours for FE, the statistical analysis showed a significant difference in the values of parameters like number of unique IPs, average number of requests per IP etc. Table II provides a micro-level statistics for the two events during the observation period.

In case of a FE, apart from some regular or frequent users, most are concerned with a very limited and specific piece of information related to the event. Thus, a majority of clients appeared to have a very low request per IP values. The overloading of servers in FE is mainly caused by an increase in

Table II
MICRO-LEVEL STATISTICS OF CAIDA AND WORLD CUP DATASETS

Parameters	**World Cup**	**CAIDA**
Observation time	2 hours	10 mins
Number of Requests/Packets	15,698,000	74,478,486
Number of unique source IPs	41,566	8,585
Average number of requests per IP (Mean)	377.66	8675.42
Standard deviation	958.99	7737.87
Coefficient of variation (CV)	2.53	0.89

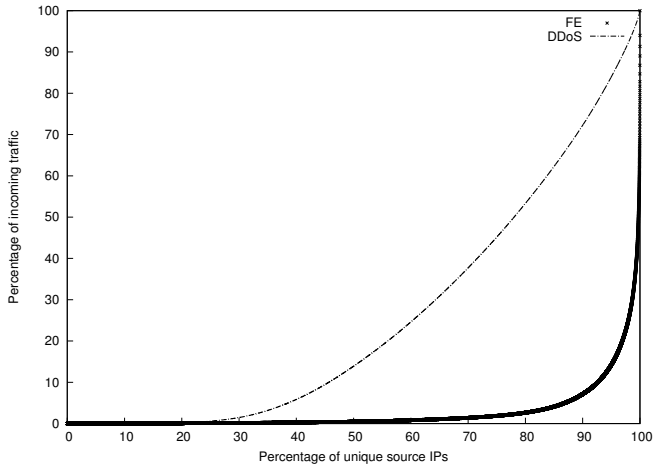

Figure 10. Requests per source IP distribution.

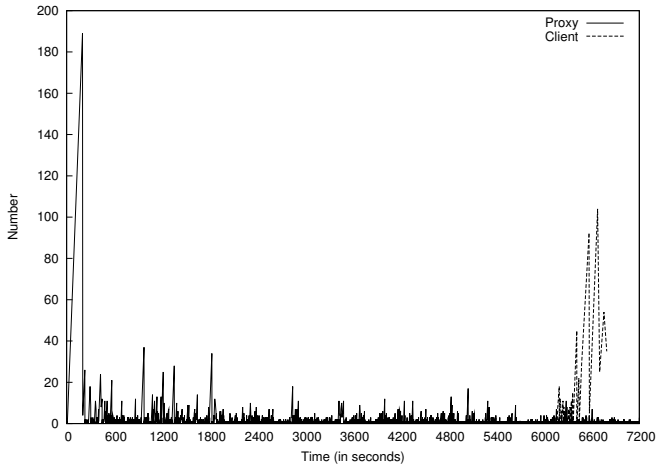

Figure 11. Packet IAT comparison for proxy and client.

the number of clients rather than requests per client. Hence, in case of a FE there are more clients generating very few requests as compared to a DDoS attack where the requests are more uniformly spread across the clients. This behavior is seen to be consistent with the values of coefficient of variation (CV) in both the cases. The value of CV greater than one (as in case of FE) indicates an Erlang Distribution whereas a value less than one (as in case of DDoS) implies an Hyper-exponential Distribution as shows in Fig. 10. It compares the distribution of requests among the source IP addresses in terms of the percentage of total outgoing traffic contributed by each source IP.

Fig. 10 shows that in case of the FE 90% of the total clients contribute less than 10% of the total outgoing traffic which is distinctly different from DDoS attack where the distribution of traffic is more uniform amongst the source IP addresses. Thus, the distribution of requests per source IP distinctly separates the two class of events.

An important observation made from this particular analysis is that the majority of traffic in case of a FE is generated by a small percentage of the overall participating clients, less than 10% in this case. The source IPs sending large number of requests are suspected to be a cluster of network proxies hiding a large number of private source IPs behind a comparatively small set of public IPs as seen by the target host. Another possible explanation could be the presence of web-crawlers or web-spiders which often send very large number of requests within a short time interval.

One of the differences between a proxy and a client as pointed out by Krishnamurthy et.al [29] is the *think time*. A proxy has a lower *think time* than a client and hence it issues more number of requests per time as compared to a client. In other words, the inter-arrival-time (IAT) for packets coming from proxy is less and fairly constant as compared to packets coming from clients which have a large and variable *think time*. This behavioral difference is shown by Fig. 11 which compares this packet inter-arrival-time for a proxy i.e. source IP sending large number of requests and a client i.e. source IP sending

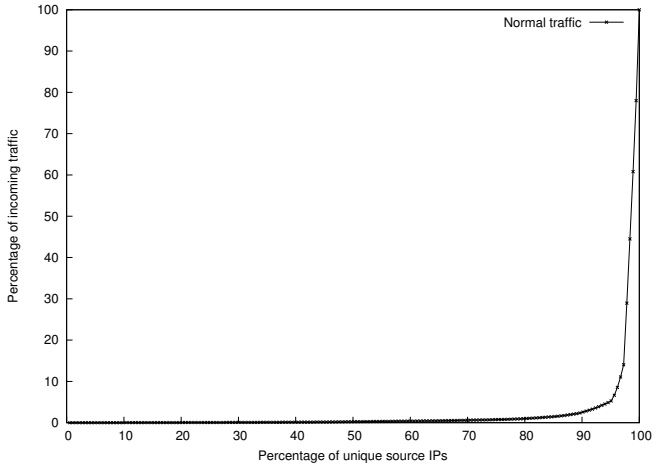

Figure 12. Requests per source IP distribution for normal traffic.

small number of requests. We conjecture that the evolution of botnet behavior to mimic the variable IATs characteristic of FEs is just around the corner and presents a significant danger to successful use of IATs to differentiate botnet traffic from FEs. We expect in future work to explore both this imminent threat and possible counter-measures based around identifying IAT distributions per IP. This is of course complicated further by the point noted above regarding a small number of set of public IPs as seen by the target host hiding a large number of private source IPs.

This led us to analyze the non-attack (first 25 minutes) traffic of the CAIDA dataset. Fig. 12 shows the distribution of outgoing traffic among the clients in case of normal traffic which looks very similar to the one observed in FE. However, the total number of unique sources responsible for the normal traffic (183) is very small as compared to the ones responsible for FEs (8585).

We see how the distribution of outgoing requests among the sources is noticeably different in DDoS attacks and FEs.

VI. Conclusion and Future Work

DDoS attacks have been one of the most destructive modes of attack in the last decade. FEs share significant characteristics with DDoS attacks and this aggravates the existing problem of how to identify DDoS attacks. The research presented in this paper attempts to provide a solution by exposing some subtle differences between these two network anomalies. This paper proposes a set of parameters (Change in Rate of Incoming Traffic, Change in Rate of New Source IP Addresses, and Distribution of Requests Among Source IP Addresses) which can be used to efficiently differentiate between DDoS attacks and FEs. It also provides a comprehensive analysis of two real-world publicly available datasets viz. CAIDA "DDoS Attack 2007" and "1998 FIFA World Cup", representing a DDoS attack and a FE respectively to validate our approach. It is acknowledged that there are marked dissimilarities between these two datasets. One of the future works of this research is to evaluate the proposed parameters on different dataset examples, datasets that are not so markedly dissimilar. This work would also attempt to classify different types of FEs.

The future work will also attempt to identify additional parameters like IAT distribution per IP, geographical location of source IP addresses etc. and use them along-with the existing parameters to build a *distinguishing function* to tell apart these two network traffic anomalies viz. DDoS attack and FE. We also expect to explore and provide a possible counter-measure to the evolution of DDoS attacks imitating the characteristic behavior of FEs e.g. a DDoS attack with a variable or random packet-rate. In addition we seek to investigate how can this parametrization of datasets be used to synthetically generate the network traffic closely approximating a DDoS attack and FE.

Acknowledgment

This research work was supported by the Australian-Indian Strategic Research Fund 2008-2011. Support for CAIDA's Internet Traces is provided by the National Science Foundation, the US Department of Homeland Security, and CAIDA Members.

References

[1] L. Garber, "Denial-of-service attacks rip the Internet," *Computer*, vol. 33, no. 4, pp. 12–17, 2000.

[2] J. Nazario, "Political DDoS: Estonia and Beyond ," in *Invited Talk, in 17th USENIX Security Symposium, July 28-Aug 1, 2008, San Jose, CA, USA*, 2008.

[3] R. Chang, "Defending against flooding-based distributed denial-of-service attacks: A tutorial," *Communications Magazine, IEEE*, vol. 40, no. 10, pp. 42–51, 2002.

[4] L. Niven, "Flash crowd," *The Flight of the Horse. Ballantine Books*, 1971.

[5] J. Borland, "Net video not yet ready for prime time," *CNET News. com. February*, vol. 5, 1999.

[6] J. Jung, B. Krishnamurthy, and M. Rabinovich, "Flash crowds and denial of service attacks: Characterization and implications for CDNs and web sites," in *Proceedings of the 11th international conference on World Wide Web*, pp. 293–304, ACM, 2002.

[7] P. Hick, E. Aben, K. Claffy, and J. Polterock, "The caida "ddos attack 2007" dataset."

[8] M. Arlitt and T. Jin, ""1998 world cup web site access logs"," August 1998.

[9] R. Mahajan, S. Bellovin, S. Floyd, J. Ioannidis, V. Paxson, and S. Shenker, "Controlling high bandwidth aggregates in the network," *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 3, p. 73, 2002.

[10] A. Cearns, *Design of an Autonomous Anti-DDoS network (A2D2)*. PhD thesis, Citeseer, 2002.

[11] L. Limwiwatkul and A. Rungsawangr, "Distributed denial of service detection using TCP/IP header and traffic measurement analysis,"

[12] Yifu Feng, Rui Guo, Dongqi Wang, and Z. Bencheng, "Research on the Active DDoS Filtering Algorithm Based on IP Flow," in *Proceedings of the 2009 Fifth International Conference on Natural Computation*, 2009.

[13] K. Kumar, R. Joshi, and K. Singh, "A distributed approach using entropy to detect DDoS attacks in ISP domain," in *Signal Processing, Communications and Networking, 2007. ICSCN'07. International Conference on*, pp. 331–337, IEEE, 2007.

[14] L. Feinstein, D. Schnackenberg, R. Balupari, and D. Kindred, "Statistical approaches to DDoS attack detection and response," in *DARPA Information Survivability Conference and Exposition, 2003. Proceedings*, vol. 1, pp. 303–314, IEEE, 2003.

[15] T. Peng, C. Leckie, and K. Ramamohanarao, "Protection from distributed denial of service attacks using history-based IP filtering," in *IEEE International Conference on Communications, 2003. ICC'03*, pp. 482–486, 2003.

[16] E. Ahmed, G. Mohay, A. Tickle, and S. Bhatia, "Use of ip addresses for high rate flooding attack detection," in *Proceedings of 25th International Information Security Conference (SEC 2010) : Security & Privacy : Silver Linings in the Cloud, Brisbane, Australia*, 2010.

[17] J. Cabrera, L. Lewis, X. Qin, W. Lee, and R. Mehra, "Proactive intrusion detection and distributed denial of service attacks: A case study in security management," *Journal of Network and Systems Management*, vol. 10, no. 2, pp. 225–254, 2002.

[18] C. Bao, "Intrusion Detection Based on One-class SVM and SNMP MIB Data," in *2009 Fifth International Conference on Information Assurance and Security*, pp. 346–349, IEEE, 2009.

[19] J. Yu, H. Lee, M. Kim, and D. Park, "Traffic flooding attack detection with SNMP MIB using SVM," *Computer Communications*, vol. 31, no. 17, pp. 4212–4219, 2008.

[20] L. Von Ahn, M. Blum, and J. Langford, "Telling humans and computers apart automatically," *Communications of the ACM*, vol. 47, no. 2, pp. 56–60, 2004.

[21] S. Kandula, D. Katabi, M. Jacob, and A. Berger, "Botz-4-sale: Surviving organized DDoS attacks that mimic flash crowds," in *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2*, pp. 287–300, USENIX Association, 2005.

[22] G. Mori and J. Malik, "Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA," 2003.

[23] H. Park, P. Li, D. Gao, H. Lee, and R. Deng, "Distinguishing between FE and DDoS Using Randomness Check," *Information Security*, pp. 131–145, 2008.

[24] Y. Xie and S. Yu, "Detecting Shrew HTTP Flood Attacks for Flash Crowds," *Computational Science–ICCS 2007*, pp. 640–647, 2007.

[25] A. Kuzmanovic and E. Knightly, "Low-rate TCP-targeted denial of service attacks: the shrew vs. the mice and elephants," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, p. 86, ACM, 2003.

[26] Q. Le, M. Zhanikeev, and Y. Tanaka, "Methods of Distinguishing Flash Crowds from Spoofed DoS Attacks," in *Next Generation Internet Networks, 3rd EuroNGI Conference on*, pp. 167–173, IEEE, 2007.

[27] S. Yu, T. Thapngam, J. Liu, S. Wei, and W. Zhou, "Discriminating DDoS Flows from Flash Crowds Using Information Distance," in *Network and System Security, 2009. NSS'09. Third International Conference on*, pp. 351–356, IEEE, 2009.

[28] M. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "My botnet is bigger than yours (maybe, better than yours): Why size estimates remain challenging," in *Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets*, p. 5, USENIX Association, 2007.

[29] B. Krishnamurthy and J. Wang, "On network-aware clustering of web clients," in *Proceedings of the conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pp. 97–110, ACM, 2000.