Modern Machine Learning for Cyber-Defense and Distributed Denial-of-Service Attacks

—RANDY C. PAFFENROTH
Worcester Polytechnic Institute, Worcester, MA 01609 USA

—CHONG ZHOU
Worcester Polytechnic Institute, Worcester, MA 01609 USA

(Corresponding author: Randy C. Paffenroth.)

IEEE DOI 10.1109/EMR.2019.2950183

Abstract—*In computer networks, denial-of-service (DoS) attacks attempt to make computers or network resources unavailable for their intended use. DoS attacks are difficult to detect and mitigate since they normally do not attempt to access the private data of their intended victim, but rather intend to disrupt the publicly available resources their victims provide. This article discusses methods for detecting and mitigating DoS attacks with a focus on techniques that leverage machine learning algorithms. Such algorithms promise to: (a) detect when computer services are being used in an adversarial fashion, (b) separate network traffic into nominal and anomalous components, and (c) provide opportunities for mitigating the attacks while maintaining the integrity of the effected services. The key ingredient of the ideas presented here is the use of correlations and dependencies in computer access patterns, and the larger context in which they exist, to separate the "wheat" – the real users of the services – from the "chaff" –the perpetrators who are attempting to disrupt the services.*

Key words: Anomaly detection, cyber-defense, deep learning, distributed denial of service (DDoS), machine learning

INTRODUCTION

IN today's business environment, the reliability and availability of computer networking services are crucial to many business enterprises. Denial-of-service (DoS) attacks have recently received much attention in the news media and the literature based upon their ability to disrupt the computer services we use every day.

DoS attacks do not aim to steal data or to access computers by breaching their safeguards. In fact, the public services which DoS attacks are intended to disrupt are being used, in large part, as they were designed to be used. However, the volume and timing of the usage pattern of the attacker is intended to have a deleterious impact on the abilities of others to use the resource, and DoS attacks are characterized by an explicit attempt by attackers to prevent legitimate use of a service.

As a simple example, consider an Internet shopping site where a web page or application programming interface (API) is used for searching for products and making purchases. A normal usage pattern would involve a given customer searching for products, placing some products into their virtual shopping cart, entering delivery information, and then proceeding to paying for the product using a credit card. Perhaps surprisingly, it is quite easy for an attacker to take advantage of such a system. The attacker merely performs each of the steps of the standard usage pattern, except at the very end of the process they never complete the final action of paying for the items. In this case, the service provider has dedicated resources including bandwidth, computing power, storage, and database accesses, to support the attacker's actions. Most importantly, these dedicated resources are not available to support legitimate customers.

A single instance of such misbehavior does not present a problem to the online store, since the online merchant likely has access to a large pool of the appropriate resources. However, what if the attacker performs the same action thousands or millions of times? One can easily imagine the available resources being exhausted or, at the very least, the performance of the web page for the actual customers being seriously degraded. These results will impact the performance of the business.

It is important to note that the public services are being used as intended, and the only adversarial aspect of the interaction is actually the *lack* of an action, namely the final purchasing step.

As we will detail later, the nature of such DoS attacks makes them difficult to detect and mitigate. When such DoS attacks are distributed across a large swath of the Internet, and involve hundreds or thousands of attackers, their mitigation often requires advanced algorithmic techniques. These algorithms can study the patterns of user behavior. They separate the legitimate users of a network service from those who are merely trying to prevent others from accessing the service. We expect that computer network services providers and their stakeholders will appreciate the following insights from this article where we:

1. promote the wider appreciation of the underlying structure and intention of DoS attacks,
2. raise awareness of the difficulties in detection and mitigating DoS attacks in modern computer networks, and
3. discuss methods for detection when anomalous usage patterns arise and separating legitimate from adversarial users.

BACKGOUND

Since DoS attacks use computer services, in large part, as their designers intended, one might rightly wonder how the attacked can achieve a deleterious effect. For example, if a particular user is found to be misusing the service, then that user's computer can simply be blocked from using the service. An Internet shopping site could merely disable a user's account who abuses the service. Unfortunately, real world DoS attacks are far more sophisticated, and are almost always *distributed* in nature.

Many attacks on personal computers, such as the viruses found in downloaded executables [Ludwig, 2017] or "phishing" attacks which use forged emails to get users to visit malicious web pages [Chiew, 2018], care surprisingly little about the machines they infect or their contents. Many attackers merely care that they can control the operation of the infected machines when the need arises to attack their true target.

Even if 99% of attacks fail, the few successes can lead to an attacker having many thousands of machines at their disposal in what is sometimes called a "bot-net" [Behal, 2017]. Bot-nets – see Figure 1 – can be used as tools by attackers and provide many anonymous computers from which large-scale attacks can be staged. It is perhaps interesting, and troublesome, to note that bot-nets are traded and sold on the Internet, providing a marketplace for DoS attackers to purchase the tools of their trade [Putman, 2018].

The difficultly of mitigating DoS attacks now becomes clear. A million aborted purchases from a single customer would raise any number of red flags and be easily mitigated by merely blocking that customer from using the services. The implementation of such a strategy is known as a "blacklist." When such DoS attacks are distributed across many attacking computers, the defender is forced to be subtler in their response.

There may be millions of aborted purchases, but each purchase appears to be from a unique customer. The defender is therefore forced to dedicate at least some resources to each visitor to the web site, since they cannot immediately discern the legitimate customers from the attackers. Such attacks are called distributed denial-of-service (DDoS) attacks, and they will comprise the focus of our discussion.

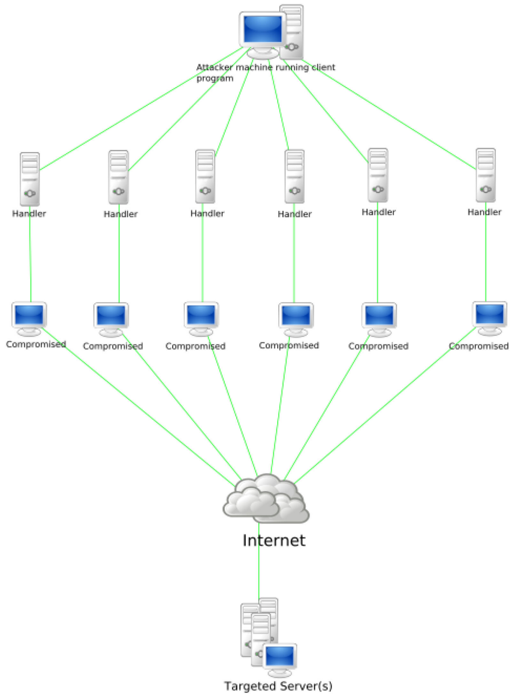

Figure 1. An example of a bot-net used for a distributed DoS attack. The attacking machine, at the top, has gained controls of a large number of "bots" as shown in the middle. The target is then overwhelmed by traffic generated by the bots. Figure curtousey of Coelho Everaldo https://commons.wikimedia.org/w/index.php?curid=3980651.

In many cases, the assailant's computer that is orchestrating the attack never actually contacts the target service. It is only through the intermediary machines in the bot-net that the attackers influence their target.

What is the defender to do? One extreme response would be to turn off the public services that are being attacked. While this response might blunt the attack, it would also achieve the precise goal of the perpetrator: the disruption of the service. Any effective response must somehow distinguish the nominal users of the service from the anomalous users of the service, especially when the owners of the anomalous computers are not even aware that their machines are being used in a nefarious fashion.

A tempting solution is to perform some type of pattern detection. An unsophisticated attacker may provide indications that they are adversarial by always preforming the same searches or purchasing the same items. A defender would be wise to not assume all attackers are so unsophisticated. More sophisticated pattern detection approaches are often required.

There are two general forms of DDoS attacks: those that crash services and those that flood services [Zargar, *et al*. 2013]. In the latter category, two illustrative examples are "SYN-flood" [Kumar, 2018] attacks and "slow read" [Hong, 2017] attacks.

Using online purchases as an example, a SYN-flood attack is analogous to large number of users trying to connect to a server in a short period of time. A SYN flood occurs when a host sends a flood of transmission control protocol (TCP) "SYN" packets, which are used by a client to request a connection to the server.

SYN packets are not only common in nominal Internet traffic but are a required part of all reliable Internet communications. Each of these packets are handled like a connection request, causing the server to spawn a half-open connection and waiting for a packet in response from the sender address. SYN-flood attacks can affect many different types of Internet resources, since they affect the underlying protocols that all computers use.

Using our online marketplace example, the slow read attack can be thought of as a group of attackers holding onto a limited set of items that prevents them being sold to legitimate users. A slow read attack sends legitimate application requests, but reads responses provided by the service very slowly, thus trying to exhaust the server's connection pool. Such attacks are often called "low and slow," since they do not overwhelm the target with a large volume of traffic all at one time, but instead slowly gather resources that they never release back to the server.

An arms race is afoot. Attackers attempt to mask their attacks as normal traffic, while the defender attempts to find commonalities in the attacks and features that differentiate them from normal users so that they can make best use of their limited resources. Automated procedures are clearly of importance given the rate and volume of attacks, and this is where machine learning approaches show their strength.

Pattern Detection for DDOS Attacks

There is a large amount of research work revolving around the application of machine learning and data mining techniques on analyzing cyber-security data [Dilek, 2015]. Most of them fall into two main categories, namely signature-based and anomaly-based [Buczak, 2016].

Here, we mention two examples of signature-based attacks in the literature. First, Hu *et al*. [2003] used a method called a robust support vector machine (SVM), a variation of the classic SVM [Drucker, 1997] which they use as an anomaly classifier in their study. Second, an application of Random Forests to anomaly detection is described by Zhang *et al*. [2008], where an anomaly (outlier) detector was employed to feed a second threat classifier.

Numerous anomaly-based DDoS detection techniques have been developed to detect such new attack vectors, and many such methods model the normal network and system behavior and identify anomalies as deviations from this normal behavior. One class of anomaly-based DDoS that bears particular note, because of their popularity and their effectiveness, are those that use deep learning and neural networks ([Niyaz, 2016]; [Yuan, 2017]; [Tang, 2016]).

Deep Learning

Here we closely follow the ideas and presentation by Zhou [2019], and we note that deep learning is part of a broad family of methods for representation learning [LeCun, *et al*. 2015]. Such methods are widely used for many tasks such as natural language processing and image understanding. Unfortunately, DDoS attacks can be quite complicated, and outliers and noise, which are common in DDoS data, may reduce the quality of representations discovered by deep learning algorithms such as autoencoders ([Lyudchik, 2016]; [Ma, *et al*. 2013]). Perhaps most importantly, classic denoising autoencoders require access to anomaly-free data which may not be available for many DDoS anomaly detection problems, and therefore many deep learning methods that are available in other problem domains are not appropriate for detecting DDoS attacks.

However, there has been recent work in robust models which address many of these issues and have been successfully applied to DDoS detection problems. For example, robust principal component analysis (RPCA) has been used to construct low-dimensional linear representations on DDoS data by filtering out outlying measurements [Paffenroth, et al., 2013]. Robust PCA is a generalization of principal component analysis that attempts to reduce the sensitivity of PCA to anomalies [Paffenroth, et al., 2013]. Robust PCA allows for the careful teasing apart of sparse anomalies so that the remaining low-dimensional approximation of the measured network data is faithful nominal functioning of the network. Finally, a class of techniques that provides the advanced capabilities of deep learning with the robustness of RPCA have been recently developed leading to a class of techniques called robust deep autoencoders (RDAs) [Zhou & Paffenroth, 2017]. Such methods have been used on both real-world data as well as high-fidelity simulators and have demonstrated effectiveness in separating nominal network traffic from anomalous network traffic arising from a DDoS attack, even when the particular DDoS attack in question has never been observed before. Figure 2 shows an example of the type of results that can be achieved using advanced algorithms such as RDAs for real-world DDoS attacks.

Managerial and Organizational Observations and Lessons

In summary, we provide observations for organizations, managers, and technologists when seeking to address these issues, especially when using advanced algorithms such as deep learning.

1. Choosing the correct tool: There are a myriad of algorithms available for machine learning and anomaly detection. However, not all algorithms are suited to DDoS problems. It is important to choose algorithms that can ingest large amounts of data, but where perhaps only a small fraction of the data is labeled as to whether it is anomalous or not.
2. Deep learning: Deep learning is an active area of research where new approaches are being developed on an almost daily basis. Having a measure of in-house expertise in such methods is important to be able to discern which algorithms and approaches are applicable in your particular context. In particular, the expertise to develop new algorithms in this domain would not be required, but the ability to faithfully assess the performance of algorithms is of prime importance.
3. Data gathering: Advanced machine learning and deep learning algorithms require large amounts of data for training. So, it is never too early to start gathering data on the performance of your particular system. Having well-defined formats and systems for data gathering is important. Starting to gather data as early as possible, even when the details of the approaches, to eventually be used are not known, can pay dividends as DDoS detection and mitigation techniques are being evaluated.
4. Principled testing of approaches: The evaluation of machine learning algorithms is surprisingly difficult, and issues such as “data snooping” can lead one to believe a system is better than it will actually be in practice. Accordingly, testing the system in a manner which is a close as possible to the actual fielded

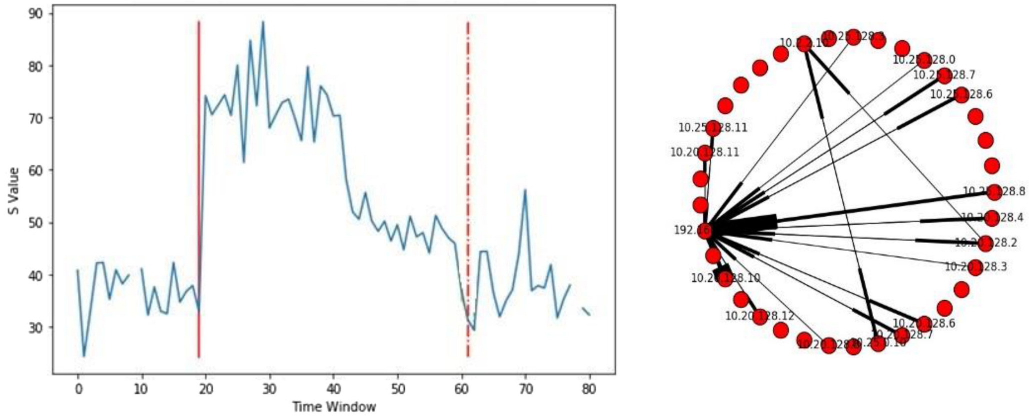

Figure 2. An example of an attack detected using an RDA. On the left we plot a time series showing a measure, called the “S Value,” of how anomalous the network traffic is over time. We overlay that plot with vertical lines that indicate when an attack begins (solid red line) and ends (dashed red line). Such methods also call for the DDoS attack to be attributed to particular network hosts. On the right we have a number of network hosts all shown at red dots, and many of these hosts are all accessing the same target host.

configuration is important. It is even more important that the data on which the system is tested is as realistic as possible.

Conclusion

Denial-of-service attacks can make computers or network resources unavailable for their intended use and they are hard to detect and mitigate since they, in large measure, use the affected services as intended.

DoS attacks do not attempt to access the private data of their intended victim, but instead disrupt publicly available resources, such as Internet commerce sites, by overwhelming them with service requests. Since, by their very nature, such attacks can be difficult to distinguish from nominal traffic, this article discussed several machine learning methods for detecting and mitigating DoS attacks.

Of particular importance for DDoS mitigation are modern methods that leverage deep learning and neural networks. Such methods have been used successfully in many problem domains, and herein we provided several examples of their use in DDoS mitigation.

Acknowledgment

The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. Government, DARPA, or Raytheon BBN Technologies. The authors would like to thank project collaborators P. Pal, N. Soule, and C. Doucette.

References

Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press.

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3, 1–122.

Buczak, A. L. (Apr.–Jun. 2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18 (2), 1153–1176.

Chiew, K. L. (2018). A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems With Applications*, 106, 1–20.

Dilek, S. A. (2015). Applications of artificial intelligence techniques to combating cyber crimes: A review. arXiv:1502.03552.

Drucker, H. A. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, (pp. 155–161).

Gehring, J., Miao, Y., Metze, F., & Waibel, A. (2013). Extracting deep bottleneck features using stacked auto-encoders. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (pp. 3377–3381).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA, USA: MIT press.

Hong, K. A. (Apr. 2018). SDN-assisted slow HTTP DDoS attack defense method. *IEEE Communications Letters*, 22 (4), 688–691.

Hu, W., Liao, Y., & Vemuri, V. R. (2003). Robust support vector machines for anomaly detection in computer security. *International Conference on Machine Learning and Applications*, (pp. 168–174).

Kumar, P. (Dec. 2018). SAFETY: Early detection and mitigation of TCP SYN flood utilizing entropy in SDN. *IEEE Transactions on Network and Service Management*, 15 (4), 1545–1669.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.

Ludwig, M. A. (2017). *The Giant Black Book of Computer Viruses*. American Eagle Books.

Lyudchik, O. (2016). Outlier detection using autoencoders. Tech. Rep.

Ma, Y., Zhang, P., Cao, Y., & Guo, L. (2013). Parallel auto-encoder for efficient outlier detection. *IEEE International Conference on Big Data*, (pp. 15–17).

Niyaz, Q. A. (2016). A deep learning based DDoS detection system in software-defined networking (SDN). *arXiv:1611.07400*.

Paffenroth, R., Du Toit, P., Nong, R., Scharf, L., Jayasumana, A. P., & Bandara, V. (Feb. 2013). Space-time signal processing for distributed pattern detection in sensor networks. *IEEE Journal of Selected Topics in Signal Processing*, 7 (1), 38–49.

Putman, C. A. (2018). Business model of a botnet. *26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, (pp. 441–445).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11, 3371–3408.

Zargar, S. T., Joshi, J., & Tipper, D. (Oct.–Dec. 2013). A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE Communications Surveys & Tutorials*, 15 (4), 2046–2069.

Zhang, J., Zulkernine, M., & Haque, A. (Sep. 2008). Random-forests-based network intrusion detection systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38 (5), 649–659.

Zhou, C. (2019). *Robust Methods for Anomaly Detection With Applications to Cyber Data*. Worcester, MA, USA: Worcester Polytechnic Institute.

Zhou, C., & Paffenroth, R. C. (2017). Anomaly detection with robust deep autoencoders. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 665–674).

Randy C. Paffenroth received the graduate degree in both mathematics and computer science from Boston University, Boston, MA, USA and the Ph.D. degree in applied mathematics from the University of Maryland, College Park, MD, USA, in June 1999. After attaining his Ph.D., he spent seven years as a Staff Scientist in applied and computational mathematics with the California Institute of Technology. In 2006, he joined Numerica Corporation, where he held the position of Computational Scientist and Program Director. He is currently an Associate Professor of mathematical sciences, computer science, and data science with the Worcester Polytechnic Institute, Worcester, MA, USA. His current technical interests include machine learning, signal processing, large scale data analytics, compressed sensing, and the interaction between mathematics, computer science, and software engineering, with a focus on applications in cyber-defense and material science.

Chong Zhou received the B.Eng. degree in computer science from Southwest University, Chongqing, China in 2012 and the Ph.D. degree in data science from the Worcester Polytechnic Institute, Worcester, MA, USA, in 2019. He is currently with Microsoft Inc., Redmond, WA, USA. He focuses on anomaly detection applications and methods.