Fundamental Limits of Volume-based Network DoS Attacks

Fundamental Limits of Volume-based Network DoS Attacks

Xinzhe Fu
LIDS, Massachusetts Institute of Technology
USA

Eytan Modiano
LIDS, Massachusetts Institute of Technology
USA

ABSTRACT

Volume-based network denial-of-service (DoS) attacks refer to a class of cyber attacks where an adversary seeks to block user traffic from service by sending adversarial traffic that reduces the available user capacity. In this paper, we explore the fundamental limits of volume-based network DoS attacks by studying the minimum required rate of adversarial traffic and investigating optimal attack strategies. We start our analysis with single-hop networks where user traffic is routed to servers following the Join-the-Shortest-Queue (JSQ) rule. Given the service rates of servers and arrival rates of user traffic, we first characterize the feasibility region of the attack and show that the attack is feasible if and only if the rate of the adversarial traffic lies in the region. We then design an attack strategy that is (i). *optimal:* it guarantees the success of the attack whenever the adversarial traffic rate lies in the feasibility region and (ii). *oblivious:* it does not rely on knowledge of service rates or user traffic rates. Finally, we extend our results on the feasibility region of the attack and the optimal attack strategy to multi-hop networks that employ Back-pressure (Max-Weight) routing. At a higher level, this paper addresses a class of dual problems of stochastic network stability, i.e., how to optimally de-stabilize a network.

KEYWORDS

Denial-of-Service Attacks; Stochastic Network Scheduling; Network Queueing Theory

ACM Reference Format:
Xinzhe Fu and Eytan Modiano. 2020. Fundamental Limits of Volume-based Network DoS Attacks. In *ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '20 Abstracts), June 8–12, 2020, Boston, MA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3393691.3394190

1 INTRODUCTION

Network denial-of-service (DoS) attacks, where an adversary seeks to make some network resource unavailable to its intended users, is one of the most serious security threats to the Internet. It often results in downtime of web services, cloud computing facilities, DNS services, etc., causing huge financial loss to institutions [1]. While some network DoS attacks exploit the vulnerabilities of protocols, the predominant type of attacks are volume-based, such as TCP SYN Flood, UDP Flood and DNS Flood [2]. They work by flooding the network with adversary traffic and blocking the service to normal users [2]. Such adversary traffic can be generated distributively from botnets and is difficult to distinguish from normal user traffic [4], which makes volume-based DoS attacks difficult to defend against. Due to the significance and prevalence of volume-based network DoS attacks, there have been a flurry of works focusing on their detection and mitigation [3, 5, 6]. However, a theoretical understanding of the limits of such attacks is still lacking, i.e., **how much resources does the adversary need for mounting a successful volume-based network DoS attack and what is the optimal attack strategy?**

In this paper, we explore the fundamental limits of volume-based network DoS attacks. Taking a network flow and queueing perspective, we translate the scenario of network DoS attacks to one where the adversary injects traffic and seeks to de-stabilize the network by overflowing network queues. Such perspective closely mirrors volume-based DoS attacks in real life and enables us to conveniently inherit the modeling and analysis tools from the network flow and queueing literature. We start our analysis with a server farm which can be modeled as a single-hop network and then generalize our results to multihop networks.

SIGMETRICS '20 Abstracts, June 8–12, 2020, Boston, MA, USA

ACM ISBN 978-1-4503-7985-4/20/06.
https://doi.org/10.1145/3393691.3394190

2 MAIN RESULTS

Consider a single-hop network with a set of parallel servers (sinks) and a set of traffic dispatchers (sources). The dispatchers are divided into two disjoint subsets: user traffic dispatchers that route user traffic to servers, and adversary traffic dispatchers, controlled by the adversary, that send adversary traffic to servers to block the user traffic. We use $S = \{s_1, \ldots, s_N\}$ to denote the set of servers, $U = \{u_1, \ldots, u_L\}$ to denote the set of user traffic dispatchers and $V = \{v_1, \ldots, v_M\}$ to denote the set of adversary traffic dispatchers. A generic server, a generic user traffic dispatcher and a generic adversary traffic dispatcher are denoted by s_n or n, u_l or l, v_m or m, respectively. Let $S_{u_l} \subseteq S$ be the set of servers that user dispatcher u_l is connected to, and $S_{v_m} \subseteq S$ be the set of servers that adversary dispatcher v_m is connected to. Each dispatcher can only route packets to the servers to which it is connected.

The network operates in discrete time with time t starting from 0. Each server has a infinite-size queue that buffers the packets, with $Q_n(t)$ representing the length of the queue of server s_n at time t. The offered service of server n at time t is denoted by $b_n(t)$. The servers do not distinguish user and adversary traffic and employ the First-Come-First-Serve (FCFS) service discipline[1]. In each time slot, $\lambda_l^u(t)$ packets arrive at user dispatcher u_l, which routes the packets to the servers following the "Join-the-Shortest-Queue" (JSQ) policy, that is, at each time slot, each user dispatcher u_l routes all its incoming packets to the server s with the minimum queue length among the ones to which it is connected ($s \in \arg\min_{s_n \in S} Q_n(t)$); Similarly, $\lambda_m^v(t)$ packets arrive at adversary dispatcher v_m, which routes the packets to servers according to some adversarial injection policy.

[1]Our results hold under all common service disciplines except priority based service with user traffic having the priority.

We assume that $b_n(t)$'s, $\lambda_l^u(t)$'s and $\lambda_m^v(t)$'s are independent sequences of i.i.d. random variables with $\mathbb{E}[b_n(t)] = \mu_n, \mathbb{E}[\lambda_l^u(t)] = \lambda_l^u, \mathbb{E}[\lambda_m^v(t)] = \lambda_m^v$. We further define $Q_n^u(t)$ and $Q_n^v(t)$ as the number of user packets and adversary packets in Q_n at t, respectively. At each time slot t, we decompose the offered service $b_n(t)$ into that offered to user traffic $b_n^u(t)$ and that offered to adversary traffic $b_n^v(t)$ with $b_n^u(t) + b_n^v(t) = b_n(t)$. Under the FCFS service discipline, the breakdown between $b_n^u(t)$ and $b_n^v(t)$ only depends on the queue composition. We further define $a_n^u(t)$ as the sum of user traffic arrivals to server n and $a_n^v(t)$ as the counterpart of adversary traffic. we also write $a_{mn}^u(t)$ ($a_{ln}^v(t)$) as the amount traffic that user dispatcher u_l (adversary dispatcher v_m) sends to n at time t. Based on the system dynamics, we summarize the queue length evolution as follows:

$$Q_n^u(t+1) = [Q_n^u(t) + a_n^u(t) - b_n^u(t)]^+,$$
$$Q_n^v(t+1) = [Q_n^v(t) + a_n^v(t) - b_n^v(t)]^+,$$
$$Q_n(t+1) = Q_n^v(t+1) + Q_n^u(t+1),$$

The adversary dispatchers inject their packets to servers in an effort to prevent user packets from getting served. A network DoS attack is considered successful if the adversary manages to block a positive fraction of user traffic from service. Formally, the goal of the adversary is that

$$\text{For some } n \in \{1, \ldots, N\}, \qquad \lim_{t\to\infty} \frac{\mathbb{E}[Q_n^u(t)]}{t} > 0, \tag{1}$$

which is equivalent to making user traffic in one of the queues mean rate-unstable [7]. Furthermore, by Little's law, (1) implies that the mean delay experienced by user traffic grow linearly with time. We say that the adversary *destabilizes* user traffic, if it achieves (1). The Network DoS Attack problem we study is **feasible** if there exists an adversary injection policy that destabilizes user traffic.

For each subset of servers $S' \subseteq S$, we define $U_{S'}$ as the user dispatchers that only have connections to servers in S', i.e., $U_{S'} = \{u_l \mid S_{u_l} \subseteq S'\}$. We further define $\Delta(S')$ as

$$\Delta(S') = \sum_{s_n \in S'} \mu_n - \sum_{u_l \in U_{S'}} \lambda_l^u.$$

$\Delta(S')$ can be interpreted as the excess service rate of S' with respect to the user traffic generated by $U_{S'}$. Finally, for each $S' \subseteq S$, we define the following linear program $LP(S')$ whose optimal value is denoted as $val(S')$.

$$val(S') = \max \sum_{m\in V} \sum_{n\in S'} f_{mn} \tag{2}$$
$$\textbf{s.t.} \sum_{n\in S'} f_{mn} \le \lambda_m^v, \qquad \forall m \in V \tag{3}$$
$$\sum_{m\in V} f_{mn} \le \mu_n, \qquad \forall n \in S' \tag{4}$$
$$f_{mn} = 0, \qquad \text{if } n \notin S_{v_m}$$
$$f_{mn} \ge 0, \qquad \forall m \in V, n \in S'.$$

Based on the definitions, we first give a necessary and sufficient condition for the feasibility of the Network DoS Attack problem in Theorem 1.

THEOREM 1. *The network DoS problem is feasible if and only if there exists a subset of servers $S' \subseteq S$ such that $U_{S'}$ is non-emptyand $val(S') > \Delta(S')$.*

Next, we propose the Min-Zero policy which works as follows: at each time slot t, the adversary maintains a target subset of user dispatchers and a corresponding target subset of servers, which are denoted by $U(t)$ and $S(t)$, with $U(t) \subseteq U$, $S(t) \subseteq S$ and $S(t) = \bigcup_{u_l \in U(t)} S_{u_l}$. All the adversary dispatchers that have connections to $S(t)$ send packets to $S(t)$ in a JSQ fashion, and other adversary dispatchers send packets arbitrarily. Then, after the servers finished their service during the current slot, the adversary checks if $\min_{n\in S(t)} Q_n(t) = 0$ (hence the name, Min-Zero). If so, then in the next slot, the adversary choose $U(t+1)$ uniformly at random from all non-empty subsets of user dispatchers and set $S(t+1)$ accordingly; otherwise, set $U(t+1) := U(t)$ and $S(t+1) := S(t)$.

We show in Theorem 2 that the Min-Zero policy does not rely on network statistics (the arrival rates and service rates) and destabilizes user traffic whenever the Network DoS Attack problem is feasible. The proof is done by showing the existence of a Lypunov function with positive drift on the Markov chain of queues [9].

THEOREM 2. *Under the Min-Zero policy, there exists a queue n with $\lim_{t\to\infty} \frac{\mathbb{E}[Q_n^u(t)]}{t} > 0$ if the network DoS attack problem is feasible.*

Finally, we extend our results to multi-hop networks that employs back-pressure routing [8]. We propose the multi-hop counterpart of the feasibility condition and a extended version of the Min-Zero policy that works in the multi-hop setting.

ACKNOWLEDGMENTS

This work was supported by DTRA grants HDTRA1-13-1-0021 and HDTRA1-14-1-0058, and NSF grants AST-1547331, CNS-1617091, CNS-1524317, and CNS-1907905.

REFERENCES

[1] https://www.msspalert.com/cybersecurity-research/kaspersky-lab-study-average-cost-of-enterprise-ddos-attack-totals-2m/
[2] https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/
[3] S. T. Zargar, J. Joshi and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks", in *IEEE communications surveys & tutorials*, Vol. 15, No. 4, pp. 2046-2069, 2013
[4] C. Kolias, G. Kambourakis, A. Stavrou and J. Voas, "DDoS in the IoT: Mirai and other botnets", in *Computer*, Vol. 50, No. 7, pp. 80-84, 2017.
[5] R. Braga, E. de Souza Mota and A. Passito, "Lightweight DDoS flooding attack detection using NOX/OpenFlow", in *IEEE LCN*, Vol. 10 pp. 408-415, 2010.
[6] A. Compagno, M. Conti, P. Gasti and G. Tsudik, "Poseidon: Mitigating interest flooding DDoS attacks in named data networking", in *IEEE LCN*, pp. 630-638, 2013.
[7] M.J. Neely, "Stochastic network optimization with application to communication and queueing systems", in *Synthesis Lectures on Communication Networks*, Vol. 3, No. 1, pp. 1-211, 2010.
[8] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks", in *IEEE Conference on Decision and Control*, pp. 2130-2132, 1990.
[9] G. Fayolle, V. A. Malyshev and M. V. Men'shikov, "Topics in the constructive theory of countable Markov chains," Cambridge university press, 1995.