Source-End DDoS Defense in IoT Environments

Samuel Mergendahl
University of Oregon
smergend@cs.uoregon.edu

Devkishen Sisodia
University of Oregon
dsisodia@cs.uoregon.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Hasan Cam
Army Research Laboratory
hasan.cam.civ@mail.mil

ABSTRACT

While the Internet of Things (IoT) becomes increasingly popular and pervasive in everyday objects, IoT devices often remain unprotected and can be exploited to launch large-scale distributed denial-of-service (DDoS) attacks. One could attempt to employ traditional DDoS defense solutions, but these solutions are hardly suitable in IoT environments since they seldom consider the resource constraints of IoT devices.

This paper presents FR-WARD which defends against DDoS attacks launched from an IoT network. FR-WARD is an adaptation of the classic DDoS defense system D-WARD. While both solutions are situated near the attack sources and drop packets to throttle DDoS traffic, FR-WARD utilizes the fast retransmit mechanism in TCP congestion control to minimize resource penalties on benign IoT devices. Based on our analysis and simulation results, FR-WARD not only effectively throttles DDoS traffic but also minimizes retransmission overhead for benign IoT devices.

1 INTRODUCTION

The majority of Internet of Things (IoT) devices lie in edge networks, so they are subject to similar security threats as any other regular end host. However, due to their nature of low power and low computational strength, IoT devices suffer from many additional vulnerabilities. For example, an IoT device may not be capable of running encryption algorithms or anti-virus software, and because of this, many IoT devices become easy targets of botnets. Due to the sheer size of IoT, these botnets can perform destructive distributed denial-of-service (DDoS) attacks. In fact, in October 2016, an IoT botnet attacked and disabled a US-based DNS provider, Dyn, on which many Fortune 500 companies rely [3]. This botnet comprised an estimated 100,000 malicious IoT devices, and the attack achieved upwards of 1 Tbps of DDoS traffic.

A classic approach against such an attack is to employ a *source-end* defense solution to detect and thwart the attack traffic before it leaves its original network [1, 2, 5]. One classic source-end defense solution, D-WARD [5], monitors both inward and outward traffic statistics of the network it polices in order to compare current network traffic information to predefined normal flow patterns. D-WARD defines the aggregate traffic from the *entire* source network to a particular end host outside the source network as an **agflow** and the aggregate traffic from *one node* in the source network to a particular host outside the source network as a **connection**. Whenever the collected traffic statistics suggest an agflow may be malicious, D-WARD throttles all connections it identifies as even slightly suspicious within this agflow in order to mitigate any DDoS attacks.

However, IoT environments significantly suffer when a security solution throttles benign traffic. In an IoT environment, when D-WARD defines a benign connection as suspicious, it throttles this benign connection which forces an IoT device to retransmit dropped segments. In a network of IoT medical devices or elderly care monitors that ensure the well-being of their users, this loss or severe delay of vital packets is unacceptable, and additionally, any retransmission required of IoT devices further reduces their precious battery life.

In this paper, we describe our system, FR-WARD (utilizing TCP's **F**ast **R**etransmit for Net**w**ork **A**ttack **R**ecognition and **D**efense). It utilizes the fast retransmit mechanism of TCP congestion control to help shape the traffic that leaves a source network, which not only limits the amount of DDoS traffic that leaves the source network but also minimizes resource penalties on benign IoT devices.

2 FR-WARD

FR-WARD has two main goals; first, it must throttle DDoS traffic that leaves the source network it polices, but second, it must avoid resource penalties on benign traffic. In order to achieve these goals, FR-WARD shapes the traffic that leaves its policed network by utilizing the fast retransmit mechanism in TCP congestion control. In TCP, if a sender receives three additional acknowledgements for a particular previous segment in a row, the sender performs congestion control and cuts its window size in half (i.e. multiplicative decrease) [4]. We define this set of three additional acknowledgements of a previously sent segment as a **signal**.

FR-WARD monitors and shapes the traffic in an IoT environment in order to mitigate DDoS attacks leaving its policed network. It labels agflows and connections as malicious, suspicious, or benign similar to D-WARD, and when FR-WARD detects a malicious agflow, FR-WARD throttles all connections it identifies as malicious within this agflow. However, for each suspicious connection within this agflow, instead of throttling it like D-WARD, FR-WARD sends a

This project is in part the result of funding provided by the 2014 I3 award of the University of Oregon and the Science and Technology Directorate of the United States Department of Homeland Security under contract number D15PC00204. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Department of Homeland Security or the US Government.

IoTS&P'17, November 3, 2017, Dallas, TX, USA

ACM ISBN 978-1-4503-5396-0/17/11.
https://doi.org/10.1145/3139937.3139954

series of signals to the sender in order to shrink the size of the sender's TCP congestion window. It then monitors the IoT device's future traffic so that when an IoT device follows TCP congestion control, FR-WARD relabels the suspicious connection as benign. This allows FR-WARD to avoid throttling a benign connection, and thus, FR-WARD avoids resource penalties on benign traffic. On the other hand, if the IoT device fails to follow TCP congestion control mechanisms and still transmits at the same (high) rate, FR-WARD relabels the connection as malicious and begins throttling the traffic. In this case, FR-WARD allows trace amounts of DDoS traffic to be sent to the destination, but only for a very short duration. In fact, FR-WARD can relabel a suspicious connection as malicious and begin throttling within the length of a couple round trip times; in most cases, this additional delay in throttling will be negligible.

However, FR-WARD must also prevent smarter DDoS attacks than ones that blindly send TCP DDoS traffic. If an attacker knows a source network utilizes FR-WARD, the attacker may design a DDoS attack correspondingly. For example, a compromised device could be programmed to follow TCP congestion control. When this occurs, if FR-WARD only sends signals to initially stop the DDoS attack, the compromised device could quickly return to a high sending rate. In order to prevent this, FR-WARD sends multiple rounds of signals to the sender in order to continuously restrict its TCP congestion window to a reasonable size. This bounds the duration of DDoS traffic from the policed network to a negligible amount without penalizing the traffic from benign IoT devices. Denoting s as the number of signals FR-WARD will send each round trip time, or RTT, we can calculate s as $s = \frac{\log_2(\frac{W}{A}+1)}{\lfloor \frac{D}{RTT} \rfloor}$, where D is the negligible amount of time for a receiver to be under DDoS (an external host may notify FR-WARD of a specific allowable DDoS time, or FR-WARD can generically bound successful DDoS time by a reasonable amount), W is the current window size of the sender, and A is the amount of allowed segments each RTT. By this definition of s, after FR-WARD detects a malicious agflow, when FR-WARD sends s signals to the sender each RTT, and the sender follows TCP congestion control, FR-WARD guarantees the DDoS attack has ended after D seconds. Since s may not be a whole number, FR-WARD sends $\lceil s \rceil$ signals for the first m RTTs and $\lfloor s \rfloor$ signals the next $\lfloor \frac{D}{RTT} \rfloor - m$ RTTs where m is the smallest integer such that $\lceil s \rceil * m + \lfloor s \rfloor * (\lfloor \frac{D}{RTT} \rfloor - m) \geq s * \lfloor \frac{D}{RTT} \rfloor$.

Furthermore, since FR-WARD shapes the traffic of all suspicious connections similarly, it must maintain minimal impact on benign devices but maximal impact on malicious devices. Whether a suspicious connection attempts to send benign traffic or DDoS, the receiver cannot handle transmission rates higher than A for a longer period than D. By definition, after FR-WARD shapes the suspicious connection, benign traffic sends segments at the quickest allowable rate, but malicious traffic sends segments at a manageable rate.

Figure 1 examines how many more retransmissions D-WARD requires than FR-WARD for a benign IoT device. In our simulation, a benign IoT device attempts to send a 15-MB file outside of the network. We observed the number of retransmissions each DDoS defense system required of the IoT device over two main parameters: the sender's congestion window size, W, at the time D-WARD or FR-WARD detects a malicious agflow, and the pre-set fraction, f, of traffic that D-WARD or FR-WARD allows to leave the

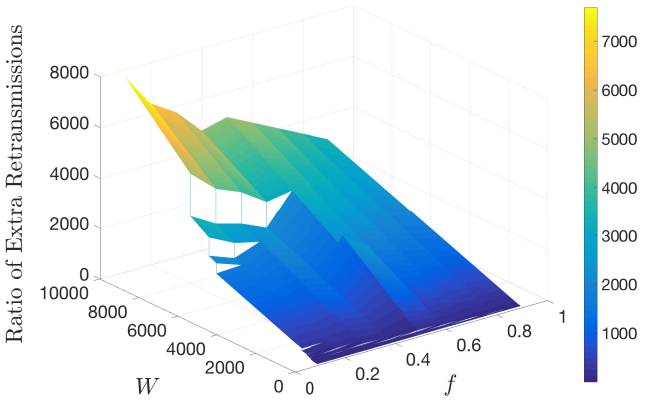

Figure 1: The ratio of extra retransmissions caused by D-WARD over FR-WARD during the transfer of a 15-MB file.

source network during a suspected DDoS attack. Upon detection of a malicious agflow, D-WARD only allows $W * f$ segments to the sender each RTT to mitigate any DDoS attacks. Therefore, when the IoT device follows TCP congestion control, D-WARD drops $W - W * f$ segments every $2RTT$. Furthermore, as W increases or f decreases, D-WARD drops more segments and thus, causes more retransmissions. For FR-WARD, as f decreases, FR-WARD must send more signals to correctly shape the traffic which leads to more retransmissions from the IoT device. This causes the "mountain peak" in Figure 1 when $f \approx \frac{1}{2}$. However, when W surpasses approximately 6500, the amount of traffic D-WARD drops causes overwhelmingly more retransmissions than the number of signals FR-WARD sends even when $f < \frac{1}{2}$.

3 CONCLUSION

FR-WARD is a source-end DDoS defense system that limits DDoS traffic from leaving a policed IoT network, but it also guarantees low negative effect on benign IoT devices. By utilizing the fast retransmit mechanism of TCP, FR-WARD ensures that any DDoS TCP flooding attack that leaves its policed network never surpasses a negligible length of time, but additionally, FR-WARD minimizes resource penalties for benign devices because a benign IoT device always sends traffic at the quickest rate manageable by the receiver and rarely must retransmit dropped packets.

REFERENCES

[1] Samuel Abdelsayed, David Glimsholt, Christopher Leckie, Simon Ryan, and Samer Shami. 2003. An Efficient Filter for Denial-of-Service Bandwidth Attacks. In *Global Telecommunications Conference*, Vol. 3. 1353–1357.
[2] Thomer M. Gil and Massimiliano Poletto. 2001. MULTOPS: A Data-structure for Bandwidth Attack Detection. In *Proceedings of the 10th Conference on USENIX Security Symposium*, Vol. 10. Article 3.
[3] Scott Hilton. 2016. Dyn Analysis Summary Of Friday October 21 Attack. https://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack.
[4] Van Jacobson. 1988. Congestion avoidance and control. In *ACM SIGCOMM Computer Communication Review*, Vol. 18. 314–329.
[5] Jelena Mirkovic and Peter Reiher. 2005. D-WARD: A Source-end Defense Against Flooding Denial-of-Service Attacks. In *IEEE transactions on Dependable and Secure Computing*, Vol. 2. 216–232.