Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

BEYOND ACCURACY: CYBERSECURITY RESILIENCE EVALUATION OF INTRUSION DETECTION SYSTEM AGAINST DOS ATTACKS USING AGENT-BASED SIMULATION

Jeongkeun Shin
L. Richard Carley

Department of Electrical and Computer Engineering
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, USA

Geoffrey B. Dobson
Kathleen M. Carley

School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, USA

ABSTRACT

Machine Learning has become increasingly popular in developing Intrusion Detection Systems (IDS) for cybersecurity. However, the focus has mainly been on achieving high detection accuracy rather than evaluating the impact on cybersecurity resiliency. In this paper, we use agent-based simulation to investigate the impact of different IDS algorithms on the cybersecurity resiliency of organizations under DoS attacks. Our simulation includes a server agent equipped with either Naive Bayes or SMO-based IDS, and a cybercriminal agent capable of launching different types of Denial of Service attacks. Our results suggest that the choice of IDS algorithm can significantly affect an organization's cybersecurity resiliency against DoS attacks. Specifically, while SMO shows better overall accuracy on the KDD Cup 1999 dataset, Naive Bayes-based IDS proves more effective in practice due to its better-balanced detection rates across different types of DoS attacks. Our findings have important implications for improving organizations' cybersecurity posture.

1 INTRODUCTION

With the increasing number of cyber attacks in recent years, organizations are now more concerned about their cybersecurity than ever before. In this paper, we focus on the Denial of Service (DoS) attack, which can result in server down, financial losses, and the organization's reputation damage. Effective detection and response to DoS attacks is essential to minimize the damage and restore normal operations. Recently, to achieve this, numerous machine learning-based intrusion detection system models have been developed (Saranya et al. 2020). However, while most studies focus on achieving high detection accuracy on pre-defined test sets, it's important to evaluate the effectiveness of these models in real-world scenarios where cybercriminals can analyze and exploit the vulnerabilities of IDS. This is challenging, as measuring an IDS's contribution to cybersecurity resiliency, the organization's ability to withstand or recover from cyber attacks (Butler 2018), depends not only on its performance but also on external factors such as the security engineer's ability, the motivation of employees for cybersecurity, the organization's security guidelines and strategies, and the availability of other security software such as firewalls. These outside factors vary greatly from one organization to another, making it challenging to measure how much an IDS can contribute to cybersecurity resiliency in a real-world environment. For example, even if an IDS has vulnerabilities, in some organizations, these vulnerabilities may be covered by other factors, while in other organizations, they may not. To overcome these challenges, we developed an agent-based simulation environment that closely replicates real-world scenarios, including cybercriminal behavior patterns that prioritize exploiting IDS vulnerabilities. By aligning all outside factors in our simulation environment, we aim to investigate the impact of different machine learning models for intrusion detection systems (IDS) on the cybersecurity

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

resiliency of organizations under various denial of service (DoS) attacks. Our study offers valuable insights into the limitations of solely relying on accuracy as a metric for evaluating the effectiveness of ML-based IDS models in enhancing cybersecurity resiliency, highlighting how the accuracy paradox (Thomas and Balakrishnan 2008) can lead to higher accuracy without necessarily improving cybersecurity resiliency.

2 DATASET & INTRUSION DETECTION SYSTEM

In this paper we used the KDD Cup 1999 dataset (Stolfo et al. 1999) to build the machine learning model for the intrusion detection systems. According to the distribution of attack types in the KDD Cup 1999 dataset (Aghdam and Kabiri 2016), we can observe that many attack types, such as Mail-bomb, Apache2, and Process-table attacks, have data available only in the test set. To ensure equal distribution of data from all labels in both the training and test sets, we merged the original KDD Cup 1999 training and test sets. We then assigned 80% of the data from each label to the new training set and the remaining 20% to the new test set. Table 2 presents the distribution of data for each label in the new training and test sets after the merging and partitioning process.

Using the aforementioned training and test datasets, we employed the Weka software (Holmes et al. 1994; Hall et al. 2009) to develop a machine learning-based intrusion detection system (Meena and Choudhary 2017) using both the Naive Bayes algorithm (John and Langley 1995) and the Sequential Minimal Optimization (SMO) algorithm for Support Vector Classifier (Platt 1999). The results of our experiments on the modified test set, including overall accuracy, accuracy for DoS detection, and specific DoS attack detection accuracy, are presented in Table 1.

Table 1: Intrusion detection system performance with Naive Bayes and SMO algorithms.

	Naive Bayes	SMO
Overall Accuracy	87.26%	99.61%
DoS Accuracy	99.69%	99.92%
Neptune	99.23%	99.99%
Smurf	99.88%	99.99%
Pod	95.77%	95.77%
Teardrop	98.99%	98.99%
Land	100.00%	100.00%
Back	96.06%	9.68%
Apache2	98.74%	95.59%
Udpstorm	0.00%	0.00%
Processtable	97.36%	99.34%
Mailbomb	97.80%	99.70%

Table 1 illustrates that the SMO algorithm outperforms the Naive Bayes algorithm in terms of both overall accuracy and DoS detection accuracy. However, both algorithms fail to detect the Udpstorm attack. On a closer examination, it is evident that while the Naive Bayes algorithm demonstrates balanced high detection accuracy across all specific DoS attacks, the SMO algorithm exhibits a significantly low detection rate in the Back attack, which may be a potential vulnerability that cybercriminals could exploit. These IDS models will be used to detect the DoS attacks in our simulation model.

3 RELATED WORKS

In recent years, there has been a growing interest in using machine learning (ML) for developing Intrusion Detection Systems (IDS). The KDD Cup 1999 dataset (Stolfo et al. 1999), which incorporates DoS attacks still relevant today, has been widely used as a benchmark dataset by ML researchers to evaluate the

Table 2: Comparison of attack type distribution in original and modified KDD Cup 99 dataset.

Category	Attack Type	Original Training Set	Original Test Set	Modified Training Set	Modified Test Set
Normal	Normal	972781	60593	826699	206675
Denial of Service (DoS)	Neptune	1072017	58001	904014	226004
	Smurf	2807886	164091	2377581	594396
	Pod	264	87	280	71
	Teardrop	979	12	792	199
	Land	21	9	24	6
	Back	2203	1098	2640	661
	Apache2	-	794	635	159
	Udpstorm	-	2	1	1
	Processtable	-	759	607	152
	Mailbomb	-	5000	4000	1000
Remote to Local (R2L)	Guess_passwd	53	4367	3536	884
	Ftp_write	8	3	8	3
	Imap	12	1	10	3
	Phf	4	2	4	2
	Multihop	7	18	20	5
	Warezmaster	20	1602	1297	325
	Warezclient	1020	-	816	204
	Snmpgetattack	-	7741	6192	1549
	Named	-	17	13	4
	Xlock	-	9	7	2
	Xsnoop	-	4	3	1
	Sendmail	-	17	13	4
User to Root (U2R)	Buffer_overflow	30	22	41	11
	Loadmodule	9	2	8	3
	Perl	3	2	4	1
	Rootkit	10	13	18	5
	Spy	2	-	1	1
	Xterm	-	13	10	3
	Ps	-	16	12	4
	Httptunnel	-	158	126	32
	Sqlattack	-	2	1	1
	Worm	-	2	1	1
	Snmpguess	-	2406	1924	482
Probe	Port-sweep	10413	354	8613	2154
	IPsweep	12481	306	10229	2558
	Nmap	2316	84	1920	480
	Satan	15892	1633	14020	3505
	Saint	-	736	588	148
	Mscan	-	1053	842	211

accuracy of IDS. However, this dataset is primarily composed of network traffic data with imbalanced classes, leading to the accuracy paradox (Thomas and Balakrishnan 2008; Valverde-Albacete and Peláez-Moreno 2014; Alabdallah and Awad 2018). To address this issue, various techniques have been proposed, such as combining machine learning models with stratified sampling and weighted support vector machine (Alabdallah and Awad 2018), or with extreme learning machines with kernel (Awad and Alabdallah 2019).

There have been several models developed to simulate human organizations and Denial of Service (DoS) campaigns. Kumar and Carley developed a network simulation model to understand the flow pattern of Internet traffic in a DDoS attack (Kumar and Carley 2017). Dobson and Carley developed the Cyber-FIT framework (Dobson and Carley 2017; Dobson and Carley 2018; Dobson and Carley 2021) to model cyber warfare and estimate the effectiveness of military cyber forces against various cyber attacks, including DoS, Phishing, and Routing Protocol Attack. The OSIRIS framework (Shin et al. 2022a; Shin et al. 2022b; Shin et al. 2023) models a human organization with realistic end-user behavior patterns and has been used as a testbed to simulate the potential overall organizational damage from various cyberattacks, such as phishing and ransomware, and to evaluate the effectiveness of cybersecurity strategies.

In this paper, we aim to measure and compare the cybersecurity resilience of different IDSs through the simulations by embedding machine learning-based IDSs into the server agent in the OSIRIS framework.

4 SIMULATION MODEL DESIGN

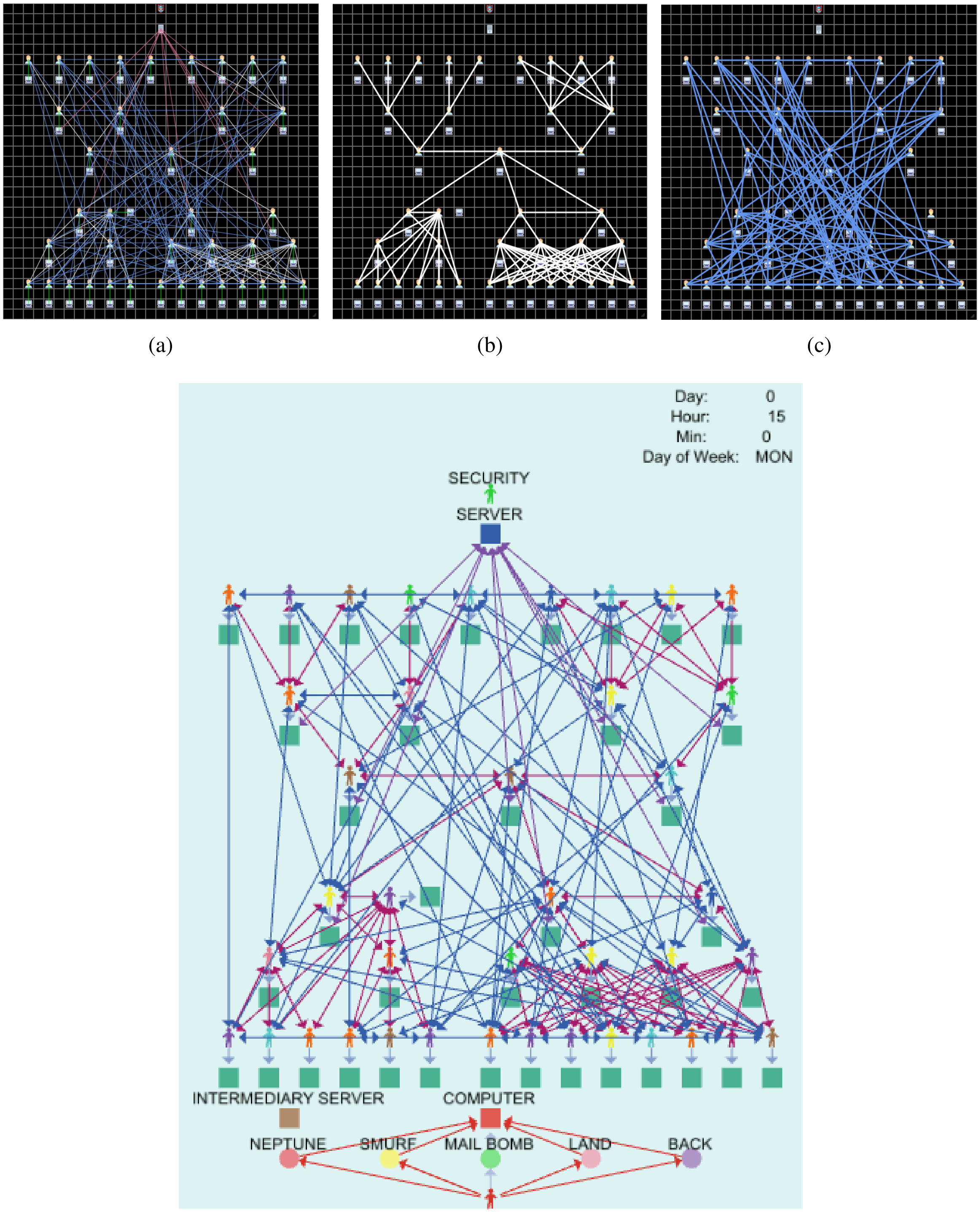
In this section, we will describe our simulation model. Specifically, we employed OSIRIS (Shin et al. 2022b; Shin et al. 2023) to construct a virtual organization that includes end-user agents, computing device agents (including server agents), networks among end-user agents, and connections between computing devices. In our simulation, we imported the attacker agents and defender agents of Cyber-FIT (Dobson and Carley 2017; Dobson and Carley 2021) as cybercriminal agents and security professional agents, respectively. The most recent versions of OSIRIS and Cyber-FIT frameworks were built using Repast Simphony (North et al. 2013). In Repast Simphony (North et al. 2013), time is measured in ticks, where each tick represents a unit of simulation time. One tick in our model corresponds to one real-world minute. This choice aligns with the empirical data (Park 2021), which measures server downtime caused by DDoS attacks in minutes. Thus, equating one tick to one minute adequately captures the dynamics and temporal aspects of cybersecurity events.

4.1 Organization Model

We used the OSIRIS framework (Shin et al. 2022b; Shin et al. 2023) to create a virtual organization. Specifically, we created a virtual small and medium-sized business with 40 end user agents, each of which was assigned a personal computing device agent. Additionally, we assigned one server agent and one security professional agent to the organization. Then, we created several types of networks within the organization (see Figure 1a), with end users who work together being connected by formal relationships (see Figure 1b). To generate the informal relationship network between end user agents (see Figure 1c), we used the Erdős-Rényi random network (Erdos 1959) generator with a probability parameter of $p = 0.1$, utilizing the random network generator available in OSIRIS. Only end user agents with high levels and security agents have direct access to the server agent. Once we finished designing the agents and networks in the organization, we executed the simulation. The OSIRIS UI automatically exported the virtual organization to the Repast Simphony simulator. When we started the simulation, the cybercriminal agent launched a cyberattack campaign against our virtual small and medium-sized company.

4.1.1 Server Agent with Intrusion Detection System (IDS)

In this model, the server agent is equipped with one of the intrusion detection systems (IDS) described in Section 2. To facilitate efficient handling of the modified test set, we separated it by attack type and saved each segment using the respective attack type name. During each tick, if the server agent is not under


(a)

(b)

(c)

(d)

Figure 1: Virtual small and medium-sized business in OSIRIS (a) Network architecture designed in OSIRIS UI (b) Formal relationships modeled in OSIRIS UI (c) Informal relationship generated in OSIRIS UI (d) Simulation environment in OSIRIS.

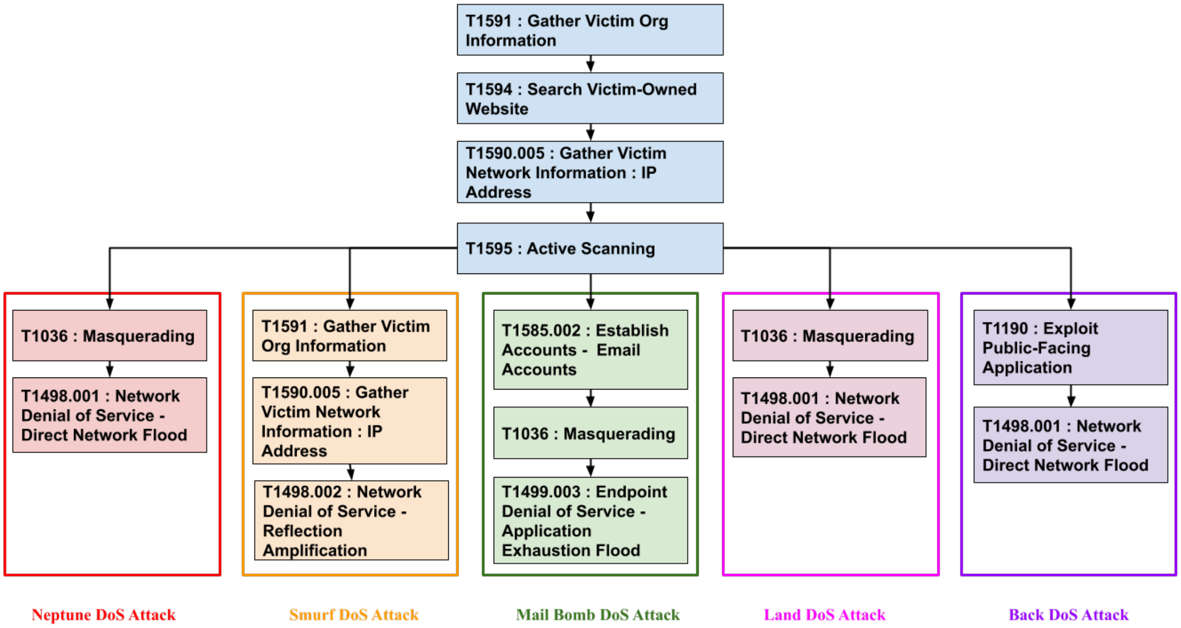

T1591 : Gather Victim Org Information
T1594 : Search Victim-Owned Website
T1590.005 : Gather Victim Network Information : IP Address
T1595 : Active Scanning
T1036 : Masquerading
T1498.001 : Network Denial of Service - Direct Network Flood
T1591 : Gather Victim Org Information
T1590.005 : Gather Victim Network Information : IP Address
T1498.002 : Network Denial of Service - Reflection Amplification
T1585.002 : Establish Accounts - Email Accounts
T1036 : Masquerading
T1499.003 : Endpoint Denial of Service - Application Exhaustion Flood
T1036 : Masquerading
T1498.001 : Network Denial of Service - Direct Network Flood
T1190 : Exploit Public-Facing Application
T1498.001 : Network Denial of Service - Direct Network Flood
Neptune DoS Attack
Smurf DoS Attack
Mail Bomb DoS Attack
Land DoS Attack
Back DoS Attack

Figure 2: MITRE ATT&CK-based Denial of Service (DoS) attack model.

attack, it randomly selects and generates a data point from the ‘normal’ dataset. However, in the event of an attack, the server agent randomly selects and generates a data point from the corresponding attack type file. The IDS equipped in the server agent promptly evaluates the type of attack being directed towards the server as soon as the generated data is received. If the generated data is classified as one of the Neptune, Smurf, Mailbomb, Land, and Back DoS attacks, the IDS immediately reports to the security professional agent that the server is currently under a Denial of Service (DoS) attack. Although IDS may fail to detect DoS attacks, organization members may eventually realize that they are under attack. Based on real-world simulation results indicating that small and medium-sized companies typically take an average of 9 minutes to detect DDoS attacks (Park 2021), we set a time frame of 7 to 11 minutes for either an organization member to recognize and directly report to the security professional agent or for the security professional agent to detect the DoS attack while monitoring the server.

4.2 Denial of Service (DoS) Attack Campaign

As shown at the bottom of Figure 1d, in the simulation, we imported the cybercriminal agent from Cyber-FIT (Dobson and Carley 2017; Dobson and Carley 2021) into the simulation environment. This agent is capable of executing five different types of Denial of Service (DoS) attacks (Neptune, Smurf, Mailbomb, Land, and Back) using a single computer. Since this campaign is not Distributed Denial of Service (DDoS) campaign, botnets are not used. Even though Cyber-FIT’s attacking troops followed the steps in the cyber-kill chain (Dobson et al. 2018), the cybercriminal agent in OSIRIS performs attacks by following the MITRE ATT&CK (Strom et al. 2018) tactics and techniques as it provides a more detailed approach to attacking steps. Based on the descriptions of Neptune, Smurf, Mailbomb, Land, and Back attacks (Haddadi and Beghdad 2018), we carefully break down each attack into several MITRE ATT&CK (Strom et al. 2018) techniques to provide a more detailed and comprehensive understanding of the attacks (Figure 2).

The denial of service (DoS) attack campaign will last for 129,600 ticks (90 days). During the first 43,200 ticks (30 days), the cybercriminal agent will randomly select one of five DoS attacks during each session, targeting the virtual organization’s server agent. A session’s attack will end when the security professional agent begins to mitigate the attack. After each session, the cybercriminal agent will update the average server downtime caused by each attack. After the first 43,200 ticks have passed, the cybersecurity team will analyze the weaknesses of the organization by identifying the most effective DoS attack that caused the longest server downtime in previous attacks. From that point on, the cybercriminal agent will concentrate on attacking the organization's server agent with the DoS attack that caused the longest server downtime until the simulation ends at 129,600 ticks.

The remaining part of this section will provide a detailed description of the five DoS attack methods (Neptune, Smurf, Mailbomb, Land, and Back) in the context of the MITRE ATT&CK framework (Strom et al. 2018).

4.2.1 Reconnaissance

Before launching any Denial of Service (DoS) attack, the cybercriminal agent gathers necessary information about the target organization using techniques categorized under the Reconnaissance tactic in the MITRE ATT&CK framework (MITRE 2020). This process is summarized in the following four attack techniques.

1. Gather Victim Org Information (T1591) : The cybercriminal agent collects critical information about the target organization that can be useful during the attack, such as the key end-user agents in the organization and their duties within the organization.
2. Search Victim-Owned Website (T1594): The cybercriminal agent obtains valuable information about the target organization by searching its website, such as contact details and employee information.
3. Gather Victim Network Information : IP Address (T1590.005) : The cybercriminal agent acquires the target organization's IP address to use during the DoS attack.
4. Active Scanning (T1595): The cybercriminal agent examines the victim's infrastructure through network traffic with the intention of acquiring data for use in DoS attacks.

4.2.2 Neptune

The Neptune attack (Marchette 2001; Chang 2002; Li et al. 2010; Haddadi and Beghdad 2018) involves flooding a TCP server with spoofed SYN packets, exploiting flaws in the TCP protocol. This causes the server to create half-open connections, eventually leading to resource exhaustion and the rejection of new connections from authorized clients.

1. Masquerading (T1036) : The cybercriminal agent spoofs or generates fake IP addresses in the SYN packets.
2. Network Denial of Service - Direct Network Flood (T1498.001) : The cybercriminal agent floods the target server with a large volume of SYN packets in order to overwhelm the target server and disrupt normal operations.

4.2.3 Smurf

The Smurf attack (Marchette 2001; Chang 2002; Haddadi and Beghdad 2018) sends ICMP echo packets to the intermediary server with the target server's IP address as the source address. Then, the intermediary server broadcasts the ICMP echo packets to all hosts in the network, causing them to respond with ICMP echo reply packets to the target server. This overwhelms the target server with traffic, causing the server shuts down or become unreachable.

1. Gather Victim Org Information (T1591) : The cybercriminal agent identifies an organization with a server that can be used as an intermediary in the attack.
2. Gather Victim Network Information - IP address (T1590.005) : The cybercriminal agent obtains the IP address of the identified intermediary server.
3. Network Denial of Service - Reflection Amplification (T1498.002) : The cybercriminal agent sends ICMP echo packets to the intermediary server, using the target server's IP address as the source

address. The intermediary server and its hosts respond with ICMP echo reply packets, overwhelming the target server with traffic.

4.2.4 Mailbomb

The Mailbomb attack (Marchette 2001; Kim et al. 2011; Haddadi and Beghdad 2018) is a type of DoS attack in which an attacker sends a large number of emails to a target email address to overwhelm or slow down the target organization's mail server. To evade spam filters, the emails are sent from different email addresses and with different messages.

1. Establish Accounts - Email Accounts (T1585.002) : The cybercriminal agent creates a large number of email accounts to be used for sending spam emails to the target organization.
2. Masquerading (T1036) : The cybercriminal agent masks their true identity by faking the sender's email address, making it difficult for the target organization to trace the source of the attack.
3. Endpoint Denial of Service - Application Exhaustion Flood (T1499.003) : The cybercriminal agent sends numerous emails to the target organization's email address, overwhelming the mail server's resources and causing a denial of service (DoS) by preventing legitimate access to the email service.

4.2.5 Land

The Land attack (Marchette 2001; Li et al. 2010; Haddadi and Beghdad 2018) sends TCP SYN packets to the target server with both the source and destination IP addresses set to the target's IP address. This can cause the server to lock up and require a reboot.

1. Masquerading (T1036) : The cybercriminal agent spoofs its identity by altering the source IP address of the TCP SYN packets to make it appear that the attack is coming from the target server's IP address.
2. Network Denial of Service - Direct Network Flood (T1498.001) : The cybercriminal agent sends a flood of TCP SYN packets to the target server to cause a denial of service (DoS).

4.2.6 Back

The Back attack (Marchette 2001; Li et al. 2010; Haddadi and Beghdad 2018) sends requests with thousands of front slashes to the Apache Web server, causing it to slow down. This attack exploits the vulnerability of Apache web server applications that cannot efficiently handle unusual input.

1. Exploit Public-Facing Application (T1190) : The cybercriminal agent identifies a vulnerability in the Apache web server that causes a slowdown when processing requests with a large number of front slash characters.
2. Network Denial of Service - Direct Network Flood (T1498.001) : The cybercriminal agent sends a large number of requests with thousands of front slashes to the target organization's server, causing a Denial of Service (DoS).

4.3 Security Professional Agent

In the previous work using the OSIRIS framework (Shin et al. 2022b; Shin et al. 2023), the security professional agents were primarily responsible for monitoring the computing devices of end-user agents or managing the human firewall to prevent cyber threats and improve cybersecurity. However, in this paper, we assigned the security professional agents a different role: mitigating Denial of Service (DoS) attacks once they receive warnings from the intrusion detection system (IDS). After receiving a warning from the IDS system, the security professional agent checks if the server is actually under a Denial of Service (DoS) attack (verifying the IDS signal as a true positive), and if so, it begins mitigating the attack and recovering the server. We have deployed a single security professional agent within our virtual organization and assume that it is capable of responding to and mitigating different types of DoS attacks by utilizing various defense mechanisms (Chen et al. 2004).

Based on the results of a simulated DDoS cyberattack in small and medium-sized businesses, where it took an average of 13 minutes to respond and mitigate attacks (22 minutes in total, including 9 minutes for detection) (Park 2021), we estimate that the security professional agent will require 11 to 15 minutes to mitigate each DoS campaign after detection.

5 VIRTUAL EXPERIMENTS

In this section, we will describe our virtual experiments using two different intrusion detection systems: Naive Bayes (John and Langley 1995) and SMO (Platt 1999) algorithms. We utilized our simulation model to conduct the experiments, and the simulation settings are summarized in Table 3. We performed a sensitivity analysis on the impact of different time intervals between attacks on server downtime in the absence of an IDS, as well as with Naive Bayes and SMO IDSs, for varying interarrival time between attacks (2880, 1440, 720, 360, and 180 minutes). Our analysis consisted of 15 cells, with three IDSs tested across five time intervals. We ran 100 simulations for each cell, resulting in a total of 1500 simulations. The experiment result is summarized in Figure 3.

Table 3: Simulation summary.

Type	Name	Implication
Input	Virtual Organization	Virtual small and medium-sized business organizations composed of 40 employees, computing device agents, and server agents
	DoS Attack Campaign	Five different MITRE ATT&CK-based DoS attacks: Neptune, Smurf, Mailbomb, Land, Back
	Security System	One security professional agent with capability to mitigate the DoS attacks
Output	Server Downtime	The amount of time (in minutes) that the organization's server agent was down during the 90 days DoS Attack Campaign
Parameter	Intervals Between Attacks	The interarrival time between attacks : 2880, 1440, 720, 360, and 180 minutes
	Intrusion Detection System	The machine learning algorithm used to build the intrusion detection system (IDS) : No IDS or Naive Bayes or SMO
	Number of Simulations	100

Based on the results of the virtual experiment in Figure 3, it can be observed that both SMO-based IDS and Naive Bayes-based IDS were effective in mitigating the DoS attack when compared to the absence of an IDS. Moreover, although the SMO-based IDS had higher overall accuracy and DoS detection accuracy compared to the Naive Bayes-based IDS as described in Table 1, it caused longer server downtime for all five different cases. In the absence of IDS, the average server downtime was approximately 21 minutes. On the other hand, when IDS was implemented, the average server downtime per one DoS attack decreased to approximately 13 minutes for the Naive Bayes model and 16 minutes for the SMO model. Therefore, both IDS algorithms reduced the server downtime per attack, but the Naive Bayes-based IDS was found to be more effective in mitigating the impact of DoS attacks. We also conducted a two-tailed t-test in the 99% significance level for each interarrival time between attacks cases to compare the results of the Naive Bayes model and the SMO model. The obtained p-values were extremely low (close to zero), indicating strong evidence that the results from the Naive Bayes model and SMO model are statistically significantly

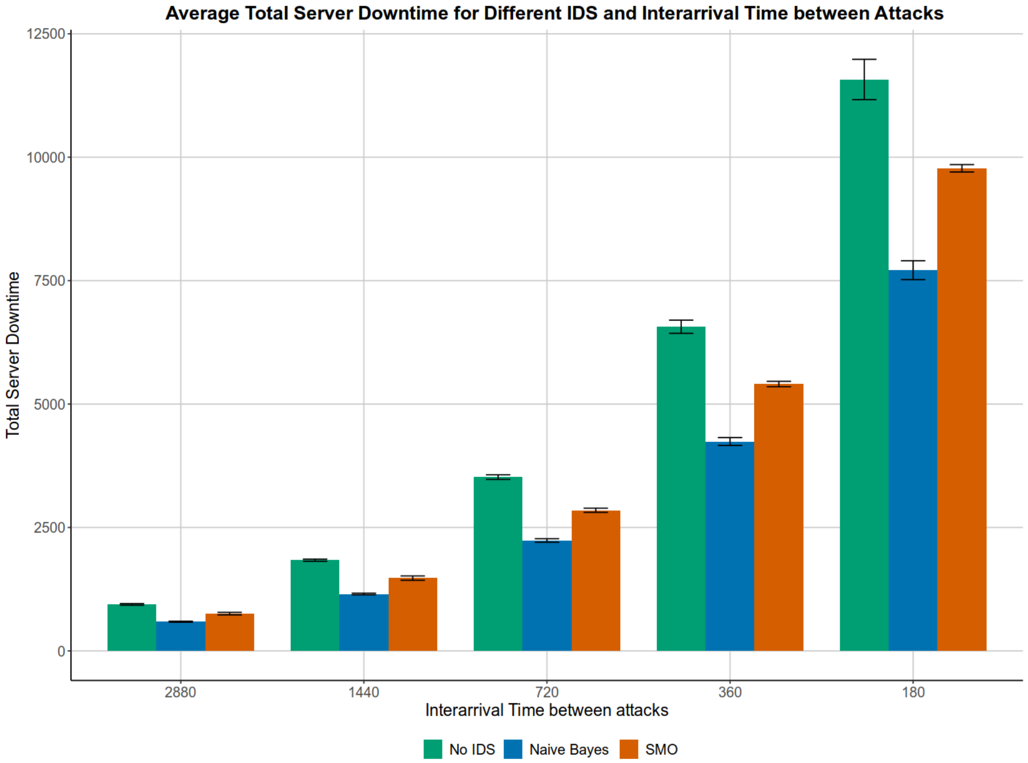


Figure 3: Virtual experiment result.

different from each other. Additionally, it was found that as the interarrival time between attacks becomes shorter, the difference in server downtime between the SMO-based IDS and the Naive Bayes-based IDS increases. This suggests that as interarrival time between DoS attacks decreases, more damage is expected for the organization. Thus, SMO-based IDS may not be the optimal choice for minimizing server downtime.

6 DISCUSSION AND CONCLUSION

The research presented in this paper has various limitations. Although the Naive Bayes-based IDS model demonstrates greater effectiveness in terms of cybersecurity resiliency compared to the SMO-based IDS model, it is also observed to produce more false alarms when dealing with normal feature vectors. Since security professionals have limited attention, these false alarms may impede their ability to promptly respond to other types of cyber attacks in the real world. In future work, we will devise methods to measure the damage caused by false alarms and strive to develop a more reliable evaluation method for the Intrusion Detection System's cybersecurity resilience. Additionally, validating the results of this study in the real world is challenging since evaluating the cybersecurity resilience of an IDS in isolation is nearly impossible. IDSs typically work in conjunction with other cybersecurity software and tools, making it difficult to isolate their effectiveness. Future studies could explore the integration of IDSs with other cybersecurity tools and assess their combined effectiveness in enhancing an organization's cybersecurity resiliency.

In conclusion, we used the agent-based simulation to evaluate the cybersecurity resilience of Intrusion Detection Systems. Our study demonstrated that while employing any IDS in the organization was helpful in mitigating server downtime during DoS attack campaigns, relying solely on overall accuracy and DoS detection accuracy can be misleading. It can result in lower cybersecurity resilience and longer server downtime during attacks compared to other IDSs with lower overall and DoS detection accuracy but more balanced detection rates among various DoS attacks. This work can inform cybersecurity professionals and organizations in making informed decisions about their IDS and overall cybersecurity strategies.

ACKNOWLEDGMENTS

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by the Minerva Research Initiative under Grant #N00014-21-1-4012 and by the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research or the US Government.

REFERENCES

Aghdam, M. H., and P. Kabiri. 2016. "Feature Selection for Intrusion Detection System Using Ant Colony Optimization". *International Journal of Network Security* 18(3):420–432.

Alabdallah, A., and M. Awad. 2018. "Using Weighted Support Vector Machine to Address the Imbalanced Classes Problem of Intrusion Detection System". *KSII Transactions on Internet and Information Systems (TIIS)* 12(10):5143–5158.

Awad, M., and A. Alabdallah. 2019. "Addressing Imbalanced Classes Problem of Intrusion Detection System Using Weighted Extreme Learning Machine". *International Journal of Computer Networks & Communications (IJCNC) Vol* 11(5):39–58.

Butler, C. 2018. "Five Steps to Organisational Resilience: Being Adaptive and Flexible during Both Normal Operations and Times of Disruption". *Journal of Business Continuity & Emergency Planning* 12(2):103–112.

Chang, R. K. 2002. "Defending against Flooding-Based Distributed Denial-of-Service Attacks: A Tutorial". *IEEE Communications Magazine* 40(10):42–51.

Chen, L.-C., T. A. Longstaff, and K. M. Carley. 2004. "Characterization of Defense Mechanisms against Distributed Denial of Service Attacks". *Computers & Security* 23(8):665–678.

Dobson, G., A. Rege, and K. Carley. 2018. "Informing Active Cyber Defence with Realistic Adversarial Behaviour". *Journal of Information Warfare* 17(2):16–31.

Dobson, G. B., and K. M. Carley. 2017. "Cyber-FIT: An Agent-Based Modelling Approach to Simulating Cyber Warfare". In *Social, Cultural, and Behavioral Modeling: 10th International Conference, SBP-BRiMS 2017, Washington, DC, USA, July 5-8, 2017, Proceedings 10*, 139–148. Springer.

Dobson, G. B., and K. M. Carley. 2018. "A Computational Model of Cyber Situational Awareness". In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11*, 395–400. Springer.

Dobson, G. B., and K. M. Carley. 2021. "Cyber-FIT Agent-Based Simulation Framework Version 4". Technical Report CMU-ISR-21-113, Center for the Computational Analysis of Social and Organizational Systems, Pittsburgh, PA.

Erdos, P. 1959. "On Random Graphs". *Mathematicae* 6:290–297.

Haddadi, M., and R. Beghdad. 2018. "DoS-DDoS: Taxonomies of Attacks, Countermeasures, and Well-known Defense Mechanisms in Cloud Environment". *Edpacs* 57(5):1–26.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. 2009. "The WEKA Data Mining Software: An Update". *ACM SIGKDD Explorations Newsletter* 11(1):10–18.

Holmes, G., A. Donkin, and I. H. Witten. 1994. "WEKA: A Machine Learning Workbench". In *Proceedings of ANZIIS'94-Australian New Zealnd Intelligent Information Systems Conference*, 357–361. IEEE.

John, G. H., and P. Langley. 1995. "Estimating Continuous Distributions in Bayesian Classifiers". In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 338–345.

Kim, W., O.-R. Jeong, C. Kim, and J. So. 2011. "The Dark Side of the Internet: Attacks, Costs and Responses". *Information Systems* 36(3):675–705.

Kumar, S., and K. M. Carley. 2017. "Simulating DDOS Attacks on the US Fiber-optics Internet Infrastructure". In *2017 Winter Simulation Conference (WSC)*, 1228–1239. IEEE.

Li, J., Y. Liu, and L. Gu. 2010. "DDoS Attack Detection Based on Neural Network". In *2010 2nd International Symposium on Aware Computing*, 196–199. IEEE.

Marchette, D. J. 2001. *Computer Intrusion Detection and Network Monitoring: A Statistical Viewpoint*. 1st ed. New York: Springer.

Meena, G., and R. R. Choudhary. 2017. "A Review Paper on IDS Classification using KDD 99 and NSL KDD Dataset in WEKA". In *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, 553–558. IEEE.

MITRE 2020. "Reconnaissance, Tactic TA0043 - Enterprise | MITRE ATT&CK". https://attack.mitre.org/tactics/TA0043/, accessed 10th April.

North, M. J., N. T. Collier, J. Ozik, E. R. Tatara, C. M. Macal, M. Bragen, and P. Sydelko. 2013. "Complex Adaptive Systems Modeling with Repast Simphony". *Complex Adaptive Systems Modeling* 1:1–26.

Park, Sae-jin 2021. "Small and Medium Companies Take Average 9 Minutes to Detect Cyberattack: Simulation Data". https://www.ajudaily.com/view/20210706153945251, accessed 10th April.

Platt, J. C. 1999. "Fast Training of Support Vector Machines using Sequential Minimal Optimization". *Advances in Kernel Methods*:185–208.

Saranya, T., S. Sridevi, C. Deisy, T. D. Chung, and M. A. Khan. 2020. "Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review". *Procedia Computer Science* 171:1251–1260.

Shin, J., L. R. Carley, G. B. Dobson, and K. M. Carley. 2023. "Modeling and Simulation of the Human Firewall Against Phishing Attacks in Small and Medium-Sized Businesses". In *2023 Annual Modeling and Simulation Conference (ANNSIM)*, 369–380. IEEE.

Shin, J., G. B. Dobson, K. M. Carley, and L. R. Carley. 2022a. "Leveraging OSIRIS to Simulate Real-world Ransomware Attacks on Organization". *2022 Winter Simulation Conference (WSC) Poster Session*.

Shin, J., G. B. Dobson, K. M. Carley, and L. R. Carley. 2022b. "OSIRIS: Organization Simulation in Response to Intrusion Strategies". In *Social, Cultural, and Behavioral Modeling: 15th International Conference, SBP-BRiMS 2022, Pittsburgh, PA, USA, September 20–23, 2022, Proceedings*, 134–143. Springer.

Stolfo, S., W. Fan, W. Lee, A. Prodromidis, and P. Chan. 1999. "KDD Cup 1999 Dataset". *UCI KDD Repository*. https://archive.ics.uci.edu/dataset/130/kdd+cup+1999+data, accessed 10th April.

Strom, B. E., A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, and C. B. Thomas. 2018. "MITRE ATT&CK: Design and Philosophy". Technical Report 10AOH08A-JC, The MITRE Corporation, McLean, VA.

Thomas, C., and N. Balakrishnan. 2008. "Improvement in Minority Attack Detection with Skewness in Network Traffic". In *Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2008*, Volume 6973, 226–237. SPIE.

Valverde-Albacete, F. J., and C. Peláez-Moreno. 2014. "100% Classification Accuracy Considered Harmful: The Normalized Information Transfer Factor Explains the Accuracy Paradox". *PloS ONE* 9(1):e84217.

AUTHOR BIOGRAPHIES

JEONGKEUN SHIN is a Ph.D. student in the Department of Electrical and Computer Engineering at Carnegie Mellon University. He is a member of the Center for Computational Analysis of Social and Organization Systems (CASOS). His research includes modeling and simulation of human and organizational behaviors relevant to cybersecurity. He holds a bachelor's degree in computer science from the University of Michigan and a master's degree in electrical and computer engineering from Carnegie Mellon University. His email address is jeongkes@andrew.cmu.edu.

GEOFFREY B. DOBSON is a Systems Scientist at the Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University's School of Computer Science. His research focuses on modeling and simulating the human behavioral and social aspects of cyber conflict. He is an officer in the United States Air Force Reserve stationed at the Air Force Research Laboratory, Wright-Patterson Air Force Base, OH where he oversees a research portfolio focused on human performance in cyber missions. His email address is gdobson@cs.cmu.edu.

L. RICHARD CARLEY received an S.B. in 1976, an M.S. in 1978, and a Ph.D. in 1984, all from the Massachusetts Institute of Technology. He is the professor of Electrical and Computer Engineering Department at Carnegie Mellon University (CMU) in Pittsburgh, Pennsylvania. Dr. Carley's research interests include analog and RF integrated circuit design in deeply scaled CMOS technologies, and novel micro-electromechanical and nano-electro-mechanical device design and fabrication. For the past several years, Dr. Carley has studied the design of efficient RF Power Amplifiers in advanced BiCMOS technologies. Dr. Carley has been granted 27 patents, authored or co-authored over 250 technical papers, and authored or co-authored over 20 books and/or book chapters. He has won numerous awards including Best Technical Paper Awards at both the 1987 and the 2002 Design Automation Conference (DAC), a Most Influential Paper award from DAC, and a Best Panel Session award at ISSCC in 1993. In 1997, Dr. Carley co-founded the analog electronic design automation startup, Neolinear, which was acquired by Cadence in 2004. His email address is lrc@andrew.cmu.edu.

KATHLEEN M. CARLEY (H.D. University of Zurich, Ph.D. Harvard, S.B. MIT) is a Professor of Societal Computing, Software and Societal Systems Department (S3D), Carnegie Mellon University; Director of the Center for Computational Analysis of Social and Organizational Systems (CASOS), Director of the Center for Informed Democracy and Social Cybersecurity (IDeaS), and CEO of Netanomics. Her research blends computer science and social science to address complex real world issues such as social cybersecurity, disinformation, disease contagion, disaster response, and terrorism from a high dimensional network analytic, machine learning, and natural language processing perspective. She and her groups have developed network and simulation tools, such as ORA, that can assess network and social media data. Her email address is kathleen.carley@cs.cmu.edu.