Computer Networks 222 (2023) 109553

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Survey paper

Distributed denial of service attack prediction: Challenges, open issues and opportunities

Anderson Bergamini de Neira [a,*], Burak Kantarci [b], Michele Nogueira [c]

[a] Department of Informatics, Federal University of Paraná (UFPR), Brazil
[b] School of Electrical Engineering and Computer Science, University of Ottawa (uOttawa), Canada
[c] Computer Science Department, Federal University of Minas Gerais (UFMG), Brazil

ARTICLE INFO

Keywords:
Cybersecurity
DDoS attack prediction
Survey
Network security

ABSTRACT

Distributed Denial of Service (DDoS) attack is one of the biggest cyber threats. DDoS attacks have evolved in quantity and volume to evade detection and increase damage. Changes during the COVID-19 pandemic have left traditional perimeter-based security measures vulnerable to attackers that have diversified their activities by targeting health services, e-commerce, and educational services. DDoS attack prediction searches for signals of attack preparation to warn about the imminence of the attack. Prediction is necessary to handle high-volumetric DDoS attacks and to increase the time to defend against them. This survey article presents the classification of studies from the literature comprising the current state-of-the-art on DDoS attack prediction. It highlights the results of this extensive literature review categorizing the works by prediction time, architecture, employed methodology, and the type of data utilized to predict attacks. Further, this survey details each identified study and, finally, it emphasizes the research opportunities to evolve the DDoS attack prediction state-of-the-art.

1. Introduction

The Internet makes various applications and services available, allowing users to access multiple forms of work and entertainment. During the coronavirus pandemic, connectivity enabled online learning, conferences, medical appointments by video call, and teleworking. Despite the benefits available through computing, it is necessary to take several precautions to avoid losses. Denial of service (DoS) attacks cause significant losses [1,2]. Among the existing cyber threats, they are one of the most dangerous [3–5]. In 2020, experts identified 1,560,000 more DDoS attacks than in 2019; this shows an average growth of 130,000 attacks per month [6]. The year 2021 was also intense in the amount of DDoS attacks. Experts identified that the number of DDoS attacks in the first half of 2021 was lower than in 2020. This decrease occurred because the beginning of the coronavirus pandemic greatly influenced the number of DDoS attacks in the first half of 2020 [7,8]. In the second half of 2021, the number of attacks increased from DDoS attacks compared to the second half of 2020 [9,10]. Even in 2022, the amount of attacks remains high. In the first half of 2022, experts identified 6,019,888 DDoS attacks coming from 190 different countries [11]. Platforms offer DDoS attacks as a service at a low price (US$ 100.00 per day [12] or US$10.00 per hour [13]), contributing to a substantial increase in the number of attacks.

DDoS attacks target organizations from government to domestic users [14,15] with different motivations. In the third quarter of 2020, 76.92% of the DDoS attacks targeted gambling sites or the gaming industry [16]. In 2020, the attackers diversified their activities, attacking health, e-commerce, and educational services [6,17]. Several teachers and thousands of students in the United Kingdom suffered interruptions in their online education platform [18]. In Brazil, a six-day DDoS attack disrupted a website reporting human rights crimes [19]. Finally, the New Zealand stock exchange has experienced a DDoS attack for two days [20]. In the first quarter of 2021, the countries that experienced the most DDoS attacks were the United States of America (USA), China, and the Netherlands. The countries that most originated attacks were the United States of America, China, and Canada [7].

Network administrators have limited time to avoid the negative effects of DDoS attacks [21] after the attack has been launched. In 2014, Gartner's estimate showed that the average cost of service disruption could reach US$ 5,600 per minute [22]. In 2020, specialists revisited this estimate showing that an hour of outage could cost around US$ 300,000 [23]. Even with these reference values, it is not trivial to quantify the losses related to service disruption. Ongoing research addresses various efforts to prevent or reduce disruption-related harm

* Corresponding author.
E-mail address: abneira@inf.ufpr.br (A.B. de Neira).

https://doi.org/10.1016/j.comnet.2022.109553
Received 4 November 2022; Received in revised form 27 December 2022; Accepted 28 December 2022
Available online 3 January 2023

caused by DDoS attacks. Examples are firewalls and other actions such as keeping infrastructure up-to-date and disabling inactive services.

Attack detection and mitigation mechanisms are also common [24]. However, alleviating DDoS attack damage after it is launched is challenging [25,26]. There are records of DDoS attacks reaching terabits per second or millions of requests per second [27–31]. In 2022, Microsoft Corporation reported that it was the target of a DDoS attack that, in one minute, reached the level of 3.47 terabits per second [32]; therefore, every second is crucial in combating DDoS attacks. Hence, the authors in [21,33] show that the best way to prevent damages caused by DDoS attacks is to prevent the attack from being launched.

Again, there is a class research works in cyber security searching to predict upcoming attacks to provide more time for service administrators to deal with them and, consequently, avoid losses. Prediction means identifying evidence of an attack preparation before it effectively starts [34], while detecting DDoS attacks requires effectively running the attack, consuming the victim's resources. For example, in [35], the authors predict a DDoS attack two hours before the attacker launches the attack, while the DDoS attack detection literature detects the attack a few seconds after the attacker launches the attack. Although predicting attacks is not a trivial task, the literature presents studies demonstrating the feasibility of proposing solutions for the prediction of cyber threats in a generalized manner [36]. Given the potential to cause losses and the non-triviality of DDoS attack prediction, this survey aims to present a comprehensive and up-to-date classification of the studies that address this subject. This classification simplifies the understanding of the state-of-the-art in DDoS attack prediction because it shows the main features of each solution. In addition, this classification encourages the debate on DDoS attack prediction and promotes the evolution of DDoS attack prediction solutions. Finally, this classification fills in the lack of studies focused on organizing literature on DDoS attack prediction.

The contributions of this survey are three-fold: (1) The classification of works from the literature based on the main criteria employed to predict DDoS attacks, (2) The identification and overview of the studies handling DDoS attack prediction, and (3) An in-depth analysis of open issues that help to evolve solutions. The classification of the existing works on DDoS attack prediction presents relevant aspects that should be considered by researchers when planning new solutions for DDoS attack prediction. Time, architecture, methodology, and data type are common aspects employed by the solutions for DDoS attack prediction. This survey highlights the state-of-the-art of DDoS attack prediction and presents a classification. Out of the 2,482 analyzed studies, it shortlists 27 studies directly related to the prediction of DDoS attacks. This survey details all 27 studies. Further, it presents an expressive analysis of open issues, highlighting research opportunities.

This survey article proceeds as follows. Section 2 presents concepts related to DDoS attacks and defense mechanisms. Section 3 presents the classification proposed in this survey article and details the existing studies on DDoS attack prediction. Section 4 highlights open issues and opportunities for future research. Finally, Section 5 concludes the survey.

2. Background

Confidentiality, integrity, and availability are three key principles of computer security [37]. Attack and defense mechanisms aim to either damage or protect these principles in digital systems. Denial of Service (DoS) attack is an event that disables or restricts the correct functioning of a network or service [38–40], making them inaccessible to legitimate users [15]. These events can be related to attackers exploiting network weaknesses to subvert, obstruct or suppress the network [39]. Moreover, these attacks usually achieve their goal by flooding the service infrastructure with an intense flow of packets [41]. In addition to the number of packets, these events can be software or hardware failures, insufficient resources, environmental conditions, or any combination of these factors [39].

In the origins of the DoS attack, it was common for attackers to attack from a single point. This action simplified the mitigation of the attack because it was relatively easy to identify and block the source of the attack [41]. For effectiveness, attackers employ DoS attack variations, such as the DDoS attack. The first DDoS attack tool was introduced in 1998 [42,43]. Since then, DDoS attacks have evolved in frequency, volume, and sophistication [44–46]. A DDoS attack aims to challenge or obstruct legitimate access to services. Then, attackers exploit infrastructure weaknesses or consume all resources utilizing multiple compromised agents [47]. This section introduces the operating mode of DDoS attacks, attack types, and the protection mechanisms against DDoS attacks.

2.1. DDoS attacks

The popularization of the high-capacity Internet and the increase of devices connected to the worldwide web-enabled the evolution of the DoS attack. Attackers started sequestering machines connected to the Internet, making it possible to circumvent the weakness of the single launch point of the attack. Therefore, the DDoS attack combines several connected devices to attack a target [15,41,48,49]. In the classical model, the DDoS attacks exhaust the computing resources of the victim's infrastructure by creating multiple connections from different sources [50]. Link Flooding Attack (LFA) is another way to degrade or disrupt the victim's service, congesting critical links to isolate the victim's network from the Internet [51,52].

Another way to cause damage to the victim is pursuing the Economic Denial of Sustainability (EDoS). In EDoS, the goal is to consume the victim's resources, forcing the victim to allocate more computing resources. This action increases the cost necessary to keep the service running [53–55]. Cloud-internal Denial of Service (CIDoS) consumes server resources utilizing multiple virtual machines hosted on the victim's physical host. The attacker increases the workload of virtual machines to consume more resources from the host and consequently stops the service [56,57]. Ransom DDoS (RDDoS) is another way to damage utilizing variations of the DDoS attack. In this case, the attackers request the payment of ransoms to suspend or not launch DDoS attacks against the victim [58,59].

The standard parties of a DDoS attack are the attackers, the infected devices, and the victim. A zombie, web robot, or bot is a malware-infected device connected to the Internet that executes programmed tasks [60]. Bots perform actions such as sending bulk electronic correspondence (spam), sniffing traffic, capturing sensitive information, phishing, click fraud, keylogging, disseminating software for cryptocurrency mining, and launching DDoS attacks [61–63]. A "robot network" or botnet is a group of several bots remotely controlled by attackers or botmasters [21,64]. A victim is a server or computer network that holds some resources for the correct operation of a service [65]. A botmaster sends commands to the botnet to initiate connections to the victim to conduct a DDoS attack [41]. The duration of DDoS attacks varies from minutes to days [7], reaches terabits per second [27–29,32] or surpass millions of requests-per-second [30,31].

Fig. 1 illustrates the operation of a DDoS attack, where a botmaster manages bots through control traffic, making the various bots send attack traffic to the victim's infrastructure. Botmaster exploits various types of weaknesses of different Internet-connected devices to spread its malicious code [46]. The targets can be devices with more resources, such as desktop computers, notebooks, servers, tablets, and smartphones [66], or devices with limited resources, such as devices that make up the Internet of Things (IoT) [66,67], such as security cameras or smart TVs. After infection, attackers send commands for their bots to execute. A DDoS attack occurs when the attacker instructs bots to create connections to the victim's infrastructure to consume all available resources.

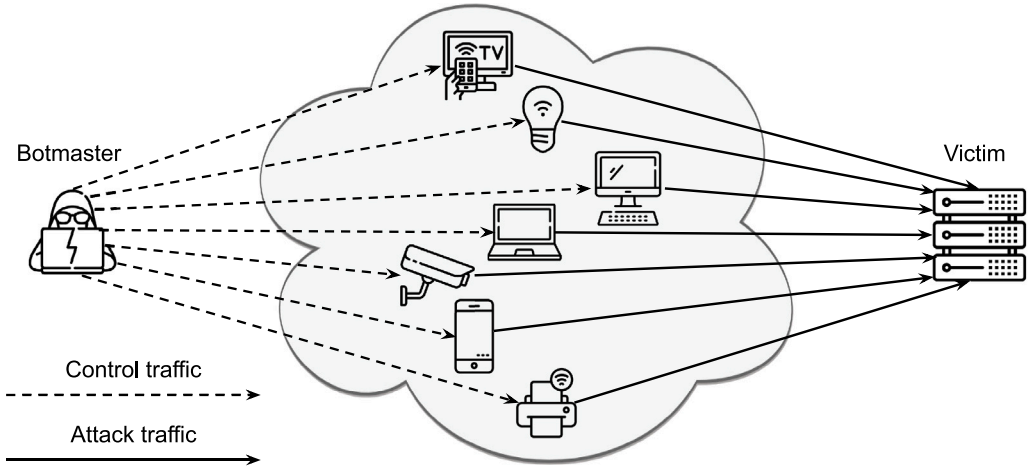


Fig. 1. Basic structure of the DDoS attack (based on [46]).

The literature presents three architectures for attackers to control botnets, centralized, peer-to-peer (P2P), and hybrid architecture [26,68,69]. In a centralized architecture, the attackers communicate with the entire botnet through a central point comprehended as a command and control (C&C) server. A limitation of this architecture is that the central communication point needs high bandwidth to manage communication with the botnet. Despite facilitating botnet management, the central point is an architectural failure point. If the attackers lose access to the C&C server, they lose control over the botnet [26,70]. In order to avoid a single point of failure, attackers can utilize P2P architecture to operate as the C&C model. In this architecture, bots act as clients and servers. The bot receives commands (client) and provides the commands to other bots (server). If a group of bots leaves the botnet, the botnet can continue operating with the rest of the bots. This architecture is more complex for the attackers to manage, and the command dissemination time can be longer [26,70]. The hybrid architecture aims to reduce the management complexity of the P2P architecture. In the hybrid architecture, only one group of bots acts as server and client. Therefore, all bots search for these servers to update themselves [26,71].

It is necessary to utilize some pre-existing protocol developed for another purpose or an exclusive protocol to implement the botnet architectures [72]. Internet Relay Chat (IRC) is one of the main protocols for communicating with botnets. The IRC protocol operates in Internet chat systems. However, attackers began to utilize it for its easy-to-implement structure, as it is widely utilized on the Internet and allows conversation between various entities. The disadvantages of this protocol are that corporate networks rarely utilize this protocol, and firewalls easily block the traffic. Another protocol commonly utilized on the Internet and by attackers is the Hypertext Transfer Protocol (HTTP). HTTP operates in the client–server model where the client creates a request, and the server responds to it [41]. As with IRC communications, HTTP communications are simple to implement. However, they have higher latency when compared to IRC and do not allow communication between groups, as the client must initiate the request, in this case, the bot. Furthermore, attackers utilize the Server Message Block (SMB) protocol to communicate between botnets. SMB is a protocol designed to share resources such as printers, files, and serial ports between computers [73]. Attackers utilize this protocol for communication over local networks as this protocol can be blocked directly at the Internet gateway.

Several attackers utilize protocols based on P2P architecture to maintain communication with their bots. P2P communication is an alternative to centralized servers that reduces download time. Devices present in the P2P network share different text, audio, or video files. Therefore the file is available on multiple servers. When the client needs to download a file, it gets pieces from the closest servers and not necessarily the entire file from the same place. When the client finishes receiving the file parts, it becomes a file server, increasing the number of servers [41]. Attackers utilize this communication method to distribute their commands among bots. Therefore, bots distribute the attackers' commands to other bots as soon as they download the commands. Botmasters can create protocols exclusively for communication with bots. Attackers often implement these new protocols based on the Transmission Control Protocol (TCP) or User Datagram Protocol (UDP). However, there are cases where attackers utilize the Internet Control Message Protocol (ICMP) for C&C communication. The aim of proposing exclusive protocols for botnets is to make detection difficult. However, if they reveal rarely utilized patterns on the network, the security systems easily detect the communication [72].

DDoS attacks have at least four steps: reconnaissance, recruitment, command and control, and launch of the attack [74,75]. Reconnaissance involves collecting information about the victim and the bot candidates to form the botnet. Identifying Internet-connected devices helps the attacker define which devices could act as C&C servers or clients that carry out the attack. In addition, the information helps attackers to create strategies for recruiting these devices. In order to recruit bots, attackers exploit vulnerabilities to infect devices or create campaigns to spread malicious software in email attachments or web downloads [76]. The attack evolves into the control step upon reaching thousands of controlled devices. Attackers can conduct timely maintenance on botnet code to update the bots' code and synchronize them according to the attackers' intentions. Also, in the control step, attackers perform tests lasting seconds or minutes to assess the effectiveness of the attack and make any necessary corrections. The last step is to launch the attack where all active bots in the botnet start sending malicious traffic to the victim. Attackers can avoid the preparation process of a DDoS attack if they utilize an online DDoS attack service, such as [12,77]. In these cases, other attackers have already performed the initial steps of the attack and already have their botnets ready to launch the DDoS attack for a fee [78].

2.2. DDoS attack types

The literature presents three types of DDoS attacks: volume-based attack, protocol-based attack, and application layer-based attack. Volume-based attacks overload the victim's network bandwidth by utilizing a large volume of malicious data. Protocol-based attacks exploit vulnerabilities in network and transport layers of the TCP/IP model reference to overload the victim's computing resources. The application layer-based attack denies web application service reaching the fifth layer of the TCP/IP protocol stack [26,79–83].

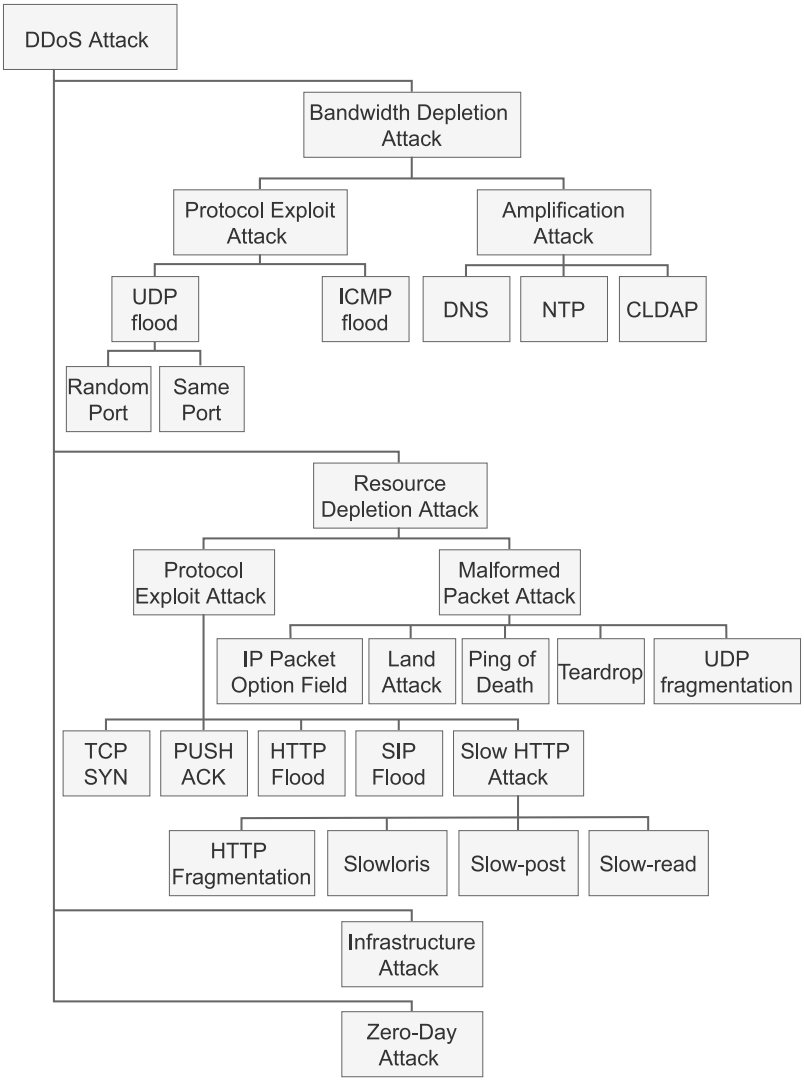


Fig. 2. Classification of DDoS attack mechanisms (based on [14,65]).

Due to the number of variations in DDoS attacks, the literature has different classifications according to distinct points of view [33,50,84–86]. The degree of attack automation, the vulnerability exploited by the attack, and the attack rate are the criteria for classifying existing DDoS attacks [26]. This survey presents DDoS attacks utilizing a classification based on attack impacts similar to the study in [14,65,87]. In [14,65], the authors have reviewed DDoS attacks in four categories: bandwidth depletion, resource depletion, infrastructure attack, and zero-day attack. This survey chose this classification because in [14], the authors proposed it to ensure state-of-the-art is covered thoroughly. In addition, this survey considers that presenting the classification utilizing attack impact facilitates the understanding of specialists in different areas, not only specialists in DDoS attacks. Finally, the literature can specialize this classification for different networks, such as IoT [65]. Fig. 2 presents the classification of DDoS attacks, and this survey describes the categories in the classification below.

2.2.1. *Bandwidth depletion attack*

A bandwidth depletion attack consumes the victim's network bandwidth by utilizing compromised bots to deny access to services. This type of attack can run for long periods before being mitigated. The literature classifies bandwidth depletion attacks as protocol exploit attacks and amplification attacks [14,43,65]. The main characteristic of the **protocol exploited attacks** is to utilize protocols in different network layers to lead the victim to bandwidth starvation. There are two sub-types of attacks in this category: UDP and ICMP flood [14,65]. Below is the description of these attacks:

- **UDP flood-based attacks** are common, making up 43% of the attacks in the first quarter of 2021 [7]. In this type, the attacker directly or indirectly requests that the botnet flood the network bandwidth of the victim with a large stream of UDP packets with the spoofed source address. This attack has two variants where the attacker decides whether the UDP packet destination port will be the *same* for the entire botnet or the port will be *random*. The attack works because the server receives a stream of UDP packets and needs to check if there is any program on the specified port capable of responding to the request. If there is no program to solve the request, the server responds to the request, warning that the destination address is not reachable. The server starts to crash as it receives, processes, and responds to many UDP packets [14,65,88].
- **ICMP flood-based attacks** consume network resources when sending and receiving ICMP requests. ICMP is a protocol to report errors to clients and is utilized by tools such as ping and traceroute [41]. The attacker exploits the network bandwidth of the victim by sending excessive ICMP requests. The server processes these requests and returns the ICMP response to the origin, taking

up the bandwidth [14,65,89]. A simple way to avoid this attack is to disable or limit the receipt of ICMP packets from other networks. The side effect is that the server will not respond to ping or traceroute requests [89].

Amplification attacks aim to deny the service by directing large responses to the victim from small requests. There are three types of amplification attacks: DNS amplification, NTP amplification, and CLDAP amplification [14,65]. Below is the description of the amplification attacks:

- The **DNS amplification attack** utilizes DNS resolvers to increase the attack load directed to the victim and exhaust the network bandwidth. In this attack, the attacker utilizes the botnet to query open DNS resolvers. In queries, the attacker enters the source address of the victim rather than the address of the bot. Consequently, the responses are not directed to the botnet devices that created the query but to the victim. The attacker configures the request to demand the maximum amount of information available. Therefore, with a small request, the attacker generates substantial traffic towards the victim. If the response originated by the DNS server is valid, then the responses are challenging for defense systems to filter. In order to enable this type of attack, the attacker needs to identify open DNS resolvers [14,65,90].
- The **NTP amplification attack** utilizes Network Time Protocol (NTP) servers to increase the attack load directed towards the victim and exhaust network bandwidth. Machines connected to the Internet utilize the NTP protocol to synchronize their internal clocks. In order to amplify the attack, the attacker utilizes an NTP server function that returns data from the last 600 queries performed on the server. This attack creates a response several times larger than the original request with just one request. Therefore, the attacker utilizes a botnet to create many queries with responses directed to the victim's server. One way to avoid this type of attack is to disable the function that returns data from the last queries, which is the default action for NTP servers with versions 4.2.7 or higher [14,65,91].
- The **CLDAP amplification attack** utilizes the Connection-less Lightweight Directory Access Protocol (CLDAP) protocol to increase the attack load directed at the victim and exhaust the network bandwidth. The CLDAP protocol is an option to the Lightweight Directory Access Protocol (LDAP) and provides the information in the directories. In order to amplify the attack, the attacker performs multiple queries utilizing the victim's address. The query response the victim receives is larger than the query causing a denial of service [65,92].

2.2.2. Resource depletion attack

Another approach to deny service to legitimate users is to consume resources other than network bandwidth. The Central Processing Unit (CPU), memory, or sockets are common examples of resources that attackers consume. Resource depletion attacks can be Protocol Exploit Attacks and Malformed Packet Attacks [14,43,65]. **Protocol Exploit Attacks** utilize various protocols to consume the vital resources for the delivery of the service. There are five protocol exploit attacks: TCP SYN attack, TCP PUSH+ACK attack, HTTP flood attack, SIP Flood attack, and Slow HTTP attack [14,65]. Below is the description of the protocol exploit attacks:

- **TCP SYN** is an attack where the attacker exploits a weakness of the TCP protocol to consume the memory resource and consequently deny service. It is necessary to accomplish the handshake process to initiate a TCP connection. The client, who wants to establish the connection, sends an SYN-type packet to the server. The server responds with an SYN/ACK packet to acknowledge the communication [41]. The server waits for the client to send back an ACK packet to establish communication; at this moment, the attacker acts because the expected ACK packet never reaches the server. In addition, the attacker creates numerous TCP connections with spoofed Internet Protocol (IP) addresses until the server becomes overloaded and unable to establish new connections. Therefore, users who try to utilize the service will not be able to establish connections [14,65,93].
- In **TCP PUSH+ACK attack**, the attacker forces the victim's server to drop packets originated by legitimate users through a flood of TCP PUSH ACK packets. Utilizing the botnet, the attacker sends several packets with the PUSH and ACK flags with the value "1" in the header. This action requires extra processing from the server to check the need to clear memory and forward a response. Hence, the attacker makes the destination server unable to process all TCP requests besides discarding legitimate packets [14,65].
- The **HTTP flood attack** exhausts the victim's computing resources by flooding the server with too many HTTP requests. HTTP is an application layer protocol that allows communication between clients and web servers [41]. Attackers utilize two approaches to exhaust a victim's computing resources: HTTP GET and HTTP POST. In an HTTP GET attack, attackers utilize botnets to create many requests for images, videos, or any file on the target server. This action makes the server need to find the requested file, load it into memory, and split it to deliver to the requesters. However, the server runs out of processing resources or memory to serve all the requests created by the attacker. In an HTTP POST attack, the attacker attaches parameters to the POST request to cause a significant processing load with heavy operations on the victim's server. Eventually, resources vanish, and the attack denies service to legitimate users [14,65,94].
- The **SIP flood attack** exhausts the computational resources of the victim by flooding the server with many requests of Session Initiation Protocol (SIP). Voice-over IP (VOIP) applications utilize the SIP protocol [41]. The attacker sends many SIP messages utilizing a botnet, such as SIP NOTIFY or SIP INVITE. The SIP server receives these messages and tries to resolve the requests of bots. However, the SIP server consumes all its resources trying to resolve all requests; this causes the real users to have delays or be unable to utilize the SIP server [14,65].
- The **Slow HTTP attack** slowly consumes the computing resources of the victim. In general, the attacker searches for ways to keep client–server connections open as long as possible. This attack does not require abundant resources to launch and produces traffic similar to the real one. Consequently, this mechanism of attack is difficult to identify. There are four approaches to conducting this attack: *HTTP fragmentation*, *slowloris*, *slow-post*, and *slow-read*. The attacker makes the botnet split HTTP packets into small fragments in HTTP fragmentation. By sending the fragments as slowly as possible to the victim's server, the bot keeps the connection open as long as possible. Slowloris attack opens a connection to the victim's server and slowly sends partial HTTP headers; this makes the server wait till the end of the message, consuming a connection and denying the service for legitimate users. Slow-post attack, also known as a RUDY attack, keeps the server waiting for several HTTP POST requests that the attacker sends to the server slowly. The server keeps these connections open for a long time and causes service interruption. Finally, the purpose of slow-read is to read server responses as slowly as possible, forcing the server to keep connections open as long as possible [14,65,95].

Malformed packet attacks aim to interrupt the victim's operation utilizing malformed packets. They can be one of five types IP Packet option field attack, Land attack, Ping of Death, Teardrop attack, and UDP fragmentation attack [14,65], as described below:

- The **IP Packet option field attack** consumes server resources by increasing the processing load required to analyze each packet. In order to do this, attackers modify the optional fields of the IP packet by entering additional information. Botnets send these altered packets to the victim to cause system interruption [14,65].
- The **Land attack** consumes the victim's server resources, causing loops between requests and responses. In order to do this, attackers modify the IP packet header, then the source and destination become the IP address of the victim. The victim responds to itself and eventually consumes the available resources [14,65].
- The **Ping of Death** (PoD) causes the victim's server to crash by constructing ICMP packets larger than the maximum size. The packet breaks into smaller segments and forces the victim to rebuild it. This rebuilding process causes memory overflow and crashes the victim's server [65,96].
- The **Teardrop attack** causes the victim's server to crash by exploiting TCP/IP packet fragmentation and reassembly. In order to do this, attackers modify the "fragment offset" field of TCP/IP packets to overlap them. The victim's server may crash when it receives these invalid packets and tries to reassemble them [14,65,97].
- The **UDP fragmentation attack** aims to consume the resources of the server. In order to do this, attackers send numerous UDP packets larger than the network accepts. The packet breaks into smaller segments and forces the victim to rebuild it. This rebuilding process is costly for the victim's server, which consumes all available resources [65,96].

2.2.3. Infrastructure attack

The purpose of the infrastructure attack is to deny access to services by consuming all the bandwidth and computing resources of infrastructure critical to the functioning of the Internet. A classic example of this attack is the attack on DNS servers [14]. If DNS cannot resolve a request, the user may not be able to access the intended service. In October 2016, the company DynDNS was the target of one of the biggest infrastructure attacks. In this case, thousands of IoT devices flooded the servers of the company to deny access to important services such as GitHub, Twitter, and Netflix [14,98].

2.2.4. Zero-day attack

Zero-day DDoS attacks are attacks with uncatalogued vectors causing unprecedented attacks. In this attack type, attackers exploit vulnerabilities or security breaches not yet utilized to conduct the DDoS attack. In addition to the utilized vectors, the impact of the attack is also unknown. The attack takes this name because only after the attack on day zero is it possible to recognize the attack vector and propose the appropriate defense or response [14,83,99,100]. Keeping systems up-to-date and correctly configured can be a way to avoid unknown vulnerabilities [58], decreasing the chances of zero-day attacks. In addition, companies offer rewards for researchers to report vulnerabilities and security breaches [101–103]. Experts have identified a zero-day vulnerability in the phone system sold by Mitel MiCollab. Utilizing this vulnerability, attackers amplify DDoS attacks by up to 220 billion percent [104].

2.3. DDoS defense mechanisms

Defense mechanisms against DDoS attacks define actions to prevent or reduce the damage caused by attacks. The literature presents approaches to classify defense mechanisms based on criteria related to how, where, and when it is applied. Examples of classification criteria are activity level, attack response strategy, cooperation degree, deployment location, and time of the mechanism operation [33,84,105]. This survey presents a classification of DDoS defense mechanisms based essentially on when the defense mechanism operates, as before or during the attack [25,33,99,106–108]. This survey follows this criteria because it is the central reference to define DDoS attack prediction, which is the main focus of this survey article [26,109]. In [25], the authors have classified DDoS defense mechanisms into three categories: prevention, detection, and mitigation (response). Currently, environments such as the Cloud [108], IoT [65], and Software-defined networking (SDN) [83] have specialized defense mechanisms to support specific characteristics of each environment. In order to be generic and present an overview of the defense mechanisms against DDoS attacks, Fig. 3 presents the proposed classification at [25], updated based on [26,34,106–108]. This survey describes the categories in the classification below.

2.3.1. Attack prevention

The ultimate goal of combating a DDoS attack is to prevent it [25,33]. Handling DDoS attack after its beginning is not trivial [26]. Actions such as disabling unnecessary services, keeping software protocols and firewalls updated and correctly configured, solving bugs, and replicating services in different places are general prevention mechanisms [24,25]. Prevention must occur before the attack is launched to avoid or narrow the attack negative effects [26,106,108]. In [26,106,108], the authors present challenge response, hidden servers/ports, restrictive access, and resource limits as additional approaches to prevent DDoS attacks. Below is the description of the prevention mechanisms:

- The **challenge response** mechanism aims to distinguish legitimate users from bots to prevent attacks. In order to prove legitimacy, users must solve a challenge in the form of a Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). Currently, there are several types of CAPTCHAs with many different complexity levels. Text-based, image-based, video-based, and audio-based are the existing types of CAPTCHA [110]. Balancing security with simplicity is a challenge for this approach, whereas a disadvantage is excessive resource consumption [26,106,108]. The result of a CAPTCHA test can serve as input for other prevention mechanisms, such as firewalls that benefit from those results to improve the access control rules and block network traffic from bots when those fail to resolve the challenge.
- The purpose of **hidden servers/ports** is to hide servers and avoid direct contact between clients and servers. Intermediary devices, also known as a proxy, are deployed between the clients and servers. The proxy is responsible for managing the connection between the clients and servers and monitoring or forwarding traffic between servers. Hidden ports, moving targets, and intermediary servers are examples of approaches utilized to hide servers. Although this mechanism hides possible weaknesses of servers, proxies can be overwhelmed while redirecting traffic [26,106,108].
- The **restrictive access** mechanism aims to manage access by prioritizing potentially trusted clients. The server delays the response to prioritize clients. Therefore resources become available to the clients that have the best reputation score. There are several strategies to define the reputation of users. One strategy to calculate reputation is to utilize access history. Another strategy is to analyze the time to solve cryptographic puzzles. Even though this is an approach aimed at regulating server resources, it is necessary to deal with some disadvantages of this mechanism. The cost of keeping connections delayed, issues related to scalability, and increased processing time to analyze puzzle answers are examples of problems with this technique [26,106,108].
- The purpose of the **resource limits** mechanism is to avoid huge costs in server accounts caused by DDoS attacks. In this approach, service administrators impose server replication limits and restrictions on available resources for each service. Therefore, a DDoS attack against the network will activate a maximum number of servers, ensuring the account does not exceed the pre-defined

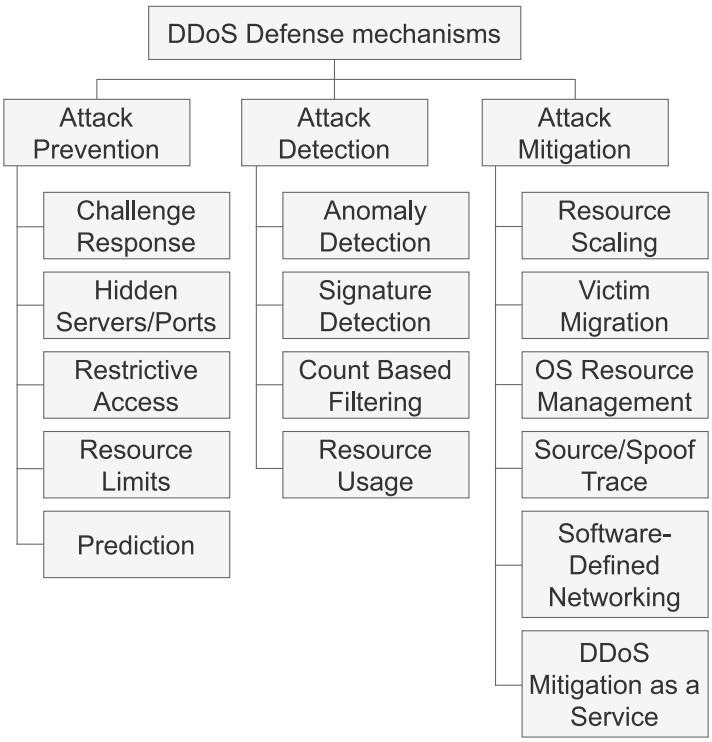


Fig. 3. Classification of DDoS defense mechanisms (based on [25,26,34,106–108]).

limits. This mechanism only protects the server against huge bills because whether the DDoS attack consumes all network/service resources, it will deny the service to legitimate users [26,108].

- The literature introduces **prediction** as a mechanism to assist attack detection and mitigation because prediction occurs before the launch of the attack [34]. Prediction techniques output evidence of upcoming attacks; the evidence can be an alert message, the probability of an attack occurring, or any artifact aiming at network administrators that could represent a DDoS attack that the attacker does not launch yet. Fig. 4 highlights the difference between prediction and detection. The prediction appears before the launch of the attack regarding the attack chronology, while the detection mechanisms need the attack to be running to recognize it. Section 2.1 defines the beginning of the attack as its fourth step. The main advantage of prediction is providing more time for the network administrators to prepare for an attack, decreasing the losses caused by the attack. That is because the damage can be irreversible once the attack is underway. Prediction techniques can operate with little information volume about future incidents compared to detection and mitigation solutions. In other words, prediction solutions must produce evidence about future attacks by processing information where the signals do not determine an attack. Detection techniques depend on signals of depletion caused by the attack to detect an attack. Despite the difficulty in identifying these signals, predictions will be more accurate as more signals can be identified [34]. Section 3 presents the classification proposed in this survey and details the solutions contributing to the prediction mechanism.

2.3.2. Attack detection

Once the prevention procedures are determined, the next step is to define attack detection mechanisms. Attack detection occurs after the attacker launches the attack and can happen at the first overload signals or when the DDoS attack is causing damage. In order to detect DDoS attacks, solutions may utilize metrics such as processing and memory consumption, response times, and information related to server performance. Detection can occur at the victim's infrastructure, intermediaries network, or attack source. The most viable place for attack detection is in the victim's infrastructure because it is the victim who receives all the malicious traffic. Further, detection near the source of the attack reduces the amount of damage suffered by the victim and intermediaries network [25,33,106,108]. The existing detection mechanisms are anomaly detection, signature detection, count-based filtering, and resource usage [26,106,108]. Below is the description of the detection mechanisms:

- The **anomaly detection** mechanisms search for unusual patterns which deviate from the standard to detect DDoS attacks. The solutions track and analyze packets, analyze access and Internet connections logs, and compare this information to usual predefined standards. A DDoS attack may be in progress if the mechanism identifies an anomaly. Machine learning-based solutions have been gaining prominence in the literature. However, there are solutions based on data mining, statistical methods, soft computing, and data stream algorithms [111]. The biggest issue with anomaly-based mechanisms is the high rate of false positives, as an anomaly does not always represent an ongoing attack [26,106–108].
- The **signature detection** mechanisms detect attacks with known patterns. Solutions utilizing this mechanism compare current data to past attack signatures to detect attacks. Despite being effective against known attacks, keeping the attack base up to date is not trivial. Another limitation of solutions that implement signature detection is that solutions are unlikely to detect attacks with patterns not yet registered [107].
- The **count-based filtering** mechanism aims to filter malicious traffic by utilizing network resources. The solutions that implement this mechanism count the number of connections, the number of requests made by each origin, or the number of hops. The solution marks the packet as malicious whenever the result exceeds a predetermined threshold. This mechanism is simple to implement and provides a quick response to the attack. However, it is necessary to monitor the success rate because the solutions can misclassify many packets [26,108].

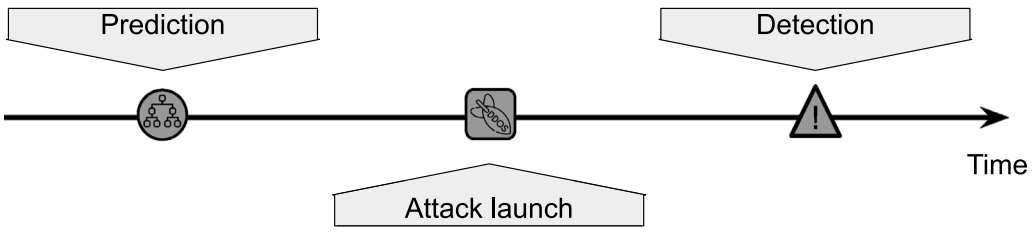


Fig. 4. Relation between prediction, detection, and attack start.

- In the **resource usage** mechanism, solutions monitor server resources frequently, and if the server starts to utilize more resources than usual, a DDoS attack may be running. As several attacks aim to exhaust server resources, metrics such as processing and memory consumption help detect DDoS attacks. The main challenge of the solutions based on this mechanism is to correctly identify when resources are utilized by legitimate users or by bots as a result of a DDoS attack [106–108]. Another difficulty with these mechanisms is that when resource consumption is relevant to detecting a DDoS attack, there may not be enough time to take precautions against the attack.

2.3.3. Attack mitigation

After identifying the attack, it is necessary to take the appropriate actions to respond and mitigate it. Mitigation consists of actions to avoid or minimize the damage experienced by the service. Furthermore, these mechanisms define actions to recover the service at the end of the attack. This survey presents mitigation mechanisms available to the victim. Likewise, it points out the need for a collective effort of other entities in the network to protect their assets and consequently increase global security against DDoS attacks. Once an attack has occurred, prevention and detection methods can be updated utilizing the lessons learned during attack mitigation [25,33,106,108]. In these papers, the authors present six mechanisms to mitigate DDoS attacks, which are as follows: resource scaling, victim migration, OS resource management, software-defined networking, DDoS mitigation as a service, and source/spoof trace. Following is the description of the mitigation mechanisms:

- **Resource scaling** defines the increase of resources following the progression of the attack. In this case, auto-scaling solutions allocate more resources to existing virtual machines or install new virtual machines on the same or different server(s) to process all requests while minimizing damage. Resource scaling can save the service from a DDoS attack, but it increases the overhead due to utilizing extra resources. This overhead to keep the service available is the EDoS attack's goal [26,106–108].
- **Victim migration** moves the attacked service to a different physical server isolated from the attack. Consequently, the service is available to legitimate users, and the attack continues to consume the old server. The mechanism depends on the possibility of having a server available to the clients but not under attack to be effective. There are cases where attackers prefer this change because this increases the costs related to the allocation of a new server [26,106–108].
- **Operating systems (OS) resource management** restricts the consumption of resources by the attacked service. Therefore, this mechanism aims to stop the excessive consumption of resources to avoid degrading the performance of other mitigation methods. The main limitation of this mechanism is that the service can quickly become overloaded, denying the service to potential users [108].
- The **source/spoof trace** mechanism aims to identify the request address. The expected result is not the attacker's address but the address of the devices utilized by the attackers to conduct the attack [112]. Mitigation mechanisms utilize the actual origin of the request to contain the attack [113]. ICMP traceback, probabilistic packet marking, and hash-based IP traceback are examples of source/spoof trace mechanisms [112]. For this mechanism to operate correctly, cooperation between different network entities is necessary, limiting the adoption of this mechanism [26,108].
- **Software-defined networking** provides some network reconfiguration opportunities to mitigate DDoS attacks. These reconfiguration opportunities are due to the physical decoupling of the routing and control functions of the network, helping the management of services [41,114]. In addition to the reconfiguration power of SDN, it simplifies the use of cybersecurity techniques. Despite its benefits, SDN is not immune to DDoS attacks and requires specialized defense mechanisms [26,106,108].
- **DDoS mitigation as a service** requires network managers to recruit specialized solutions to mitigate attacks. One of the ways to perform mitigation is to add an intermediary device that filters and directs packets to the server. Another way is to utilize cloud-based solutions to mitigate attacks. Utilizing remote approaches can cause privacy issues if user data becomes public. Another limitation of remote mitigation approaches is that mitigation actions can delay attack containment [26,107,108].

2.4. Literature review methodology

This subsection describes the search string, research sources, and inclusion and exclusion criteria utilized in the literature review. The search string is vital for the success of this work since it must represent the terms used to identify the existing solutions, reaching all studies that constitute the literature related to DDoS attack prediction. This work has utilized the following search string: distributed denial of service OR DDoS AND predict OR early OR forecast OR forecasts OR predicting OR prediction AND solution OR approach OR approaches OR framework OR frameworks OR method OR methodologies OR methodology OR methods OR procedure OR procedures OR solutions OR system OR systems OR technologies OR technology OR tool OR tools.

The research sources also impact the final result of this work. This work uses five research sources: ACM Digital Library, IEEE Xplore, ScienceDirect – Elsevier, Scopus, and SpringerLink. IEEE Xplore, SpringerLink, and ScienceDirect are research sources that only index the studies published by the platforms. Scopus is a research source that operates only with a search engine, indexing studies published on other platforms. Furthermore, the ACM Digital Library operates in a hybrid way, indexing its studies and studies published on other platforms. The objective of choosing these bases is to identify the totality of studies relevant to the adequate composition of this work.

The selection criteria determine the acceptance or rejection of the results found during the searches. There are three inclusion criteria and nine exclusion criteria. The first two inclusion criteria define the study type and the period of activity included in this work. Thus, to be included in this work, the search result must be a study proposing a solution for DDoS attack prediction and must be published by May 2022. The third criterion defines the acceptance of studies that propose

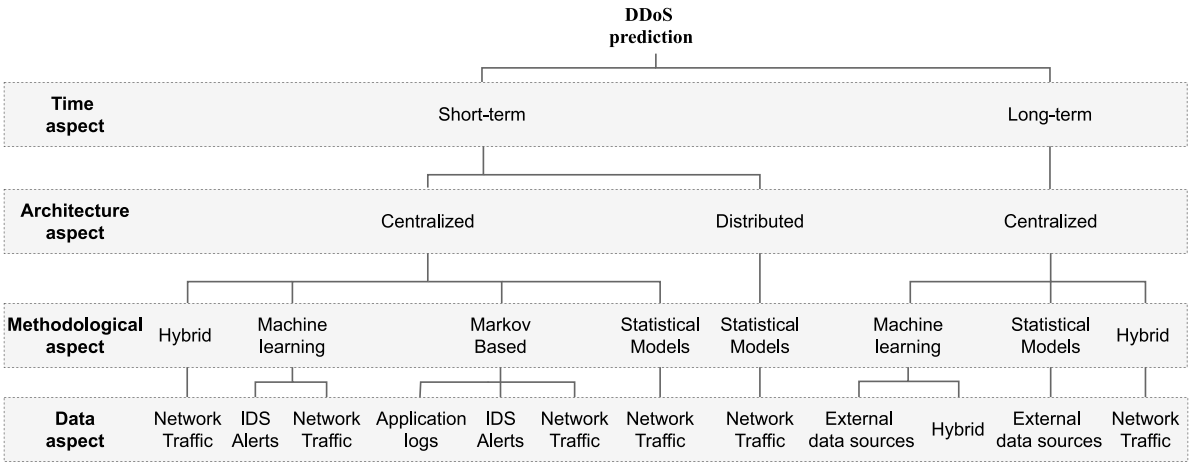


Fig. 5. DDoS attack prediction classification.

solutions to the DDoS attack prediction problem using prediction or forecasting techniques.

The exclusion criteria complement the inclusion criteria and add some specific restrictions. Results were removed from the analysis if they were slides from a presentation, studies written in any language other than English, studies that did not present results for the proposal, studies without an abstract, or only with the available abstract. Some research sources can index the same study, thus causing duplication of results, so only one of the occurrences was considered. If two or more results correspond to older versions of newer studies, this survey has discarded the oldest versions. This survey has removed the studies that full text is unavailable for access. Finally, the last exclusion criterion removes results that do not address DDoS attack prediction.

3. DDoS attack prediction

This section presents a classification of existing studies that propose solutions to DDoS attack prediction. It is worth mentioning that the classification proposed in this survey only considers the studies related to DDoS attack prediction. This is the main contribution and difference between this work and other related surveys from the literature. Therefore, this survey does not focus on showing DDoS attack detection or mitigation solutions. For surveys addressing specifically DDoS attack detection or mitigation, please, refer to [115–119]. This classification helps researchers identify how the state-of-the-art addresses DDoS attack prediction [69] and unveils the open issues [120]. This survey identifies the common aspects of the studies, such as the time before the attack begins, solution architecture, solution methodology, and the type of data the solution utilizes to predict attacks, which leads to a classification of studies based on time, architecture, methodology, and data aspects. This classification in four aspects provides different points of view on the identified studies, reflecting the main aspects of the evolution of research in DDoS attack prediction.

3.1. DDoS prediction aspects

Fig. 5 presents the classification of the studies utilizing the aspects mentioned above, presenting them in layers. The first layer classifies the studies concerning the time aspect. The second layer utilizes centralized versus distributed architectures as the classification criteria. The studies classified under the category of short-term follow both approaches, whereas those in the long-term category, until now, only follow a centralized approach. In the third layer, the studies classified under the short-term and centralized category follow four possible methodological aspects: Machine Learning/Deep Learning, Markov-based, purely Statistical Model, and Hybrid. The only study classified as short-term and distributed category utilizes Statistical Models as a methodological aspect. The studies classified under the long-term and centralized architecture category utilize Machine Learning/Deep Learning, Statistical Models, or a hybrid approach based on Machine Learning and Statistical Models. Finally, the studies utilize network traffic, Intrusion Detection System (IDS) alerts, application logs, and data collected in external data sources in the data aspect layer.

The **time aspect** is associated with the time between predicting and launching the attack. The literature commonly employs short- and long-term terms, referring to the mentioned time. However, this survey could not spot a precise short and long-term definition concerning the prediction in the studies. In order to illustrate how literature employs these terms in [121], the word short-term refers to actions to be carried out within six months, medium-term incorporates actions to occur in the next six or 18 months, and long-term includes the actions to occur beyond 18 months. Since DDoS attackers utilize various methods to avoid prediction and detection [122,123], it is important to set a suitable standard for these terms.

This survey has identified two behaviors regarding attack prediction in the studies to define short and long-term terms. Firstly, this survey has identified a group of studies performing the prediction close to the moment the attack is launched, with a maximum of a few hours before the attacker launches it. In this case, the prediction occurs less than a day in advance. Concerning the second group, this survey has identified that prediction occurs one or more days in advance. As a conclusion of this analysis, this survey employs the 24-hour frontier to divide the groups in this work.

Fig. 6 illustrates the two behaviors identified concerning the time aspect. In the figure, the arrow represents a timeline where the predictions and attacks happen. The filled circles before the 24-h frontier represent the six studies that predicted an attack more than a day in advance. These studies range from one day to three months in advance. The filled diamonds represent the 20 studies that predicted an attack less than 24 h in advance. These studies range the creation of evidence from 14 h to seconds close to the start of the attack. Therefore, if solutions that perform attack prediction up to 23 h, 59 min, and 59 s before the beginning of the attack, this survey considers the solution as a short-term solution. This survey considers the solution as long-term in case the study performs the prediction at least one day before the attacker launches the attack.

The DDoS attack prediction field seeks to prevent network administrators and security teams from being caught unprepared by the attacks, providing extra time to avoid losses [124]. Therefore, solutions with long-term aspects are desirable since they provide more time for security teams. However, only six out of the 27 studies identified in this work have presented characteristics related to the long-term

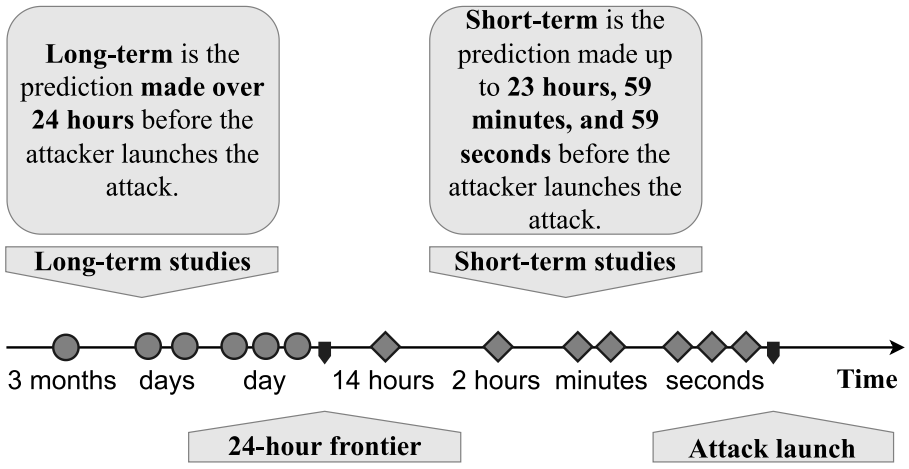


Fig. 6. Relationship between short and long-term.

classification. This fact highlights the difficulty of predicting the attacks multiple days in advance. However, this survey has identified 20 out of 27 studies that predict less than one day in advance. Despite all this, performing DDoS attack prediction is still a non-trivial task, given the low number of studies published. Finally, this survey did not identify the time aspect in one study [125]. However, this survey included these studies due to the potential to predict DDoS attacks that these solutions showed.

The **architectural aspects** refer to how data is processed. The classical approach is to collect the data utilized by solutions and centralize its processing. There are cases in which this action is essential because the union of the data can represent signals of an attack. For example, a single IDS alert indicating unusual communication may not be sufficient to represent preparedness for an attack because this alert may be a false positive. However, several alerts representing unusual communication with different sources can represent the preparation of an attack. Centralizing data processing is an option to ensure that the solution operates with all available data to predict DDoS attacks. This approach's difficulty is that centralizing data processing requires vast computational power, depending on the type, the amount of utilized data, and the method to perform DDoS attack prediction [126]. In addition to centralized architecture, this survey has identified distributed architectures. In distributed architectures, processes at different points in the infrastructure split the problem by processing small data portions. With the processed data, the processes collaborate to identify and extract evidence about DDoS attacks.

The distributed architecture is desirable for DDoS attack prediction solutions because of the complexity of the network infrastructure and the large volume of data caused by some types of attacks. However, proposing and testing distributed solutions to solve the DDoS attack prediction problem is not a trivial task. This survey only identifies one of the 27 studies as distributed [127]; the other 26 studies process information centrally. Despite being more common in the literature, these works must consider the amount of data to be analyzed since the resources may not be enough to complete the tasks.

The **methodological aspects** involve the techniques the studies employed in the solutions. This aspect is essential because the results of methodologies characterize the time and architectural aspects. This survey has identified four techniques: Machine Learning/Deep Learning, Markov-based, purely Statistical Model, and Hybrid. Statistical models were the first approaches followed to predict DDoS attacks. In general, the goal was to utilize techniques such as ARIMA or Gray Theory to predict the behavior and to know in advance if an attack could occur. This type of technique has evolved, and other approaches have appeared in the literature to make it possible to identify signals before the attacker launches the attack. The evolution of computational power and the creation of thousands of gigabytes of data every month [128] supported the utilization of machine learning techniques in several contexts [129], including the prediction of DDoS attacks. Machine learning techniques utilize the available data to build models able to predict the occurrence of a DDoS attack. In addition to properly handling large amounts of data, machine learning techniques have a variety of models that can inspire solutions to predict DDoS attacks. As in [130–132], this survey considers deep learning as a branch of machine learning. Therefore, this survey classifies all solutions that utilize deep learning techniques such as machine learning. Solutions based on Markov chains [133] are also part of the classification presented in this work. Studies utilize the possible states of a service or an attack to build Markov chains and predict DDoS attacks. In addition, Markov chains show the probability of changing states. Suppose the solution can identify signals that an attack is in the phases of attack coordination, such as communication with the C&C center [134]. In that case, the solution presents a probability that this attack evolves to the attack state before the attack even launches, achieving the prediction of DDoS attacks. Finally, hybrid solutions combine two or more techniques previously mentioned to predict DDoS attacks. For example, in the study by [135], the authors combine the predictions performed with ARIMA and neural networks to build a solution for predicting DDoS attacks.

The **data aspects** layer highlights the type of data the solutions employed to make predictions. This survey has identified that the studies utilize IDS alerts, network traffic, application logs, data collected from external data sources, or a hybrid approach. IDS alerts are warnings created by different IDS that some solutions utilize to identify possible actions, such as malware spread attempts and unusual communications. Therefore, solutions utilizing IDS alerts try to identify sets of breaches to predict the attack before the attacker launches it. The network traffic is the most utilized data aspect among the studies analyzed. Studies try to infer characteristics that represent attacks even before the attacker launches the attack. Two studies utilize application logs to build the attack prediction model [136,137]. These logs can be device operations, login attempts, or application states. Some studies utilize data collected from external sources such as social networks, news sites, blogs, forums, and blacklist sites to predict DDoS attack targets and campaigns. Finally, one study enriches the network traffic with data collected externally. In this case, this survey defines this combination as a hybrid approach.

In addition to the time that solutions can predict attacks, some studies complement the evaluation of solutions with accuracy, precision, recall, f1-score, and area under the receiver operating characteristic curve (ROC curve). In order to define them, it is necessary to measure the total number of true positives (TP) and true negatives (TN), the

total false positives (FP) and false negatives (FN), and the total of observations (N) to calculate the metrics. The accuracy evaluates the classifications of the system (Eq. (1)), and it is a standard metric that helps understand the results. However, few signs of attack preparedness exist as attackers hide their actions. Therefore, accuracy above 90% may present a false sense of good results because the system can correctly classify all observations of the majority class and misclassify all observations of the minority class and have accuracy above 90%. Some works utilize precision, recall, and f1-score to complete the analysis of the results. Precision indicates the relationship between the observations labeled by the system for a specific type and how many were of the assumed type (Eq. (2)). The recall presents the relationship between all the expected observations of the specific type and how many observations of this type the system correctly classified (Eq. (3)). The F1-Score (Eq. (4)) is the harmonic mean between precision and recall. Finally, some studies utilize the area under the curve (AUC) to evaluate the results of their proposals. The AUC value equal to 1 means that the correct classify all samples. It is necessary to create the ROC curve to calculate the AUC metric. The ROC curve is based on different threshold values, true positive and false positive rates. Therefore, the AUC condenses the relationship between thresholds, true positive rate, and false positive rate into just one measure.

$$Accuracy = \frac{TP + TN}{N} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = 2. \frac{precision\ .\ recall}{precision + recall} \tag{4}$$

3.2. Short-term prediction

This subsection presents all the studies classified as short-term predictions. This survey has distinguished centralized and distributed as architectural aspects for these studies. Also, this work has identified that studies with centralized aspects utilize machine learning, Markov-based, statistical models, or hybrid as methodological aspects. This work has identified that the study with distributed aspects employs statistical models as a methodological aspect. This survey divides them according to the methodological aspect utilized in each study to simplify the presentation of the studies. However, each study has its complete classification throughout the text.

3.2.1. Hybrid models

In [138], the authors propose a solution to predict DDoS attacks in healthcare environments, despite failing to test the solution with a dataset representing healthcare environments. The proposal utilizes the possibility of predicting DDoS attacks using neural networks and Exponential Smoothing. Consequently, this survey classifies the study of [138] as a hybrid methodology. The authors proposed a solution with three modules: traffic analyzer, time-series predictor, and attack detector. The first action of the solution in the traffic analyzer module is to collect the network traffic in the form of the number of packets per second; the solution stores this data in a time series. The solution subtracts from each item in the time series the value of the item in the past, to remove unusual trends. The solution predicts the subsequent values of the pre-processed time series using Exponential Smoothing. The solution compares the predicted time series results with the actual values. As a result, the solution builds a new time series with the difference between the predicted and the actual values. Therefore, the output of this module is a time series representing the prediction errors. The time series of errors is important because the error values can present anomalies in the network traffic, and the solution utilizes these possible anomalies to predict the attacks. The time-series predictor module utilizes the time series of prediction errors to predict errors in the next few seconds. In this module, the solution employs a Recurrent Neural Echo State Network named SCESN to perform the prediction. Therefore, the output of this module is a time series that contains the prediction of errors for the next second. The attack detector module analyzes the time series of predicted errors and calculates the Lyapunov exponent for the time series. The solution emits an alert predicting a DDoS attack when the value of the Lyapunov exponent is positive. Due to the lack of datasets with DDoS attacks in healthcare environments, the authors evaluated the solution using Defense Advanced Research Projects Agency (DARPA) 1998 [139]. As a result, the solution creates evidence of DDoS attacks up to 50 s before the attack. The solution achieved a precision of 88.97%, a recall of 77.44%, and an F1-score of 82.81% predicting attacks that will occur in the next 20 s.

3.2.2. Machine learning

In [140], the authors propose the analysis of alerts created by IDS for DDoS attack prediction. The goal is to identify groups of alerts that can symbolize the orchestration of a DDoS attack before the attacker launches the attack. The proposal has three phases: pre-processing alerts, constructing the model, and detecting attacks. The solution transforms alerts from different IDSs into standardized objects in the pre-processing phase. The solution collects the alerts' description, priority level, protocol, sensor information, source/destination IP and Port, time, and type. The solution merges and eliminates duplicate or irrelevant alerts to finalize alert pre-processing. During the model building phase, the solution calculates the entropy of the pre-processed data. Entropy measures the level of uniformity in the distribution of each feature. The authors chose K-means [141] as the unsupervised machine learning technique to compose the solution because K-means can find spherical-shaped clusters and quickly converge. The model building phase ends when K-means clusters the IDS alerts based on entropy. In the detection phase, the solution checks whether the clusters defined by K-means are normal or malicious; for this, the solution calculates the average entropy of each cluster. If the average entropy is close to zero, the cluster is malicious. If the average entropy is near the highest values of the base, the cluster is normal. Finally, the authors developed a solution to utilize real-time alerts. Therefore, the solution updates the clusters as new IDS alerts become available. The authors evaluated the solution on a simulation containing 140 h of network traffic and attacks such as port scanning, brute force, and DDoS attacks. The authors evaluated the results using only two features: the message destination port number and the alert type. At the end of the simulation, the solution presents three clusters. The average entropy of the first was close to zero, which indicates the preparation of attacks. However, the authors analyzed the alerts and found that 84.4% of the alerts are not related to attack preparation. Despite the proportion of errors obtained in this cluster, this cluster only has 1.17% of all data. The second cluster contains 54% of all alerts, the most significant cluster. The average entropy of this cluster is zero; consequently, the solution also classifies these alerts as a signal of DDoS attack preparation. The authors identified that 22.55% of the alerts of this cluster are not related to DDoS attacks. The third cluster has high average entropy; therefore, the solution classifies this cluster as normal. Future works aim to improve the solution by decreasing the error rate. Finally, K-means requires that, before performing the clustering, the authors define the expected number of clusters, making it challenging to utilize the solution.

The study in [142] presents a solution to predict when DDoS attacks will consume all the resources of a network's authentication server. The solution utilizes the Gaussian Process Regression (GPR) for DDoS attacks prediction utilizing network traffic collected from sensors. GPR is a supervised machine learning technique to execute classification and regression tasks [143]. The sensors send the number of defective hosts and the number of authentication attempts by time unit to a processing center. The processing center utilizes this data to perform

the GPR training and predict the timing of the victim's overload. The authors centralize the data processing instead of distributing the data processing in sensors due to the low processing power of the sensor. The authors performed simulations varying the training size and attack rate to evaluate the solution. During the simulations, the solution could predict how long it would take for an attack to overload the victim, usually a few seconds before the overload. During simulations, the authors train the solution utilizing data after launching the attack, and the result indicates when the overload will happen. However, this survey considers the study [142] in this work because, after training, the solution can predict when a new attack will overload the server. Therefore, the solution may not predict the first attack, but the solution can predict the subsequent attacks. The presented results indicate that the solution depends on the quality and quantity of the training data; because the solution may not predict when the overload will occur if the authors do not train the solution correctly.

In [144], the authors utilize the Learning from Examples based on Rough Sets (LERS) strategy to build a set of rules capable of predicting attacks and intrusions. Although not a specific solution for DDoS, it is reported that the solution can operate with DDoS attacks. First, the solution processes the information in a centralized way. The data processing consists of dividing the historical data, reducing the number of available features utilizing the Principal component analysis (PCA) [145], and performing a covariance analysis. The solution applies the set of rules created utilizing LERS. LERS is an intelligence acquisition mode based on machine learning [146]. Therefore, by utilizing network traffic before the beginning of the attack, the solution may predict an attack that will occur in the next few seconds. The authors evaluate the solution utilizing the KDD CUP 1999 [147] dataset; they utilize all 41 available features in the dataset.[1] Unfortunately, the authors do not present accuracy, precision, or error rate metrics to describe the DDoS attack prediction.

In [148], the authors utilize multivariate linear regression analysis, a machine learning technique [149], to predict the volume of a DDoS attack that might occur in the next few hours. The solution needs the importance level of servers that can be targets of DDoS attacks and must be protected and the number of possible bots to predict DDoS attacks. The level of importance of servers that must be protected ranges between one and ten. The higher the level, the greater the importance of the service or data provided by the servers. In order to estimate the number of bots, the authors utilize the network traffic to check the number of non-duplicated IP addresses that try to connect via Secure Shell (SSH) on the servers. The authors performed a correlation analysis on Keimyung University network traffic data between December 2010 and August 2011. Then they applied the output of this correlation analysis over the multivariate linear regression analysis to obtain a model to estimate the volume of attacks that may happen during the next few hours. In order to perform the prediction, the authors utilize the model obtained jointly with the number of possible active bots and the value of the servers that must be protected. Despite estimating the values for the volumes of attacks, the authors do not present metrics to show the accuracy of the model.

The work by [150] compares three regression techniques for predicting DDoS attacks. The authors have trained the techniques with normal data and attack data to predict the occurrence of DDoS attacks. The techniques compared are the Logistic Regression (LGR), Support Vector Regression (SVR), and Kernel Ridge Regression (KRR). The authors chose LGR because it is appropriate for solving binary problems, SVR because it is the regression version of Support Vector Machine (SVM) [151], and KRR because it is faster than SVR. To perform the comparison, the authors utilized the network traffic of the DARPA 1999 dataset [152], and in some moments, the authors manually increased the number of packets to introduce DDoS attack traffic. The authors trained the regression techniques using the number of packets per 10 s in 80% of the dataset. They tested the solution with the remaining 20% of the dataset. LGR obtained the best results, achieving an accuracy of 98.60% to predict the occurrence of DDoS attacks in the next 15 min. The LGR and KRR predicted attacks with up to 4 h of lead with a low error rate. However, the LGR and KRR made predictions in cases with no attack. This error is serious because network administrators can spend time and money to stop an attack that will not happen. In addition, the authors consider that the solution is suitable for IoT environments; however, the tests do not support this because the data employed are from 1999.

The study in [153] predicts DDoS attacks utilizing samples of network traffic and RBF neural networks [154], considered a machine learning model. This study proposes one solution to centrally processing data and predicting traffic variation more faithful to reality than other classic techniques such as ARIMA. The proposed solution starts by collecting historical network traffic samples. However, the authors do not specify the features utilized in the historical network traffic samples. After collection, the solution performs data preparation. The preparation consists of a step of denoising the network traffic and the reconstruction of the chaos of network traffic. Local projection and wavelet threshold are standard technologies for traffic denoise reduction. This data preparation is essential because the authors mention that, for network traffic predictions to be accurate, it is necessary to highlight the traffic chaos hidden by the high dimensional noise of network traffic. With the data prepared, RBF neural networks predict network traffic behavior for the future. Prediction of traffic in the next seconds is input to a DDoS attack detection algorithm. In other words, the complete solution does not present evidence to indicate whether the system notifies about the attack before or after the attack starts. However, predicting future traffic can generate evidence of a possible DDoS attack before it even happens. In addition, the proposed solution can become a solution that addresses the DDoS attack prediction requirements if it is possible to change the detection algorithm to utilize the prediction of future traffic and generate evidence before the start of the DDoS attack. The authors evaluated the proposed solution by simulating the network traffic of a DDoS attack. The proposal outperformed ARIMA by predicting the amplitude of network traffic with near-zero error rates.

The study in [125] proposes a solution focused on detecting botnets during the initial stage of C&C communication. The first step conducted by the authors was the selection of features. In order to select features, the authors utilized PCA and Information Gain. At the end of the feature selection process, the solution obtained the 40 most representative features. The five most important features are the proportion of received packets whose size is between 1200–1299 bytes, the proportion of received packets whose size is more than 1400 bytes, the proportion of packets with the RST flag, the proportion of the packets with the FIN flag, and the shortest time interval between received packets. Next, the authors centrally analyze the network traffic containing traces of botnets' actions. The authors utilized the Cyber Clean Center (CCC) dataset, the 40 best features, and compared the accuracy between four machine learning techniques: Random Forest (RF) [155], SVM, Logistic Regression [156], and Multilayer Perceptron (MLP) [157]. The authors evaluated the comparison using accuracy, the True Positive Ratio (TPR), and the False Positive Ratio (FPR). Using 37 out of 40 features, the authors obtained 97.8% accuracy, 0.99 TPR, and 0.007 FPR using RF to detect botnets prematurely. Using the 40 features, RF achieved an accuracy of 99%, 0.97 TPR, and 0.02 FPR. Although the solution does not explicitly predict a DDoS attack, detecting C&C communication is evidence of a possible attack occurring in the future. Consequently, this survey included the study of [125]. It is unclear how long before the attack the solution can identify the C&C communication and consequently create evidence of the future attack. Therefore, this survey did not identify the classification for the time aspect of the study [125]. However, this survey identifies

[1] kdd.ics.uci.edu/databases/kddcup99/kddcup.names

and classifies all other aspects of [125]. Finally, attackers can utilize centralized, P2P or hybrid architectures to build botnets (Section 2.1). The proposed solution focuses on a centralized architecture based on IRC and HTTP. Therefore, the solution may obtain different results in different scenarios.

3.2.3. Markov based

In the study in [136], the authors propose a solution to utilize Markov chains to predict cybersecurity events. The solution presented does not focus on DDoS, but the authors cite the possibility that the solution works in the context of DDoS attacks. The solution performs the modeling of the Markov chain based on application logs generated at the sensor. Each log has the capture time, the source and destination, the log size, and the event type. The event type indicates the action performed by the sensor. The options are reading the report, disconnecting, reconnecting, upgrade. With the logs, the authors define the possible states of the network, and with a Markov chain, the solution predicts the next state using the current state of the network as input. In order to evaluate the solution, the authors collected event logs over two weeks on a network without attacks. The authors utilized the first week for training the solution, and the second was for testing. The authors generated different attacks with different victims and combined the logs emitted by the sensors during the attack with the normal logs. It is worth mentioning that the authors mixed a maximum of 10% of attack logs with benign data. The attack simulations include the attempt to infect the sensors and launch the attacks by these sensors. The authors conducted two tests. In the first test, the authors trained the solution utilizing one week of data. In the second test, the author trained the solution using only one hour of data. The solution predicts the change between states with less than 2% error in both tests. Therefore, if the state variation prediction represents a change to an attack state, the solution can predict attacks seconds before they occur. For the solution to predict attacks, the data utilized in training must represent the attack preparation and launching. Furthermore, correctly configuring the Markov chain for each scenario may be challenging.

In [137], the authors propose a solution for predicting DDoS attacks in IoT networks by analyzing each device's behavior. The solution consists of three models: the collector model, the prediction and detection model, and the reaction model. The authors designed a model to collect all data sent by IoT devices in the access layer. The access layer includes the gateways that receive data from devices, transmit the data to the MAC layer and monitor the devices in real-time. Due to the volume of data generated by IoT devices, the first action of the solution is to clean the data, removing duplicates and keeping only representative samples. The solution pre-processes data from heterogeneous devices into standardized logs in the prediction and detection model. The solution classifies logs into three categories: read, update, and delete activities. The solution processes logs using a mathematical model based on the stochastic Markov process. The solution identifies a malicious state of the device when the solution verifies that the amount of logs exceeds the pre-defined limit. The state is suspect when the solution analyzes device logs, and the number of actions performed is greater than the threshold, but the system tolerates this number of actions. During the suspect state, the solution analyzes the logs and checks the amount of time the device brings to transition between states. If this observed time exceeds the time verified in the histories, the solution alerts network administrators that an attack may happen in the future. Finally, the authentic state of the devices occurs when the amount of logs is compatible with the limits. The authors analyzed historical information such as users' logins, device state, and system operations to define the limits. In the reaction model, the solution blocks the access of some devices to avoid overloading. During the attacks simulated by the authors, the solution predicts the DDoS attacks that would happen in the next few seconds. However, the authors presented the prediction results together with the detection results. Therefore, the attack prediction accuracy is unclear. Finally, it can be challenging to analyze the behavior of IoT devices. Because, in general, IoT networks concentrate many devices and generate extensive volumes of data.

The study in [158] aims to identify and model the typical behavior of botnets in a Markov chain. Therefore, the authors proposed a methodology to predict attacks based on the evolution probability from the current state to an attack state soon. The authors identified that botnets could have eight states: inbound scan, social engineering, drive-by download, exploit, binary download, C&C discovery, C&C communication, and attack. The authors utilized the Exploit, Binary Download, C&C Communication, and Attack states to build the solution because they are the states that the alerts generated by intrusion detection systems such as Snort IDS[2] can represent. The authors evaluated the solution on SysNet Dataset and ISCX Botnet Dataset. SysNet Dataset contains 24 h of network traffic where ten types of botnets carried out attacks. ISCX Botnet Dataset contains three days of network traffic from different botnets. Using Snort IDS to analyze the network traffic of these datasets, the authors collect multiple IDS alerts and utilize them for training the Markov chain solution. The solution analyzes new IDS alerts and predicts command-and-control (C&C) communication with 99% accuracy during the experiments. This result can be evidence to predict future DDoS attacks. The results still show that the prediction of the attack can vary from a few seconds to 14 h before the beginning of the attack. The proposed solution utilizes four of the eight possible botnet states; this limits the generalizability of the solution because it may not have mapped the botnet state. Another limitation is the lack of details regarding the attack. The solution notifies the network administrators that an attack will happen soon. If the solution could notify the type of attack or when the attack will happen, the administrator can handle the attack better.

The study in [159] proposes the utilization of alerts produced by IDS to predict attacks utilizing the Hidden Markov Model (HMM). The authors' first action was to create a dataset of Common Vulnerabilities and Exposures (CVE) to associate them with alerts. The solution combines the dataset with IDS alerts to avoid overfitting. Therefore, the features utilized to predict the attacks are the alert description, severity, and the CVEs associated with the alerts. The subsequent action is to train the solution in offline mode. With this training, the solution will identify the evolution of attacks to predict them. Therefore, the solution works for different multi-step attacks on the condition that the training reflects the attacks. Using data from a DDoS attack, the authors extract IDS alerts to train the solution. During training, the solution analyzes the different IDS alerts, identifies the attack states, and calculates the probability matrix. The probability matrix describes the transition probability between the network states that IDS alerts represent. The authors proposed state identification and the probability matrix generation in a supervised and unsupervised way. After training, the solution can identify the current state and measure the probability of the network evolving into an attack state. The authors evaluated the solution using the DARPA 2000 dataset [160], precisely scenario 1 (LLDOS 1.0). In order to test the solution, the authors utilized the *tcpreplay*[3] tool to reproduce the data captured in the dataset in real-time. The authors utilized the Snort IDS tool to generate alerts regarding the attack traffic. However, the solution works using alerts from other IDSs. The authors found that the technique predicts a DDoS attack about 11 min before the beginning of the attack. The solution obtained these results using the supervised training approach, while the solution using the unsupervised training approach did not predict the attack. Finally, the proposed solution requires training data to represent the evolution of different DDoS attacks. Therefore, the proposed solution cannot predict the DDoS attacks that differ from the attacks utilized in training.

[2] www.snort.org

[3] https://tcpreplay.appneta.com

The study in [161] proposes utilizing the Markov chain to predict and detect intrusions in the network. The proposed solution has three phases: the definition of states, the definition of the probability matrix, and the definition of the occurrence of events. The Markov chain depends on the possible states of the observed system to operate correctly. Therefore, the phase of defining the network states is essential for the proposal. The authors utilize a dataset that includes all the possible states of a network to define the possible states of the network. First, the solution clusters the normal data. Then the authors add the attack data, and the solution considers this data as outliers. Each cluster identified by the solution represents a system state. In the second phase, the solution calculates four probability matrices. The first matrix presents the probability of normal states transitioning to normal states. The second matrix contemplates the probability of a normal state transitioning to an attack state. The third matrix presents the possibility of an attack state transitioning to a normal state. Finally, the fourth matrix presents the probability of the system transitioning between attack states. The solution utilizes probability matrices and the Markov chain in the third phase to compute the transition between states. The solution notifies administrators if it identifies that the network may evolve from a normal state to an attack state. The authors conducted an experimental test using a DDoS attack scenario (DARPA 2000 [160]), precisely scenario 1 (LLDOS 1.0). In order to represent the clusters and the network states, the authors utilized nine features: the entropy of source/destination port number, the entropy of source/destination IP address, the entropy of packet type, the amount of ICMP, UDP, TCP SYN packets, and the total amount of packets per time. After training the solution, the experimentation results indicate that the solution has detected DDoS attacks. However, it is unclear whether the solution produces evidence of attack prediction. The solution may show the probability of the network evolving into an attack state seconds before the beginning of the attack. Therefore, this survey added the study [161] to this work. Although the authors emphasize the detection of DDoS attacks, this survey considers this study to be a prediction of DDoS attacks due to the possibility of the solution producing evidence of the occurrence of an attack before the attacker launches the attack since the authors observed that the solution could predict the occurrence of abnormal activities.

3.2.4. Statistical models

The study of [162] utilizes an incremental improvement over Autoregressive integrated moving average (ARIMA) with the Adaptive Conditional Score (ACS) Models to predict the throughput (rate of packets per minute of the network). The author hypothesizes the possibility of improving the prediction of ARIMA by making it adaptable to fluctuations. Therefore, the author added an ACS filter to ARIMA. This ACS filter has a scoring function that provides adaptability to fluctuations. In order to evaluate the proposal, the author utilized seven days of network traffic from Internet traces provided by Wand Group.[4] The proposed solution centralizes the processing of network traffic to predict the throughput of a network. The author trained the solution using two hours of network traffic throughput. For the remaining network traffic, the solution predicted the throughput with Mean Absolute Percent Error (MAPE) between 15% to 20%. In order to emphasize the importance of this result, the author performed the same experiment with the Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH). The MAPE for GARCH was between 30% and 50%. Although it is not specific to predict DDoS attacks, the proposed solution can create evidence near the launch of a DDoS attack.

The study in [163] predicts DDoS attacks using indices related to different types of threats. The authors utilize the risk levels divided into three layers where the risk indices of the lower layers are the input for the definition of the higher risk indices. The top layer represents the overall risk of the organization being cyber-attacked (Cyber Security Situation Integrated Index - CSSII). The second layer has at least three indices that define the CSSII. The second layer indexes are the infrastructure security index (ISI), the system vulnerability risk index (SVI), and the security threat index (STI). In order to define the ISI, the authors need to quantify the network traffic, service status, and resource consumption. The authors quantify SVI utilizing the third layer indexes called vulnerability index and software protection index. Finally, the trojan and DDoS indexes are STI security's third-tier indexes. Each of the seven third-tier indices has its statistical methodology to be built. However, the authors only present the statistical methodology to obtain the VSI. After defining the methodology of each third-tier index, the solution will utilize this statistical methodology to predict the variation of the indices in the future. The authors define this prediction as a prediction in the time dimension. The authors complement the in-time dimension prediction with the spatial dimension prediction. For that, they utilize Fuzzy Cognitive Maps (FCM) together with the third layer indices to perform the prediction in the spatial dimension. The authors evaluated the proposal using the DARPA 2000 [160] dataset. The results indicate an increase in the index in the first (CSSII) and second layer (ISI, SVI, and STI) indices before the onset of the attack. This increase shows that a DDoS attack will happen in the future. Especially since the STI, which comprises the DDoS index, reaches the highest value observed in the experiment before the start of the attack. Although the solution focuses on the overall risk of the organization being a cyberattack victim, this survey considered the study [163] because the DDoS index represents the risk of attacks occurring in the next few seconds. However, the authors of [163] do not present a method for obtaining the DDoS index.

In [164], the authors propose a solution for predicting network traffic close to the real one for detecting DDoS attacks. The authors divide the solution into two parts: deterministic and stochastic. In the deterministic part, the solution has a module that utilizes previous observations to train the autoregressive model (AR) incremented with the Kalman filter. The authors chose the Kalman filter to complement the AR because the Kalman filter does not need the time-series data to be stationary to perform the predictions. The solution utilizes AR to predict network traffic volume (bits per second) in the next five minutes. The authors designed the stochastic part to have three components: a module for calculating the adaptive diffusion coefficient, a module for calculating the square root of the stochastic process itself, and a module for capturing the dynamics of the evolution of the AR result. The solution utilizes the two-part modules (deterministic and stochastic) and the Mean–Reversion Stochastic Process (MRSP) to predict second-by-second traffic over the next five minutes. The output of this prediction represents normal or abnormal activities in the next five minutes. The authors utilized a real dataset to which they added simulated DDoS attacks to evaluate the solution. The results indicate that the proposed solution can predict traffic close to the real one. Although the authors emphasize the detection of DDoS attacks, this research considers this study a prediction of DDoS attacks due to the possibility of the solution predicting the occurrence of abnormal activities before the attacker launches the attack.

The study in [35] utilizes metastability theory (statistical models) to identify signals before the beginning of the attack. Metastability theory relates the equilibrium of an observable system as the parameters or conditions of the system change. A system is in equilibrium when the system's state does not change even though system parameters or conditions change. However, there are cases where the system transitions between equilibrium states as conditions change. This survey presents the following example to illustrate the transition between states. Given that a server provides a hypothetical service, the service is in a equilibrated state even when the number of users increases (observed conditions vary), and the server usually provides the service. Suppose the number of users exceeds the limits of the server's resources. In that case, a service can transition between states, going from a

[4] www.wand.net.nz/wits

state where the service is presented correctly to an error state where the server does not correctly provide the service. The authors of [35] propose a solution that collects and stores network traffic utilizing the average packet size feature in a time series to identify the transition between steady states. The proposed solution calculates four statistical indicators with the time series: asymmetry, autocorrelation, coefficient of variation, and return rate. Asymmetry checks whether the distribution of time series data is symmetric, where the distribution is similar to the normal distribution. Alternatively, if the data distribution is skewed, having biases to the left or the right. Autocorrelation estimates the level of similarity between consecutive observations. The coefficient of variation estimates data dispersion in the time series. The return rate measures the probability of the observed system returning to its previous state and recovering from a disturbance. After measuring these statistical indicators, the solution calculates the *Kendall tau* for each time series. If *Kendall tau* shows strong trends, the system issues an alert. The authors defined a threshold function that utilizes the *Kendall tau* of all indicators to define when to emit the alert. Scanning the network, exploiting vulnerabilities, and coordinating attacks are actions before the attack that justifies the variation in statistical indicators and, consequently, the possibility of the proposed solution predicting the attack. The proposed solution was evaluated experimentally in two datasets, Defense Advanced Research Projects Agency (DARPA 2000) [160] and scenario four provided by the Czech Technical University dataset (CTU-13) [165]. The scenarios contain ICMP and UDP flood attacks. As a result, the system could identify evidence of the attack two hours before the attacker launched the attack. Finally, the authors do not present accuracy, precision, or error rate metrics. Therefore, it is impossible to identify when or if the solution fails to predict the attack, for example, generating a false prediction.

In [166], the authors analyze the feasibility of detecting DDoS attacks by predicting the individual trajectory changes of users. The authors propose a solution that analyzes changes in each user's network traffic to predict whether that user will be part of a DDoS attack. The proposed solution collects the source/destination IP addresses, packet arrival time, and TCP window size to predict the average delay between transmitted packets to predict changes in each user's network traffic. The proposal classifies the user as malicious if the predicted value exceeds a threshold. The prediction happens via a statistical analysis, where the solution analyzes historical user data, determines the extrapolation of network traffic for the future, and classifies the user. The authors evaluated the solution using a simulation. During the simulation of a Slow-post attack, the proposed solution predicts the users' behavior considering the average delay between transmitted packets. Since the solution utilized data collected near the beginning of the attack, the solution became more accurate, achieving an error rate of less than 1%. The prediction of the behavior performed by the solution before the attack starts can be evidence of DDoS attack prediction. The proposed solution can fail if there is insufficient data to predict user traffic, for example, when an attacker spoofs IP addresses.

The study in [167] proposes predicting the cyber risk coefficient to predict attacks, including DDoS attacks. Centrally, the authors calculate the information entropy in the packet header, such as source and destination IP and source and destination port. The authors report the entropy of precursory moments to the current entropy to obtain the current risk coefficient. With this risk coefficient and information entropy, the authors compute the risk coefficient in the next few seconds using Gray Theory. This future risk may indicate the occurrence of an attack before its launch. Therefore, this study is apt to be considered in this work. In order to evaluate the proposal, the authors performed a simulation in MATLAB.[5] The simulation results indicate that the solution can predict the future risk coefficient close to the real risk coefficient. However, the authors do not present metrics, such as error rates or accuracy.

In [168], the authors propose a solution for predicting DDoS attacks based on the spline extrapolation of the network traffic. Spline extrapolation predicts peak traffic based on self-similar traffic to make network management efficient [169]. The authors of [168] have utilized the linear and cubic spline extrapolations to predict DDoS attacks. The linear spline consists of continuous linear functions, while the cubic spline utilizes cubic functions. The solution utilizes network traffic before, during, and after a DDoS attack to predict similar attacks. Therefore, the solution utilizes normal data and attack data as input to perform the extrapolation and predict the occurrence of future DDoS attacks. The authors performed one simulation of a DDoS attack to evaluate the solution. The feature utilized by the solution during the evaluation is the traffic intensity. However, the authors do not mention how they quantified this feature. The cubic splines presented the best results for predicting the attack in the next few seconds. Despite simple-to-implement spline extrapolation methods, the solution can only predict DDoS attacks that match the imputed ones. Therefore, future work should allow the generalization of the solution to other types of attacks and decrease dependence on labeled data.

The study in [127] is the only one identified in this survey that proposes a distributed solution for predicting DDoS attacks. The proposed solution collects and processes the network traffic in a distributed manner. The solution is installed at strategic points in the victim's infrastructure, analyzing traffic and predicting when an overload may occur. In order to select the nodes that will execute the solution, it is first necessary that the network administrators have the routing information and the network topology. The optimal approach identifies all possible paths network traffic can take to reach the victim server. With this information, the algorithm selects the nodes with the greatest network coverage possible. Fig. 7 shows a network topology with six nodes. Node 0 represents the victim server. Possible paths for network traffic to reach the victim are: {1,0}, {2,1,0}, {3,2,1,0}, {4,5,0} and {5,0}. For this scenario, nodes 1 and 5 represent the ideal combination for installing the solution because they provide the most coverage over network traffic. Although this approach identifies the best combination of nodes to install the solution, it is not recommended for larger networks as it takes much processing to identify all possible paths. The authors proposed two other approaches, the Maximum-Coverage-Node-First (MCNF) Approach, and the Weak-Path-First (WPF) Approach to select the best nodes utilizing less processing. Both approaches utilize less processing to define nodes 1 and 5 as ideal nodes to install the proposed solution utilizing the topology presented in Fig. 7. The selected nodes will execute the solution for DDoS attack prediction utilizing the network traffic history. The solution utilizes historical data in time series to mine network traffic patterns and trends. Each node executes the solution and utilizes ARIMA to create a model that predicts expected traffic based on historical data. This model calculates network traffic and infers when a possible overload may occur. If any node identifies a possible overload, this node issues an alert with the details of the possible overload. The authors evaluated the solution utilizing 60 simulations where simulations had up to 250 nodes and different network topologies. The first evaluation verified the performance for selecting the best nodes for the solution installation of the MCNF and WPF approaches. The results indicate that the approaches selected the same nodes as the optimal approach, utilizing less time and consuming fewer resources. The authors do not indicate which feature of network traffic they utilized to predict DDoS attacks during the evaluation. Furthermore, there is no evidence in the results about the timing of the attack for which the solution issues a warning. However, if this solution can create alerts seconds before the attack begins, this survey can classify this as a solution for DDoS attack prediction. Finally, the authors argue that the distributed architecture prevents the proposed solution from overloading at the single point of failure because several nodes will divide the task to process the network data. Furthermore, according to the authors' understanding, the solution that processes network data centrally cannot fully satisfy the requirements of the users, as only a distributed solution can deal with distributed attacks.

[5] www.mathworks.com/products/matlab.html

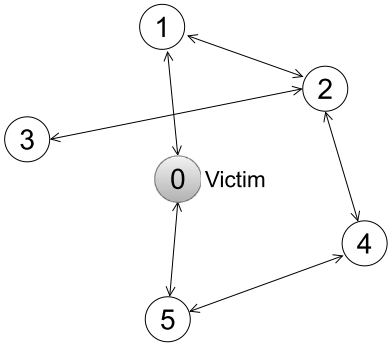


Fig. 7. Sample network topology with six nodes (based on [127]).

3.3. Long-term prediction

This subsection presents all the studies classified as long-term predictions. The centralized aspect is the architectural aspect of these studies. Furthermore, these studies utilize machine learning, statistical models, or hybrid methodological aspects. In order to simplify the exhibition of studies, this survey utilizes the methodological aspect to present the long-term prediction. However, each study has its complete classification throughout the text.

3.3.1. Machine learning

The study in [170] utilizes external data sources to train SVM to predict security events. The hypothesis underlying this study is the possibility that malicious activities originated by the same network, for example, in an Autonomous System (AS), present stable behavior. The authors collected two data types from the Internet, reputation blacklists and security events. Reputation blacklists represent the reputation of IP addresses possibly involved in attacks. Security events are attacks reported on the Internet that the authors utilized to validate the predictions. The authors collected and grouped data from 11 reputation blacklists per day between January 2013 and February 2014. Using the IP addresses of each day, the authors identified the ASs and the Border Gateway Protocol (BGP) prefix related to the IP addresses. The authors collected the security events reported by [171] between October 2013 and February 2014. Using these security events, the authors identified the domain names of the victims of the attacks and found the BGP prefixes related to the domain names of most events. The authors selected the security events in October 2013 and identified the BGP prefixes related to these security events to define the training base. After identifying the BGP prefixes, the authors searched the entire history between January and September 2013 for the BGP prefix. For each BGP prefix, the authors extracted the duration and frequency that the BGP prefix contains IP addresses in the reputation blacklists. These BGP prefixes were labeled as belonging to an attack. In order to complement the training base and define the BGP prefixes not related to attacks, the authors randomly selected 1000 BGP prefixes that are not related to attacks in October 2013. The authors process this data centrally and predict the occurrence of events related to cybersecurity with an average true positive of 69% for the next three months, from November 2013 and January 2014. Although not focused on DDoS attacks, the security events include DDoS attacks. For this reason, this survey considers this study in this work. Employing reputation blacklists makes the solution susceptible to errors because, with the current dynamics of IP addresses, an IP address included in the blacklist can start to be utilized by normal users. Therefore, the solution can deny service to users without malicious interests. Furthermore, not focusing on DDoS attacks undermines the accuracy of the solution for DDoS attacks, which is the context of this survey. Therefore, solutions specialized in DDoS attack prediction can be more accurate than non-specialist solutions.

In [172], the authors propose a solution that monitors relevant texts from external data sources. Social networks such as Twitter are relevant data sources for the proposed analysis. The aim of monitoring social media texts is to centralize the processing of related tweets and utilize them to build deep learning models to predict the probability of a DDoS attack. The solution creates a matrix in which each position stores a word present in the tweet. The solution enriches the data by performing a sentiment analysis on each tweet. The solution performs the prediction of DDoS attacks using the texts of the tweets aggregated daily, weekly, and monthly. The goal is to improve results by identifying real trends in historical data; this is possible because the model has information at different degrees of granularity. The authors evaluated the solution experimentally using a proprietary dataset available on GitHub.[6] The dataset contains the target and date of 170 DDoS attacks. The authors identified these attacks on news sites. In order to predict attacks, the authors collected 17 760 tweets from August 2015 to April 2016. The authors utilized the area under the curve (AUC) to assess the results. The authors utilized the Long short-term memory (LSTM) to perform the prediction. The DDoS attack prediction AUC for the next day was 0.305, less than 0.5. Therefore, the solution does not predict many of the attacks present in the test base. In addition to the difficulty in predicting DDoS attacks, another limitation to the solution is that attackers will not always expose the intention to attack. Therefore, the solution may fail to predict some DDoS attacks for a lack of data. While it is possible to predict DDoS attacks using data from social networks, attackers can manipulate this strategy if they threaten to conduct an attack and do not launch it. Therefore network administrators can waste time and money preparing for an attack that will not happen.

In [173], the authors utilize an Artificial Neural Network (ANN) [174] to predict which botnets could perform attacks on the following day. The authors utilized an ANN with three layers, the input layer, an intermediate layer called the hidden layer, and the output layer. Because the ANN is a supervised machine learning technique, it is necessary to train the model using historical data. Therefore, to train the ANN, the authors utilize past attacks and make a list grouping the attacks by day. The authors utilized 24 features to represent the attack data, including the IP address of origin and destination of the attack, the source and destination port of the attack, the payload of the message, the type of botnet that conducted the attack, geographic data of the IP addresses present in the attacks, and data of the HTTP protocol. The authors evaluated the ANN with different combinations of neurons in the hidden layer and utilized historical data obtained from a dataset provided by Cybersecurity Malaysia. This dataset contains several DDoS attacks collected in 2016. For each combination of neurons in the hidden layer, the authors performed 15 tests. The combination of neurons in the hidden layer that presented the best average result was the ANN with 500 neurons. The Mean Square Error (MSE) for the ANN with 500 neurons was 0.0053. Therefore the proposed solution identifies attacks that will occur the next day with a low error rate. The proposal has two limitations related to the data utilized. Firstly, it is opportune to preserve users' privacy by avoiding using the packet payload. In addition, using geographic data makes it challenging to utilize the proposed solution because attackers can spoof the IP address, making the information unreliable. Finally, the authors performed several tests to obtain an ANN configured to minimize the occurrence of errors. However, an ANN with 500 neurons may not be the ideal solution for other scenarios. Therefore, identifying the correct ANN configuration is still an open issue.

[6] https://github.com/wangzq870305/ddos_forecast

3.3.2. Statistical models

The study in [175] proposes a solution for predicting events related to cybersecurity. Although it does not focus on predicting DDoS attacks, it can perform accordingly. The proposed solution includes data collection, text mining infrastructure, and the warning generation methodology. Data collection consists of searching for information on the Internet. Twitter is the first data source. The authors collected the tweets of 69 cybersecurity experts using the official Twitter API. In addition to Twitter, the authors selected 290 security blogs to collect information about vulnerabilities, exploits, and other issues in cybersecurity to enrich the data collected on Twitter. The last data source is dark web forums. In this case, the authors selected 263 sites not accessible using standard browsers. This data collection aims to obtain information about malicious hackers, fraud, and data breaches. The authors utilize the data collected on Twitter and security blogs as input for alert generation. The solution stores this data in vectors where each position is a word from the tweet or forum post. The solution utilizes data collected on dark web forums to monitor possible threats found by systems. The solution identifies and maintains the vectors in English to complete the data collection. The solution also removes symbols, numbers, URLs, and duplicated words. The first action performed in the text mining infrastructure step is to reduce the number of positions in the vectors by filtering out irrelevant terms. The solution removes English terms such as "hello" and "because" and stopwords such as "on" and "for". The solution also removes terms known to cybersecurity as "DDoS" and "phishing"; however, the solution marks that the vector has this type of word. The solution then analyzes each collected vector to remove duplicated terms. The solution analyzes the data hourly and daily. When it encounters a new term, the solution issues a warning that could represent a future attack. The authors evaluated the solution between September 2016 and January 2017. The solution showed 81% accuracy in detecting cybersecurity-related events. The solution was able to identify potential threats before the attacks became public. Despite the accuracy being 81%, the solution does not predict attacks if there are no mentions of attacks in the analyzed bases.

The study in [176] evaluates how to predict the trend of DDoS attacks utilizing political, diplomatic, informational, military, and economic events collected from external data sources. In order to predict the trend of DDoS attacks, the authors run simulations with events aligned with data to represent the relation between countries to verify whether the DDoS attack tendency will increase or decrease. The proposal is to model each event according to the destination, hostility, severity, origin, and type. For example, in 2016, the United States of America (USA) publicly claimed that Russia influenced the presidential election [177]. The authors simulated the effects of this action on the number of attacks. In these cases, the authors utilize statistical methods to correctly predict that there would be an increase in DDoS attacks in the USA and Russia more than one day in advance. Besides this, the authors verified that the model did not represent the reality to predict the DDoS attack variation using military and economic events. For example, simulation results indicated an increase in DDoS attacks against the US and Russia when Russia decided to leave the Nuclear Security Pact [178]. However, the real data indicated a decrease in attacks on the USA and Russia. The purpose of [176] is to predict the variation in the number of DDoS attacks caused by different events. Although [176] does not predict that an attack will occur soon, network administrators can utilize the trend of increasing attacks to prepare for attacks; this is why this survey considers the study [176] in this work. One significant limitation was that the evaluation had only five events; this makes it difficult to generalize the results obtained in [176].

3.3.3. Hybrid models

The authors of the study in [135] analyzed a non-public dataset containing 50,704 real DDoS attacks[7] to design three data-driven models to predict DDoS attacks. The authors divided the dataset into training and testing. The training base consists of 40,563 attacks, and the test base contains 10,141 attacks. The first proposed model utilizes the ARIMA model to predict temporal features such as activity level, attack magnitude, and source distribution. Activity level represents the average number of realized attacks in the dataset daily. Attack magnitude represents the total number of IP addresses that attackers utilize in attacks. The source distribution is an average of the distances between the ASs where the IP addresses involved in the attacks were registered; the authors measure this distance in hops. The authors extracted these features from all DDoS attacks and organized them into time series forms. The first model utilizes ARIMA to analyze the training base and predict the behavior of these variables for the next attack (test base). The results indicate that it is possible to utilize ARIMA to predict these features since the predictions are similar to the real values. The second model utilizes a neural network to predict spatial features such as targets' localization, DDoS attack time, and the onset of attacks. The authors obtain the targets' localization based on the autonomous system numbers (ASN) obtained by analyzing the IP addresses that participated in the attacks. The DDoS attack time is the duration of attacks. The start of the attacks represented when the bots launched attacks against the victim. The second model utilizes these features and considers each bot a neural network neuron. The neural network utilizes an input layer, a hidden layer, and an output layer. It is worth noting that calibrating the neural network is essential to obtain the best predictions, and the configuration identified by the study authors [135] may not be the optimal value for other datasets. The results indicate that the second model predicts the origin of attacks at the AS-level with an error rate near zero. The authors proposed a third model, the union of the first two, to predict DDoS attacks and not just attack features. For this, the authors trained a Regression Tree combining the output of the first two models. The results indicate that when comparing the prediction of the third model with the test data, the model predicts DDoS attacks with a root-mean-square deviation (RMSE) of 2.72 days, more than a day before the launch of the attack. Therefore, by combining the first two models, the third model decreases the error rate and improves the accuracy in predicting when the next DDoS attack will occur. Despite the results presented, the dataset includes data on DDoS attacks between 2012 and 2013. Therefore, this dataset does not include DDoS attacks from modern botnets such as Mirai, first recorded in 2016 [179]. Because the authors based the proposed models on training data, the models may not obtain the same results in current attack scenarios. In addition, the models utilize data based on the location of the IP address of the bots participating in the attacks. However, the attacker can mask the bots' real IP address making this information unreliable.

Table 1 summarizes the classification of all studies covered in this survey. Each row in this table describes one of the studies and the classification for all aspects. In addition to the classification criteria, the table reports the evaluation methods and the publication year of the study.

3.4. Discussion

Predicting DDoS attacks increases the time available to network administrators and security teams to avoid losses caused by attacks. Therefore, identifying the best time to predict attacks is critical. Long-term predictions are desirable because they provide more time to avoid losses caused by DDoS attacks. However, making long-term predictions is not trivial. The studies [135,170], and [172] illustrate this. In [170], the authors have achieved an average true positive of 69% for predicting cybersecurity-related events that would occur in the next three months using data collected from the Internet. In [172], the authors predicted the occurrence of attacks for the next few days with a low AUC (less than 50%). Even using data from more than 40 thousand attacks to reverse-engineer the attacks, the study of [135] presented an RMSE of 2.72 days. The authors of [138] obtained an F1-score of

[7] www.team-cymru.org

Table 1
Summary of DDoS prediction studies.

Ref.	Time	Architecture	Method	Data	Evaluation	Year
[138]	Short-term	Centralized	Hybrid	Network Traffic	Experimentation	2021
[140]	Short-term	Centralized	Machine learning	IDS Alerts	Simulation	2015
[142]	Short-term	Centralized	Machine learning	Network Traffic	Simulation	2011
[144]	Short-term	Centralized	Machine learning	Network Traffic	Experimentation	2017
[148]	Short-term	Centralized	Machine learning	Network Traffic	Experimentation	2017
[150]	Short-term	Centralized	Machine learning	Network Traffic	Experimentation	2021
[153]	Short-term	Centralized	Machine learning	Network Traffic	Simulation	2018
[125]	–	Centralized	Machine learning	Network Traffic	Experimentation	2020
[136]	Short-term	Centralized	Markov based	Application logs	Experimentation	2013
[137]	Short-term	Centralized	Markov based	Application logs	Simulation	2021
[158]	Short-term	Centralized	Markov based	IDS Alerts	Experimentation	2016
[159]	Short-term	Centralized	Markov based	IDS Alerts	Experimentation	2020
[161]	Short-term	Centralized	Markov based	Network Traffic	Experimentation	2013
[162]	Short-term	Centralized	Statistical Models	Network Traffic	Experimentation	2015
[163]	Short-term	Centralized	Statistical Models	Network Traffic	Experimentation	2018
[164]	Short-term	Centralized	Statistical Models	Network Traffic	Simulation	2019
[35]	Short-term	Centralized	Statistical Models	Network Traffic	Experimentation	2018
[166]	Short-term	Centralized	Statistical Models	Network Traffic	Simulation	2020
[167]	Short-term	Centralized	Statistical Models	Network Traffic	Simulation	2009
[168]	Short-term	Centralized	Statistical Models	Network Traffic	Simulation	2022
[127]	Short-term	Distributed	Statistical Models	Network Traffic	Simulation	2015
[170]	Long-term	Centralized	Machine learning	External data sources	Experimentation	2015
[172]	Long-term	Centralized	Machine learning	External data sources	Experimentation	2017
[173]	Long-term	Centralized	Machine learning	Hybrid	Experimentation	2018
[175]	Long-term	Centralized	Statistical Models	External data sources	Experimentation	2018
[176]	Long-term	Centralized	Statistical Models	External data sources	Simulation	2017
[135]	Long-term	Centralized	Hybrid	Network Traffic	Experimentation	2017

82.81% predicting attacks that will occur in the next 20 s, and in [158], 81% of predictions occurred less than one minute before the start of the attack, and the maximum value obtained was 865 min (14 h and 25 min). These results reinforce the hypothesis of a trade-off between the prediction time and the accuracy of the solutions. Increasing the prediction time can cause an increase in the error rate. Decreasing the prediction time can make solutions more accurate. However, the time to deal with attacks decreases. Finally, the literature may confirm this hypothesis in the future because, with the low number of identified studies, this survey does not have representative data to confirm it.

Imbalanced data is an inherent problem in predicting DDoS attacks. As mentioned, solutions for DDoS attack prediction must predict attacks utilizing minimal information related to attacks [34]. Therefore, it is reasonable to argue that the solutions operate in environments where the most significant amount of data originates from normal users, not attackers. This difference in the amount of normal and malicious data makes it difficult to create solutions to predict attacks because authors must demonstrate that the solutions can handle potential data imbalances. Machine learning can illustrate this problem. Classical machine learning techniques can present high error rates when the imputed data is imbalanced. One way to solve this is to pre-process data. Under-sampling is a strategy to equalize the class data by decreasing the number of instances of the majority class, while over-sampling increases the amount of data of the minority class [180]. Although these strategies assist in the learning process, caution is imperative because these sampling techniques can negatively influence the result of the solutions, or the solution may be overfitted for training data, not working in real environments. Another opportunity to deal with imbalanced data is to utilize outlier detection [181]. Finally, employing correct evaluation metrics is essential. Accuracy is a metric widely utilized in the literature to evaluate the results of solutions. However, when there is imbalanced data, it is essential to complement it with other metrics. Receiver operating characteristic (ROC) curve, precision, recall, and F1-score are metrics that complement the evaluation of solutions [182].

The mechanisms of DDoS attacks the solutions can predict is a critical discussion to the evolution of attack prediction solutions. Table 2 highlights the data aspect, the datasets employed during the evaluation, the attacks evaluated, and the features utilized in each study, presenting them in the same order as in Table 1. Table 2 shows that the solutions predict different attack mechanisms employing diverse combinations of features. The attack mechanisms with more studies identified in this survey were TCP SYN with five studies and ICMP flood with two. This survey did not identify the type of attack evaluated in some studies, such as in [140,153]. Furthermore, studies such as [125,158] utilize botnet actions to predict attacks regardless of the mechanism utilized by the attacker. The literature complements the discussion about which attack mechanisms it is possible to predict, presenting the possibility for attackers to perform tests against the victim before effectively launching the attack with the whole botnet [144]. Furthermore, the phases before the attack can cause variation in network traffic, providing the signals to predict the attacks [183]. Therefore, defining which attack mechanisms the literature can predict is not a trivial task because there are several attack mechanisms, attackers constantly evolve attack mechanisms, and different network topologies need protection. Despite the limited amount of studies identified and the non-triviality in identifying the types of attacks that can be predicted, in the future, the literature can clearly define the limits for the prediction of DDoS attacks.

It is also essential to discuss how to evaluate the proposed solutions. The authors evaluate the solutions differently to demonstrate that the solutions are feasible, and Table 2 summarizes the datasets employed during the evaluations. In order to conduct the experimentation, some studies evaluate solutions with different datasets available in the literature, such as DARPA 1998, KDD CUP 1999, and CTU-13. Four studies evaluate their solutions utilizing the DARPA 2000 dataset because its documentation clearly describes the stages of the attack. Other studies collect their data and may not make the dataset available. Furthermore, some authors utilize real data gathered from collaboration with private corporations; this data is generally unavailable to other researchers to prevent sensitive information from becoming public. In addition to experimentation, some authors utilize software such as MATLAB and NS-3[8] to simulate DDoS attack scenarios and evaluate solutions. Utilizing data in the literature encourages replication, comparison, and inspection of studies. Simulation allows the evaluation of scenarios in

[8] www.nsnam.org

Table 2
Features and datasets utilized in studies to predict DDoS attacks.

Ref.	Data aspect	Dataset	Attack	Feature
[138]	Network Traffic	DARPA 1998	ICMP flood	Number of packets per second
[140]	IDS Alerts	Simulated	–	Alerts' description, priority level, protocol, sensor information, source/destination IP and Port, time, and type
[142]	Network Traffic	Simulated	Malformed Packet	Number of defective hosts and the number of authentication attempts by time unit
[144]	Network Traffic	KDD CUP 1999	TCP SYN	Duration, protocol, service, flag, source, destination, and the number of failed logins, among others
[148]	Network Traffic	Collected	Bandwidth Depletion	The number of non-Duplicated IP addresses that try to connect via SSH on the servers and importance level of servers that can be targets
[150]	Network Traffic	DARPA 1999	Bandwidth Depletion	Number of packets per 10 s
[153]	Network Traffic	Simulated	–	–
[125]	Network Traffic	CCC dataset	C&C communication	The proportion of received packets whose size is between 1200-1299 bytes, the proportion of received packets whose size is more than 1400 bytes, the proportion of packets with the RST flag, the proportion of the packets with the FIN flag, and the shortest time interval between received packets, among others
[136]	Application logs	Collected	–	Capture time, the source and destination of the log, the log size, and the event type
[137]	Application logs	Simulated	–	Users' logins, device state, and system operations
[158]	IDS Alerts	SysNet and ISCX Botnet	Bot traffic	–
[159]	IDS Alerts	DARPA 2000	TCP SYN	The alert description, severity, and the CVEs associated with the alerts
[161]	Network Traffic	DARPA 2000	TCP SYN	The entropy of source/destination port number, the entropy of source/destination IP address, the entropy of packet type, the amount of ICMP, UDP, TCP SYN packets, and the total amount of packets per time
[162]	Network Traffic	Wand Group	–	Rate of packets per minute
[163]	Network Traffic	DARPA 2000	TCP SYN	–
[164]	Network Traffic	Simulated	–	Bits per second
[35]	Network Traffic	CTU-13 and DARPA 2000	ICMP and UDP flood, and TCP SYN	Average packet size per second
[166]	Network Traffic	Simulated	Slow-post attack	Source/destination IP addresses, packet arrival time, and TCP window size
[167]	Network Traffic	Simulated	–	Source/destination IP and port
[168]	Network Traffic	Simulated	–	Traffic intensity
[127]	Network Traffic	Simulated	–	–
[170]	External data sources	Collected	–	IP addresses involved in attacks, security events, and BGP prefixes
[172]	External data sources	Collected	–	Text of tweets
[173]	Hybrid	Cybersecurity Malaysia	Botnet Attacks	IP address of origin and destination of the attack, the source and destination port of the attack, the payload of the message, the type of botnet that conducted the attack, geographic data of the IP addresses present in the attacks, and data of the HTTP protocol
[175]	External data sources	Collected	–	Text of security blogs and tweets
[176]	External data sources	Simulated	–	Destination, hostility, severity, origin, and type
[135]	Network Traffic	Team Cymru	Bot traffic	Activity level, attack magnitude, source ASN, duration of attacks, and the start of the attack

that the literature does not have datasets. Evaluating the solutions is essential because it demonstrates that the solution works under the conditions presented, clarifies the limitations of the solution, and helps to comprehend the proposal.

The features employed by existing solutions to predict the attacks are essential because their analysis shows that the solutions can identify the signals of the preparation of the attacks. Each data aspect defines the limits for obtaining the features. For example, it is possible to get more than 80 features from analyzing network traffic [184,185]. Table 2 highlights the features utilized in the studies. Firstly, it is worth noting that this survey did not identify the features utilized by the studies [127,153,158,163]. These studies generally mention that they utilize network traffic or IDS alerts but do not specify the features extracted from the network traffic or IDS alerts. Table 2 shows the diversity of features utilizes for the prediction. The study [144] utilizes all 41 features available in the dataset.[9] Studies, such as [138, 162,164], and [35], employ only one feature to predict attacks. The source/destination IP and port are the most utilized features, present in 14,81% of the studies [140,161,167,173] identified in this survey. Finally, this survey highlights that the studies [159,161], and [35] use the same dataset (DARPA 2000) but they rely on different features.

4. Open issues

Due to attacks constantly evolving, defense mechanisms must evolve as well. Therefore, it is essential to decrease the error rate of solutions by making them more accurate and adaptable to different contexts. Avoiding errors culminates in better system adoption, avoiding wasting money and time. This section reports the open issues identified by this survey study. Open issues present research opportunities that the literature has not explored and may evolve state-of-the-art. With this in mind, this section presents several research opportunities and follows the structure of Fig. 5.

[9] kdd.ics.uci.edu/databases/kddcup99/kddcup.names

4.1. Short-term prediction

4.1.1. Machine learning-based predictions

This survey has identified ten studies utilizing machine learning for DDoS attack prediction. These studies utilize techniques such as SVM and K-means, as well as neural networks such as ANN and RBF neural networks. However, these studies can be extended by following at least seven hypotheses. (1) Decreasing the dependence on labeled information: Solutions can utilize semi-supervised [186] or unsupervised machine learning techniques such as DBScan [187] and Self-Organizing Map (SOM) [188] to predict DDoS attacks without utilizing labeled data. (2) Ensemble techniques combine machine learning classifiers with attempting to improve accuracy and precision, decreasing the error rate [189]. AdaBoost [190] and Gradient Boosting [191] are examples of ensembles that could increase the accuracy of machine learning-based solutions. (3) Utilizing deep learning techniques to predict DDoS attacks: Recently, deep learning-based solutions have gained momentum [192] in the solution of problems in various domains, such as visual recognition [193] and robotics [194]. Modern deep learning techniques, such as Gated recurrent unit (GRU) neural networks [195] and Deep belief networks (DBN) [196], can avoid over-fitting [197], do not demand manual feature selection [198], and improve results. (4) Explainable artificial intelligence models: Explainability is a critical issue in machine learning-based solutions while preventing losses caused by potential prediction errors. That is because these models focus on transparency and interpretability [199]. Therefore, whether the administrators recognize suspicious predictions, they can interpret how the model performed the prediction. (5) Self-configurable systems: This represents the systems capable of self-configuration [200,201]. Consequently, the solution extracts the best from the machine learning technique with the minimum human interaction possible [202]. (6) Federated learning: Federated learning is an option to encourage the creation of distributed solutions utilizing machine learning. The federated learning solutions create machine learning models using data collected at different locations. Consequently, solutions utilize various pieces of data to perform classification tasks [203]. (7) Reinforcement learning (RL): RL is an alternative to decrease dependence on labeled data. The literature presents the possibility of detecting attacks using RL [204,205]. Therefore, this survey hypothesizes the possibility of solutions using RL to predict DDoS attacks.

Another challenge to consider in future solutions is adversarial machine learning [206]. Machine learning is beneficial in favor of information security [207] due to advantages such as the ease of dealing with large amounts of data and the diversity of existing techniques. On the other hand, attackers can influence the training of machine learning techniques or learn how solutions that utilize machine learning operate [208]. Therefore, attackers can exploit vulnerabilities to deceive them, making the result of DDoS attack prediction incorrect or even creating new ways to conduct the attack to avoid the prediction [209]. Finally, attackers can create false alarms to activate mitigation mechanisms, increasing service-related costs. Therefore, it is opportune for the literature to discuss manners to avoid the consequence of adversarial machine learning on DDoS attack prediction solutions.

4.1.2. Markov-based predictions

This survey has identified five studies that utilize techniques based on Markov theory, three utilize a Markov chain, one utilizes a stochastic Markov process, and one utilizes an HMM. One way to advance the previous studies is to test variations of the Markov chain and the HMM, such as Markov random field, Hierarchical Markov models, Tolerant Markov model, and Markov-chain forecasting models. Another way to advance previous studies is to resolve the limitations pointed out in current solutions. Some limitations are due to the limited number of states that do not wholly represent reality [158], the difficulty in generalizing the solution to predict different attack mechanisms [159], and the need to improve the accuracy of predictions [161]. Furthermore, this survey argues that utilizing the Markov decision process [210] and the partially observable Markov decision process [211] can advance the state-of-the-art supporting feasible solutions with low error rates.

4.1.3. Statistical models

This survey has identified ten studies utilizing statistical DDoS attack prediction models, one of the most studied aspects. An opportunity for improvement and evolution is to create new solutions is to utilize ARIMA variations [162]. Seasonal Autoregressive Integrated Moving Average (SARIMA) and Seasonal ARIMA with eXogenous Factors (SARIMAX) are variations of ARIMA that deal with seasonal data. Therefore, this survey hypothesizes that solutions can utilize the SARIMA and SARIMAX to predict DDoS attacks considering seasonal, for example, network traffic [212]. The literature reinforces this opportunity by showing that SARIMA and SARIMAX can be employed to predict electricity consumption [213,214], and natural gas consumption [215]. Furthermore, due to the success achieved with solutions based on neural networks, this survey shows the possibility of new solutions predicting DDoS attacks utilizing Bio-inspired computing. Bio-inspired computing has received the attention of researchers, being utilized in areas such as Astronomy, Computer Science, and Mathematics [216]. Studies on combating DDoS attacks also utilize Bio-inspired computing [217,218]. In [219], the authors propose an attack detection solution based on an ANN with its initial parameters defined by a bee colony algorithm. Thus, Bio-inspired computing can compose solutions for predicting attacks and evolve the state-of-the-art.

4.2. Long-term prediction

4.2.1. Machine learning

This survey has identified three studies capable of producing long-term predictions utilizing machine learning [170,172,173]. This survey highlights some issues, such as the low accuracy [172], the need to present probabilities of the occurrence of DDoS attacks [170], and the concern with computational performance [173]. Producing correct long-term predictions is one of the most challenging open issues. Incorrectly performed predictions can cause unnecessary costs, and unrealized predictions can leave a service vulnerable. Therefore, the minimization of the error rate needs constant evolution. One way to shift responsibility and alleviate this problem is to produce probabilities about a future attack. In the Sklearn library [220], techniques such as SVM and Decision Trees can show the outputs of predictions and the probability for these predictions. Consequently, security teams can decide what actions to take in different cases. Finally, the performance to make attack prediction is also crucial. If the solution delays the production of the prediction, part of the prediction advantage is lost.

4.2.2. Statistical models

This survey highlights the possibility of proposing long-term solutions for predicting DDoS attacks inspired by statistical weather forecasting models. The weather forecasting has gained huge investment [221]. The developed countries forecast the weather a few days in advance, and some solutions forecast floods up to ten days in advance by correlating the weather forecast with hydrological models [222]. Although statistical weather forecasting models are complex and specific, the area of weather forecasting is mature. Therefore, it is possible to increase the time to predict DDoS attacks by utilizing the lessons learned from weather forecasting.

4.2.3. Markov based

This survey has identified that studies that are Markov based as a methodological aspect utilize network traffic or IDS alerts in the data aspect and are short-term in the time aspect. One hypothesis to change the time aspect of these solutions to the long-term is the utilization of data from external data sources, such as Twitter. The hypothesis could be confirmed if the study could map the precursor stages to the attack and produce an evolution probability to an attack state before the attacker launches the attack. The consolidation of this hypothesis is not a trivial action. However, as there are solutions that utilize IDS alerts, it is possible to utilize these studies as a basis for proposing new studies that utilize external data sources.

4.2.4. Hybrid

The union of different studies is an open issue that can result in new solutions that can alleviate each solution's intrinsic disadvantages. An example is the union of solutions that utilize and process different data types, generating a hybrid solution about the data and the methodological aspects. A plausible hypothesis is the union of solutions that utilize network traffic with solutions that utilize data available from external data sources; this will increase the solutions' prediction time and improve accuracy. This possibility exists because this survey has classified four studies that utilize external data sources as long-term, which indicates that they can predict DDoS attacks at least one day in advance. The disadvantage of these studies is the possibility that the solution does not identify an attack or the elevated chance of generating false positives. Furthermore, this survey has classified 15 out of 16 studies that utilize network traffic as short-term. Thus, to increase accuracy, the literature can combine solutions that utilize data available in external data sources with solutions that utilize network traffic. Because when one solution fails to predict the attack, the other complements the prediction.

4.3. Others issues

One of the most critical open issues is the limited number of papers concerned with distributed architecture. Among the 27 studies, this survey has identified only one study with distributed aspects. In cases where the information processed by the solution is extremely voluminous, it is plausible to imagine that solutions that centralize all processing in a single point may not analyze all the information in time to perform the DDoS attack prediction. However, it is challenging to propose solutions capable of processing information at different points that still manage to predict DDoS attacks. One opportunity to evolve the study of [127] is to replace ARIMA with lightweight machine learning techniques, such as decision trees, logistic regression, and SVM with linear kernels [223], to run the solutions on routers. Another opportunity is to execute solutions on hosts instead of analyzing data from a network, such as in routers. Therefore, the solution would have a distributed aspect because all network hosts will process their network traffic data. However, this research opportunity can operate only in cases where it is necessary to analyze the internal traffic of the networks. In addition, coordinating the issuance of alerts of possible DDoS attacks can be a challenging task. Even with the challenges related to creating solutions that operate in a distributed manner, they have advantages such as ease of scaling. They require less computing power, and installing them on devices with fewer resources, such as routers, is possible. Furthermore, the solution only partially stops working if a part becomes unavailable.

The selection of features utilized for predicting DDoS attacks can generate many contributions. Section 2 presents the diversity of attack mechanisms and the typical phases of a botnet. Section 3 shows that there is still no consensus on the ideal features to predict DDoS attacks. Therefore, identifying the features most affected by the phases preceding the beginning of each type of attack is essential to evolving the prediction of DDoS attacks and decreasing the error rate. This open issue is relevant for all data aspects (Fig. 5) because the solutions can invest computational resources only to process the data that help predict DDoS attacks. Besides, the features may vary according to the attack mechanisms, and they may vary according to the methodology employed by the solution (Fig. 5). For example, the ideal feature set for building machine learning models may not be ideal for building solutions based on the Markov chain. Therefore, each solution may have a different feature selection process. The selection of features for detecting DDoS attacks is a topic addressed in the literature [224–228]. Therefore, it is possible to begin the analysis for selecting features collected in the network traffic for the prediction by having as a model the studies that carried out the selection of features for the detection of DDoS attacks. Techniques such as RF [229], genetic algorithm [230], and analysis of variance [231] can select the most relevant features for the prediction of DDoS attacks. Finally, each attack phase can be a specialized feature selection; for example, the study of [232] focuses on selecting features for the C&C communication.

In addition to improving the success rate of solutions using the selection of features, it is opportune that the new solutions for predicting DDoS attacks provide details about the attacks. The results of the analysis of the studies presented in Section 3 indicate that, generally, the solutions only notify network administrators of a future attack. In some cases, as in solutions based on Markov chains, they present the probability of occurrence of the attack. This survey hypothesizes that the non-triviality of the topic and the limited amount of study that performs the prediction of DDoS attacks cause the lack of details about DDoS attacks. Therefore, new research opportunities may point out: (*i*) the probability of the attack happening; (*ii*) the moment of occurrence of the attack; (*iii*) the magnitude of the attack; (*iv*) the attack mechanism, among other details about the attack. These details made the solutions easier to utilize because network administrators can take different actions as the details vary.

Another open issue is the need for solutions to adapt to different scenarios. Section 3 shows that studies evaluate solutions through experiments or simulations. However, in real environments, the statistical distribution of the system's input data can vary throughout the execution of the system. Typical reasons such as the beginning of the workday or the evolution of attacks can cause variation in the statistical distribution of the input data. Therefore, it is essential to evaluate the behavior of solutions in the face of changes and create mechanisms that avoid the degradation of results if the dynamics of the data changes. In order to identify changes in concept and adapt to new scenarios, solutions can utilize techniques such as Adaptive Windowing, Concept Drift Detection, and Early Drift Detection Method [233]. The literature reinforces the utilization of change-of-concept detection techniques in cybersecurity because intrusion detection solutions [234] and botnet detection solutions [235,236] can utilize these techniques.

The IoT has provided a cyber revolution in recent years [237] and is applied to solutions in different areas [238], such as industry [239], health care [240], and agriculture systems [241]. However, IoT has also revolutionized the level of cyber threats [242,243], introducing new challenges for cybersecurity solutions. A relevant example of new cybersecurity challenges is the attack performed in 2016, where attackers utilized IoT devices to conduct an attack by disrupting access to important services such as GitHub, Twitter, and Netflix [14,98]. Of the studies identified in this work, only three consider IoT networks [137,150,175]. Although few studies consider the specifics of IoT networks, the studies mentioned above may operate for IoT networks. Nevertheless, there may be a performance degradation due to the peculiarities of IoT networks. Therefore, it is appropriate that new studies consider this new paradigm and produce specific solutions to deal with IoT networks' unique characteristics and challenges [244], or at least present results that consider the IoT networks.

As with IoT networks, SDN architecture is getting more attention in industry and academia. SDN architecture provides some advantages, such as cost savings related to equipment acquisition and installation [245], network reconfiguration opportunities [108], and deep packet inspection [108]. As in IoT networks, specialized solutions for the SDN paradigm can outperform the accuracy of generalist solutions. This survey has identified at least two studies on predicting DDoS attacks in SDN [135,162]. However, new efforts are still needed to deal with the specifics of SDN architecture to provide administrators more time to deal with upcoming DDoS attacks.

In [25,33], the authors highlight the need for cooperation between different network entities to create strategies to improve defense against DDoS attacks. Studies such as [246,247] propose cooperative solutions for detection and/or mitigation. Collaboration solutions to defend against DDoS attacks can be between victims' devices and the attack source or between devices allocated on the core networks and

the victim [248]. In [249], the authors cite the possibility of detecting attacks in their early stages by combining information from multiple networks. Distributed architectures, alert correlation, user privacy, and high accuracy are challenges that collaborative solutions must address [249]. Even with many challenges, this survey hypothesizes the possibility of proposing collaborative solutions to predict DDoS attacks.

The theory of graphs solves problems in areas such as physics, chemistry [250], and geometry [251]. Graph-based solutions predict traffic demand [252] and future traffic speeds [253]. DDoS attack detection solutions can also utilize graphs [254]. This solution maps communication patterns between bots and victims [255] to identify relationships representing attacks. Due to the maturity of the techniques based on the theory of graphs, this survey hypothesizes the possibility of proposing solutions for predicting DDoS attacks based on graphs. For example, graph-based solutions can identify C&C communications.

Protecting users' privacy should be an inherent concern of all DDoS attack prediction solutions. The General Data Protection Regulation (GDPR) [256] is valid for the countries of the European Union. The state of California in the USA utilizes the California Consumer Privacy Act (CCPA) (CCPA) [257]. Furthermore, the General Data Protection Law (LGPD) [258], valid in Brazil are laws that guarantee the privacy of users. For example, network traffic is the data aspect most utilized by the solutions identified in this work. These solutions must be concerned with handling network traffic. Given that, if data related to network traffic becomes public, the solution may infringe on the privacy of network users. Therefore, less invasive solutions are desirable and possible [259]. It is more prudent to propose solutions that avoid manipulation and storing payload of network packets.

5. Conclusion

Attack prediction is a proactive defense mechanism attracting attention in the literature. Prediction techniques highlight evidence of a DDoS attack that has not happened yet. It aims at providing enough time for administrators to deal with the attack. This survey presented the state-of-the-art in predicting DDoS attacks. It shortlisted the identified studies that propose solutions to DDoS attack prediction. Based on these studies, this survey article presented a classification following the main aspects of the solutions, such as time, architecture, methodology, and data types. Furthermore, it highlighted important research open issues, such as the instantiation of solutions for different environments, such as IoT and SDN, the reduction of dependence on labeled data for model training, the application of deep learning techniques, and the proposal of new studies to carry out long-term prediction with a low error rate. Finally, this survey emphasized that designing distributed solutions is still an open issue to evolve prediction techniques and combat DDoS attacks.

The oldest study highlighted in this survey is from 2009; this demonstrates that the concern with DDoS attack prediction is a relatively recent action. In this survey, 66% of the identified studies have a publication date in the last six years, between 2017 and 2022. The publications peaked between 2017 and 2018, with five studies each year. Therefore, the research field is beginning, but it has increasingly attracted researchers. Hence, the terms prediction, early detection, and detection must be utilized with caution to avoid misunderstandings. The machine learning field commonly employs prediction in two circumstances: (*i*) the prediction can classify an event that has not happened yet; (*ii*) classifying an event that has already happened. This survey identified studies utilizing the term prediction to refer to works that identify attacks after their effective start. Utilizing the term prediction in these last cases follows the machine learning field concept of prediction. However, this survey did not include these studies.

This work follows the idea that prediction identifies DDoS attack evidence before the attacker launches it. Although the literature presents several studies for DDoS attack detection, the same does not happen with DDoS attack prediction. Of the 2482 analyzed studies, there are only 27 studies that perform prediction. Proposing solutions for predicting DDoS attacks is a challenging task, but it is necessary. The interest and necessity for DDoS attack prediction will grow in the coming years, contributing to the cybersecurity field. The first reason for this potential growth is attack prediction's advantage to network administrators over attackers. Predicting attacks is one of the few ways to allow network administrators not to be surprised by attackers. Consequently, network administrators can have more time to deal with the attack, acting to reduce the damage caused by the attackers. Although predicting DDoS attacks is not trivial, the literature presents the feasibility of predicting attacks. Despite the difficulty in developing solutions to predict DDoS attacks, research is paramount that there is an investment in this area. The second reason DDoS attack prediction will evolve is many research opportunities.

Although the literature has already evaluated the classical methods for attack prediction, there are many opportunities for new contributions. DDoS attacks constantly evolve, requiring up-to-date studies to collaborate with the literature, presenting new and better ways to predict attacks. Future works would follow this principle, exploring open issues such as the adoption of distributed systems and the utilization of deep learning to evaluate a solution that provides more time for network administrators to deal with DDoS attacks.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Anderson Bergamini de Neira reports financial support and article publishing charges were provided by State of Sao Paulo Research Foundation. Anderson Bergamini de Neira reports financial support was provided by National Council for Scientific and Technological Development. Anderson Bergamini de Neira reports financial support was provided by Coordination of Higher Education Personnel Improvement.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by National Council for Scientific and Technological Development (CNPq/Brazil), grants #309129/2017-6 and #432204/2018-0, by the Coordination for the Improvement of Higher Education Personnel (CAPES/Brazil), grant #88887.509309/2020-00, by São Paulo Research Foundation (FAPESP), grant #2018/23098-0.

References

[1] A. Bendovschi, Cyber-attacks – trends, patterns and security countermeasures, ICFC 28 (2015) 24–31.
[2] J.M. Biju, N. Gopal, A.J. Prakash, Cyber attacks and its different types, IRJET 6 (3) (2019) 4849–4852.
[3] Z. He, T. Zhang, R.B. Lee, Machine learning based DDoS attack detection from source side in cloud, in: CSCloud, IEEE, USA, 2017, pp. 114–120.
[4] T.-K. Luong, T.-D. Tran, G.-T. Le, DDoS attack detection and defense in based on machine learning, in: NICS, IEEE, Ho Chi Minh City, Vietnam, 2020, pp. 31–35.
[5] N. Jyoti, S. Behal, A meta-evaluation of machine learning techniques for detection of DDoS attacks, in: INDIACom, IEEE, New Delhi, India, 2021, pp. 522–526.
[6] R. Hummel, C. Hildebrand, H. Modi, C. Conrad, S.B. Roland Dobbins, J. Belanger, G. Sockrider, P. Alcoy, T. Bienkowski, Netscout Threat Intelligence Report, Netscout, 2021, [Online]. Available: https://www.netscout.com/sites/default/files/2021-04/ThreatReport_2H2020_FINAL_0.pdf. (Accessed in: 08/2021).
[7] A. Gutnikov, O. Kupreev, E. Badovskaya, DDoS Attacks in Q1 2021, Kaspersky, 2021, [Online]. Available: https://securelist.com/ddos-attacks-in-q1-2021/102166. (Accessed in: 07/2021).

[8] A. Gutnikov, E. Badovskaya, O. Kupreev, Y. Shmelev, DDoS Attacks in Q2 2021, Kaspersky, 2021, [Online]. Available: https://securelist.com/ddos-attacks-in-q2-2021/103424. (Accessed in: 12/2022).

[9] A. Gutnikov, O. Kupreev, Y. Shmelev, DDoS Attacks in Q3 2021, Kaspersky, 2021, [Online]. Available: https://securelist.com/ddos-attacks-in-q3-2021/104796. (Accessed in: 12/2022).

[10] A. Gutnikov, O. Kupreev, Y. Shmelev, DDoS attacks in Q4 2021, Kaspersky, 2022, [Online]. Available: https://securelist.com/ddos-attacks-in-q4-2021/105784. (Accessed in: 12/2022).

[11] Netscout, Issue 9: Findings from 1st Half 2022, Netscout, 2022, www.netscout.com/threatreport/global-highlights. (Acesso em: 11/22).

[12] D. Stress, DDoS-as-aService, DDoS Stress, 2021, [Online]. Available: https://ddos.services. (Accessed in: 07/2021).

[13] A.D. Inc, Armor's 'Black Market' Report Highlights The Big Business of Cybercrime, Armor Defense Inc, 2018, [Online]. Available: https://res.armor.com/resources/press-release/black-market-report-highlights-cybercrime. (Accessed in: 07/2021).

[14] T. Mahjabin, Y. Xiao, G. Sun, W. Jiang, A survey of distributed denial-of-service attack, prevention, and mitigation techniques, Int. J. Distrib. Sens. Netw. 13 (12) (2017) 1550147717741463.

[15] Cybersecurity, I.S.A. (CISA), Understanding denial-of-service attacks, 2019, [Online]. Available: https://us-cert.cisa.gov/ncas/tips/ST04-015. (Accessed in: 07/2021).

[16] T. Miu, R. Yeung, K. Cheung, D. Li, DDoS threat report 2020 Q3, Nexusguard, 2020, [Online]. Available: https://blog.nexusguard.com/threat-report/ddos-threat-report-2020-q3. (Accessed in: 07/2021).

[17] J. Vijayan, DDoS Attacks Spiked, Became More Complex in 2020, Information Week IT Network, 2020, [Online]. Available: https://www.darkreading.com/attacks-breaches/ddos-attacks-spiked-became-more-complex-in-2020/d/d-id/1339814. (Accessed in: 07/2021).

[18] B. Deas, DDoS cyberattacks on Bradford Council's school network caused 'chaos', The telegraph and argus, 2021, [Online]. Available: https://www.thetelegraphandargus.co.uk/news/19205512.ddos-cyberattacks-bradford-councils-school-network-caused-chaos/. (Accessed in: 07/2021).

[19] C. to protect journalists, Investigative outlet Repórter Brasil targeted with cyberattacks, threats, attempted break-in, Committee to protect journalists, 2021, https://cpj.org/2021/01/investigative-outlet-reporter-brasil-targeted-with-cyberattacks-threats-attempted-break-in/. (Accessed in: 07/2021).

[20] BBC, New Zealand stock exchange halted by cyber-attack, BBC, 2020, [Online]. Available: https://www.bbc.com/news/53918580. (Accessed in: 07/2021).

[21] A.A. Santos, M. Nogueira, J.M.F. Moura, A stochastic adaptive model to explore mobile botnet dynamics, IEEE Commun. Lett. 21 (4) (2017) 753–756.

[22] A. Lerner, The Cost of Downtime, Gartner, Inc, 2014, [Online]. Available: https://blogs.gartner.com/andrew-lerner/2014/07/16/the-cost-of-downtime. (Accessed in: 08/2021).

[23] I.T.I. Consulting, ITIC 2020 Global Server Hardware, Server OS Reliability Report, Information Technology Intelligence Consulting, 2020, [Online]. Available: https://www.ibm.com/downloads/cas/DV0XZV6R. (Accessed in: 08/2021).

[24] B.B. Gupta, R.C. Joshi, M. Misra, Distributed denial of service prevention techniques, Int. J. Electr. Comput. Eng. 2 (2) (2010) 268–276.

[25] S.T. Zargar, J. Joshi, D. Tipper, A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks, IEEE Commun. Surv. Tutor. 15 (4) (2013) 2046–2069.

[26] B.B. Gupta, A. Dahiya, Distributed Denial of Service (DDoS) Attacks: Classification, Attacks, Challenges, and Countermeasures, CRC Press, USA, 2021.

[27] I. Ilascu, Extortion DDoS Attacks Grow Stronger and More Common, Bleeping Computer, 2022, https://www.bleepingcomputer.com/news/security/extortion-ddos-attacks-grow-stronger-and-more-common/. (Accessed in: 01/2022).

[28] D. Menscher, Exponential Growth in DDoS Attack Volumes, Google, 2020, [Online]. Available: https://cloud.google.com/blog/products/identity-security/identifying-and-protecting-against-the-largest-ddos-attacks. (Accessed in: 07/2021).

[29] A. Shield, Threat Landscape Report – Q1 2020, Amazon.com, Inc., 2020, [Online]. Available: https://aws-shield-tlr.s3.amazonaws.com/2020-Q1_AWS_Shield_TLR.pdf. (Accessed in: 08/2021).

[30] O. Yoachimik, Cloudflare thwarts 17.2m rps DDoS attack — the largest ever reported, Cloudflare, Inc., 2021, [Online]. Available: https://blog.cloudflare.com/cloudflare-thwarts-17-2m-rps-ddos-attack-the-largest-ever-reported/. (Accessed in: 08/2021).

[31] A. Marrow, G. Stolyarov, Russia's Yandex Says It Repelled Biggest DDoS Attack in History, Reuters, News agency company, 2021, http://www.reuters.com/technology/russias-yandex-says-it-repelled-biggest-ddos-attack-history-2021-09-09. (Accessed in: 10/2021).

[32] A. Toh, A. Vij, S. Pasha, Azure DDoS Protection—2021 Q3 and Q4 DDoS attack trends, Microsoft, 2022, https://www.azure.microsoft.com/en-us/blog/azure-ddos-protection-2021-q3-and-q4-ddos-attack-trends/. (Accessed in: 01/2022).

[33] B.B. Gupta, O.P. Badve, Taxonomy of DoS and DDoS attacks and desirable defense mechanism in a cloud computing environment, Neural. Comput. Appl. 28 (12) (2017) 3655–3682.

[34] M. Abdlhamed, K. Kifayat, Q. Shi, W. Hurst, Intrusion prediction systems, in: Information Fusion for Cyber-Security Analytics, Springer, Cham, 2017, pp. 155–174.

[35] M. Pelloso, A. Vergutz, A. Santos, M. Nogueira, A self-adaptable system for DDoS attack prediction based on the metastability theory, in: GLOBECOM, IEEE, UAE, 2018, pp. 1–6.

[36] M. Husák, J. Komárková, E. Bou-Harb, P. Čeleda, Survey of attack projection, prediction, and forecasting in cyber security, IEEE Commun. Surv. Tutor. 21 (1) (2019) 640–660.

[37] A. Avizienis, J.-C. Laprie, B. Randell, C. Landwehr, Basic concepts and taxonomy of dependable and secure computing, IEEE Trans. Dependable Secure Comput. 1 (1) (2004) 11–33.

[38] N. Gruschka, N. Luttenberger, Protecting web services from DoS attacks by SOAP message validation, in: S. Fischer-Hübner, K. Rannenberg, L. Yngström, S. Lindskog (Eds.), IFIP, Springer, USA, 2006, pp. 171–182.

[39] A. Wood, J. Stankovic, Denial of service in sensor networks, Computer 35 (10) (2002) 54–62.

[40] V.D. Gligor, A note on denial-of-service in operating systems, IEEE Trans. Softw. Eng. SE-10 (3) (1984) 320–324.

[41] D.E. Comer, Computer Networks and Internets, Pearson Education, USA, 2015.

[42] S.-C. Lin, S.-S. Tseng, Constructing detection knowledge for DDoS intrusion tolerance, Expert Syst. Appl. 27 (3) (2004) 379–390.

[43] R.V. Deshmukh, K.K. Devadkar, Understanding DDoS attack & its effect in cloud environment, ICAC3 49 (2015) 202–210.

[44] W. Zhijun, L. Wenjing, L. Liang, Y. Meng, Low-rate DoS attacks, detection, defense, and challenges: A survey, IEEE Access 8 (2020) 43920–43943.

[45] K. Lee, J. Kim, K.H. Kwon, Y. Han, S. Kim, DDoS attack detection method using cluster analysis, Expert Syst. Appl. 34 (3) (2008) 1659–1665.

[46] S. Bhatia, S. Behal, I. Ahmed, Distributed denial of service attacks and defense mechanisms: Current landscape and future directions, in: Versatile Cybersecurity, Springer, Cham, 2018, pp. 55–97.

[47] P.J. Beslin Pajila, E. Golden Julie, Detection of DDoS attack using SDN in IoT: A survey, in: S. Balaji, A. Rocha, Y.-N. Chung (Eds.), ICICV, Springer, Cham, 2020, pp. 438–452.

[48] T. Peng, C. Leckie, K. Ramamohanarao, Survey of network-based defense mechanisms countering the DoS and DDoS problems, ACM Comput. Surv. 39 (1) (2007) 3–es.

[49] A. Keshariya, N. Foukia, DDoS defense mechanisms: A new taxonomy, in: J. Garcia-Alfaro, G. Navarro-Arribas, N. Cuppens-Boulahia, Y. Roudier (Eds.), Data Privacy Management and Autonomous Spontaneous Security, Springer, Berlin, Heidelberg, 2010, pp. 222–236.

[50] C. Douligeris, A. Mitrokotsa, DDoS attacks and defense mechanisms: Classification and state-of-the-art, Comput. Netw. 44 (5) (2004) 643–666.

[51] R. ur Rasool, H. Wang, U. Ashraf, K. Ahmed, Z. Anwar, W. Rafique, A survey of link flooding attacks in software defined network ecosystems, J. Netw. Comput. Appl. 172 (2020) 102803.

[52] J. Xing, W. Wu, A. Chen, Ripple: A programmable, decentralized link-flooding defense against adaptive adversaries, in: USENIX, USENIX Association, USA, 2021, p. 16.

[53] M. Monge, J. Vidal, L. Villalba, Entropy-based economic denial of sustainability detection, Entropy 19 (12) (2017) 649.

[54] N. Vlajic, A. Slopek, Web bugs in the cloud: Feasibility study of a new form of EDoS attack, in: GLOBECOM, IEEE, USA, 2014, pp. 64–69.

[55] P. Singh, S. Manickam, S.U. Rehman, A survey of mitigation techniques against Economic Denial of Sustainability (EDoS) attack on cloud computing architecture, in: ICRITO, IEEE, India, 2014, pp. 1–4.

[56] S. Alarifi, S.D. Wolthusen, Robust coordination of cloud-internal denial of service attacks, in: CGC, IEEE, Germany, 2013, pp. 135–142.

[57] J.S. Ribin, N. Kumar, Precursory study on varieties of DDoS attacks and its implications in cloud systems, in: ICOEI, IEEE, India, 2019, pp. 1003–1008.

[58] Cloudflare, What is a Zero-Day Exploit? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/ransom-ddos-attack/. (Accessed in: 08/2021).

[59] T. Sasaki, C.H. Gañán, K. Yoshioka, M.V. Eeten, T. Matsumoto, Pay the piper: DDoS mitigation technique to deter financially-motivated attackers, IEICE Trans. Commun. E103.B (4) (2020) 389–404.

[60] F.T. Ngo, A. Agarwal, R. Govindu, C. MacDonald, Malicious software threats, in: The Palgrave Handbook of International Cybercrime and Cyberdeviance, Springer, Cham, 2020, pp. 793–813.

[61] P. Amini, M.A. Araghizadeh, R. Azmi, A survey on Botnet: Classification, detection and defense, in: IES, IEEE, Indonesia, 2015, pp. 233–238.

[62] H. Choi, H. Lee, Identifying botnets by capturing group activities in DNS traffic, Comput. Netw. 56 (1) (2012) 20–33.

[63] D.Y. Huang, H. Dharmdasani, S. Meiklejohn, V. Dave, C. Grier, D. McCoy, S. Savage, N. Weaver, A.C. Snoeren, K. Levchenko, Botcoin: Monetizing stolen cycles, in: NDSS, The Internet Society, USA, 2014, p. 16.

[64] Y.D. Mane, Detect and deactivate P2P Zeus bot, in: ICCCNT, IEEE, Delhi, India, 2017, pp. 1–7.

[65] M.M. Salim, S. Rathore, J.H. Park, Distributed denial of service attacks and its defenses in IoT: a survey, J. Supercomput. 76 (7) (2020) 5320–5363.

[66] L.G. Wlosinski, Cybersecurity takedowns, ISACA J. 6 (2019).

[67] A. Marzano, D. Alexander, O. Fonseca, E. Fazzion, C. Hoepers, K. Steding-Jessen, M.H.P.C. Chaves, I. Cunha, D. Guedes, W. Meira, The evolution of bashlite and Mirai IoT botnets, in: ISCC, IEEE, 2018.

[68] D. il Jang, M. Kim, H. chul Jung, B.-N. Noh, Analysis of HTTP2P botnet: case study waledac, in: MICC, IEEE, Kuala Lumpur, Malaysia, 2009, pp. 409–412.

[69] A. Karim, R.B. Salleh, M. Shiraz, S.A.A. Shah, I. Awan, N.B. Anuar, Botnet detection techniques: review, future trends, and issues, J. Zhejiang Univ. Sci. C 15 (11) (2014) 943–983.

[70] H.R. Zeidanloo, A.B.A. Manaf, Botnet detection by monitoring similar communication patterns, IJCSIS 7 (3) (2010) 10.

[71] P. Wang, S. Sparks, C.C. Zou, An advanced hybrid peer-to-peer botnet, IEEE Trans. Dependable Sec. Comput. 7 (2) (2010) 113–127.

[72] G. Vormayr, T. Zseby, J. Fabini, Botnet communication patterns, IEEE Commun. Surv. Tutor. 19 (4) (2017) 2768–2796.

[73] P. Musik, K. Jaroensutasinee, Large-scale simulation using parallel computing toolkit and server message block, WSEAS Trans. Math. 6 (2) (2007) 369–372.

[74] I. Cisco Systems, A Cisco Guide to Defending Against Distributed Denial of Service Attacks, Cisco Systems, 2014, [Online]. Available: https://tools.cisco.com/security/center/resources/guide_ddos_defense.html. (Accessed in: 07/2021),

[75] S.K. Sahu, R. Khare, DDoS attacks & mitigation techniques in cloud computing environments, Gedrag Org. Rev. 33 (2) (2020) 2426–2435.

[76] J. Yadav, J. Thakur, Botnet: Evolution life cycle architecture and detection techniques, Mukt Shabd J. 9 (6) (2020) 4265–4281.

[77] S. App, DDoS-as-aService, Stresser App, 2021, [Online]. Available: https://stresser.app/. (Accessed in: 07/2021).

[78] J.J. Santanna, R. van Rijswijk-Deij, R. Hofstede, A. Sperotto, M. Wierbosch, L.Z. Granville, A. Pras, Booters — An analysis of DDoS-as-a-service attacks, in: IM, IEEE, Ottawa, ON, Canada, 2015, pp. 243–251.

[79] L. Visalatchi, P. Yazhini, M.P. Scholar, The survey DDoS attack prevention and defense technique, IJISRT 5 (2) (2020).

[80] H.F. El-Sofany, A new cybersecurity approach for protecting cloud services against DDoS attacks, IJISAE 13 (2) (2020) 205–215.

[81] Cloudflare, What is a DDoS attack? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack. (Accessed in: 07/2021).

[82] M.M. Alam, M.Y. Arafat, F. Ahmed, Study on auto detecting defence mechanisms against application layer DDoS attacks in SIP server, J. Netw. 10 (6) (2015) 344–352.

[83] K.S. Sahoo, S.K. Panda, S. Sahoo, B. Sahoo, R. Dash, Toward secure software-defined networks against distributed denial of service attack, J. Supercomput. 75 (8) (2019) 4829–4874.

[84] J. Mirkovic, P. Reiher, A taxonomy of DDoS attack and DDoS defense mechanisms, SIGCOMM Comput. Commun. Rev. 34 (2) (2004) 39–53.

[85] A. Asosheh, N. Ramezani, A comprehensive taxonomy of DDoS attacks and defense mechanism applying in a smart classification, WSEAS Trans. Comput. 7 (4) (2008) 281–290.

[86] A.R. Yusof, N.I. Udzir, A. Selamat, Systematic literature review and taxonomy for DDoS attack detection and prediction, IJDET 1 (3) (2019) 292–315.

[87] S.M. Specht, R.B. Lee, Distributed denial of service: Taxonomies of attacks, tools, and countermeasures, in: ICPADS, ISCA, USA, 2004, pp. 543–550.

[88] Cloudflare, What is a UDP flood attack? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/udp-flood-ddos-attack/. (Accessed in: 07/2021).

[89] Cloudflare, What is a Ping (ICMP) flood attack? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/ping-icmp-flood-ddos-attack. (Accessed in: 07/2021).

[90] Cloudflare, What is a DNS amplification attack? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/dns-amplification-ddos-attack/. (Accessed in: 07/2021).

[91] Cloudflare, What is a NTP amplification attack? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/ntp-amplification-ddos-attack/. (Accessed in: 07/2021).

[92] J. Arteaga, W. Mejia, CLDAP Reflection DDoS, Akamai Technologies, Inc, 2017, https://www.akamai.com/us/en/resources/our-thinking/threat-advisories/connection-less-lightweight-directory-access-protocol-reflection-ddos-threat-advisory.jsp. (Accessed in: 07/2021).

[93] Cloudflare, What is a SYN Flood Attack? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/syn-flood-ddos-attack/. (Accessed in: 07/2021).

[94] Cloudflare, What is an HTTP Flood DDoS attack? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/http-flood-ddos-attack/. (Accessed in: 07/2021).

[95] Cloudflare, What is a Low and Slow Attack? Cloudflare, Inc., 2021, [Online]. Available: https://www.cloudflare.com/learning/ddos/ddos-low-and-slow-attack/. (Accessed in: 07/2021).

[96] R. Samta, M. Sood, Analysis and mitigation of DDoS flooding attacks in software defined networks, in: ICICC, Springer, Singapore, 2020, pp. 337–355.

[97] Radware, Teardrop Attack, Radware Ltd, 2021, [Online]. Available: https://www.radware.com/security/ddos-knowledge-center/ddospedia/teardrop-attack. (Accessed in: 07/2021).

[98] C. Williams, Today the Web Was Broken by Countless Hacked Devices – Your 60-Second Summary, The Register, 2016, [Online]. Available: https://www.theregister.com/2016/10/21/dyn_dns_ddos_explained. (Accessed in: 06/2021).

[99] E. Fenil, P. Mohan Kumar, Survey on DDoS defense mechanisms, CCPE 32 (4) (2020) e5114.

[100] K. Sonar, H. Upadhyay, A survey: DDoS attack on Internet of Things, IJERD 10 (11) (2014) 58–63.

[101] Facebook, Facebook Bug Bounty Program Info, Facebook, Inc., 2021, [Online]. Available: https://www.facebook.com/whitehat. (Accessed in: 07/2021).

[102] Google, Google Vulnerability Reward Program (VRP) Rules, Google, LLC., 2021, [Online]. Available: https://www.google.com/about/appsecurity/reward-program/. (Accessed in: 07/2021).

[103] Apple, Apple Security Bounty, Apple Inc., 2021, [Online]. Available: https://developer.apple.com/security-bounty/. (Accessed in: 07/2021).

[104] O. Yoachimik, A. Forster, CVE-2022-26143: A Zero-Day vulnerability for Launching UDP Amplification DDoS Attacks? Cloudflare, Inc., 2022, [Online]. Available: https://blog.cloudflare.com/cve-2022-26143-amplification-attack/. (Accessed in: 05/2022).

[105] J.K. Chahal, A. Bhandari, S. Behal, Distributed denial of service attacks: A threat or challenge, New Rev. Inf. Netw. 24 (1) (2019) 31–103.

[106] K. Srinivasan, A. Mubarakali, A.S. Alqahtani, A. Dinesh Kumar, A survey on the impact of DDoS attacks in cloud computing: Prevention, detection and mitigation techniques, in: S. Balaji, A. Rocha, Y.-N. Chung (Eds.), ICICV, Springer, India, 2020, pp. 252–270.

[107] D. Radain, S. Almalki, H. Alsaadi, S. Salama, A review on defense mechanisms against distributed denial of service (DDoS) attacks on cloud computing, in: WiDSTaif, IEEE, Saudi Arabia, 2021, pp. 1–6.

[108] G. Somani, M.S. Gaur, D. Sanghi, M. Conti, R. Buyya, DDoS attacks in cloud computing: Issues, taxonomy, and future directions, Comput. Commun. 107 (2017) 30–48.

[109] B.L. Dalmazo, J.A. Marques, L.R. Costa, M.S. Bonfim, R.N. Carvalho, A.S. da Silva, S. Fernandes, J.L. Bordim, E. Alchieri, A. Schaeffer-Filho, L. Paschoal Gaspary, W. Cordeiro, A systematic review on distributed denial of service attack defense mechanisms in programmable networks, Int. J. Enterp. Netw. Manag. (2021) e2163.

[110] D. Brodić, A. Amelio, Types of CAPTCHA, in: The CAPTCHA: Perspectives and Challenges: Perspectives and Challenges in Artificial Intelligence, Springer, Cham, 2020, pp. 29–32.

[111] M. Nooribakhsh, M. Mollamotalebi, A review on statistical approaches for anomaly detection in DDoS attacks, Inf. Secur. J. 29 (3) (2020) 118–133.

[112] A. Belenky, N. Ansari, On IP traceback, IEEE Commun. Mag. 41 (7) (2003) 142–153.

[113] F.J. Ryba, M. Orlinski, M. Wählisch, C. Rossow, T.C. Schmidt, Amplification and DRDoS attack defense – a survey and new perspectives, 2015.

[114] O.N. Foundation, Software-Defined Networking (SDN) Definition: 07/2021, Open Networking Foundation, 2021, [Online]. Available: https://opennetworking.org/sdn-definition/.

[115] Y. Al-Hadhrami, F.K. Hussain, DDoS attacks in IoT networks: a comprehensive systematic literature review, World Wide Web 24 (3) (2021) 971–1001.

[116] L.F. Eliyan, R. Di Pietro, DoS and DDoS attacks in Software Defined Networks: A survey of existing solutions and research challenges, Future Gener. Comput. Syst. 122 (2021) 149–171.

[117] A. Bhardwaj, V. Mangat, R. Vig, S. Halder, M. Conti, Distributed denial of service attacks in cloud: State-of-the-art of scientific and commercial solutions, Comp. Sci. Rev. 39 (2021) 100332.

[118] N. Tripathi, N. Hubballi, Application layer denial-of-service attacks and defense mechanisms: A survey, ACM Comput. Surv. 54 (4) (2021).

[119] V.D.M. Rios, P.R.M. Inácio, D. Magoni, M.M. Freire, Detection and mitigation of low-rate denial-of-service attacks: A survey, IEEE Access 10 (2022) 76648–76668, http://dx.doi.org/10.1109/ACCESS.2022.3191430.

[120] R.A. Salam, J.K. Das, Z.S. Lassi, Z.A. Bhutta, Adolescent health and well-being: Background and methodology for review of potential interventions, J. Adolesc Health 59 (4, Supplement) (2016) S4–S10, Interventions to Address Adolescent Health and Well-Being: Current State of the Evidence.

[121] L. Laing, N. Salema, M. Jeffries, A. Shamsuddin, A. Sheikh, J. Waring, T. Avery, R. Keers, Understanding the implementation and medium-longer term sustainability of the primary care prescribing safety intervention, PINCER: preliminary results from a longitudinal process evaluation, IJPP 29 (Supplement_1) (2021) i8–i9.

[122] M. Abu Rajab, J. Zarfoss, F. Monrose, A. Terzis, A multifaceted approach to understanding the botnet phenomenon, in: IMC, IMC '06, ACM, USA, 2006, pp. 41–52.

[123] G.-L. Dei Rossi, M. Iacono, A. Marin, Evaluating the impact of EDoS attacks to cloud facilities, in: VALUETOOLS, ICST, BEL, 2016, pp. 188–195.

[124] M. Nogueira, A.A. Santos, J.M.F. Moura, Non-parametric early warning signals from volumetric DDoS attacks, 2016, CoRR arXiv:1609.09560.

[125] A. Muhammad, M. Asad, A.R. Javed, Robust early stage botnet detection using machine learning, in: ICCWS, IEEE, Pakistan, 2020, pp. 1–6.

[126] S. Chen, Z. Wu, P.D. Christofides, Cyber-security of centralized, decentralized, and distributed control-detector architectures for nonlinear processes, Chem. Eng. Res. Des. 165 (2021) 25–39.

[127] M. Jog, M. Natu, S. Shelke, Distributed and predictive-preventive defense against DDoS attacks, in: ICDCN, ACM, USA, 2015.

[128] T. Barnett, S. Jain, U. Andra, T. Khurana, Cisco Visual Networking Index (VNI), Complete Forecast Update, 2017–2022, (125) Cisco Systems, 2018, [Online]. Available: https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1211_BUSINESS_SERVICES_CKN_PDF.pdf. (Accessed in: 06/2021),

[129] P.P. Shinde, S. Shah, A review of machine learning and deep learning applications, in: ICCUBEA, IEEE, India, 2018, pp. 1–6.

[130] H. Liu, B. Lang, Machine learning and deep learning methods for intrusion detection systems: A survey, Appl. Sci. 9 (20) (2019).

[131] I.C. Education, AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference? International Business Machines Corporation, 2020, [Online]. Available: https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks. (Accessed in: 08/2021).

[132] T. Kaluarachchi, A. Reis, S. Nanayakkara, A review of recent deep learning approaches in human-centered machine learning, Sensors 21 (7) (2021).

[133] J.G. Kemeny, J.L. Snell, Finite Markov Chains, Springer-Verlag, USA, 1976.

[134] S.S. Silva, R.M. Silva, R.C. Pinto, R.M. Salles, Botnets: A survey, Comput. Netw. 57 (2) (2013) 378–403, Botnet Activity: Analysis, Detection and Shutdown.

[135] A. Wang, A. Mohaisen, S. Chen, An adversary-centric behavior modeling of DDoS attacks, in: ICDCS, IEEE, USA, 2017, pp. 1126–1136.

[136] M.Q. Ali, E. Al-Shaer, Configuration-based IDS for advanced metering infrastructure, in: SIGSAC, ACM, USA, 2013, pp. 451–462.

[137] H. Moudoud, L. Khoukhi, S. Cherkaoui, Prediction and detection of FDIA and DDoS attacks in 5G enabled IoT, IEEE Netw. 35 (2) (2021) 194–201.

[138] H. Salemi, H. Rostami, S. Talatian-Azad, M.R. Khosravi, LEAESN: Predicting DDoS attack in healthcare systems based on Lyapunov Exponent Analysis and Echo State Neural Networks, Multimedia Tools Appl. - (-) (2021) 1–22.

[139] L. Laboratory, 1998 DARPA Intrusion Detection Evaluation Dataset, Massachusetts Institute of Technology, 1998, [Online]. Available: https://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-dataset. (Accessed in: 06/2022).

[140] A. Olabelurin, S. Veluru, A. Healing, M. Rajarajan, Entropy clustering approach for improving forecasting in DDoS attacks, in: ICNSC, IEEE, Taiwan, 2015, pp. 315–320.

[141] J. Wu, Cluster analysis and K-means clustering: An introduction, in: Advances in K-Means Clustering: A Data Mining Thinking, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 1–16.

[142] Z.M. Fadlullah, M.M. Fouda, N. Kato, X. Shen, Y. Nozaki, An early warning system against malicious activities for smart grid communications, IEEE Netw. 25 (5) (2011) 50–55.

[143] Scikit-learn, Gaussian Processes, Scikit-learn, 2021, [Online]. Available: https://scikit-learn.org/stable/modules/gaussian_process.html. (Accessed in: 08/2021).

[144] A. Jaber, M. Zolkipli, M. Majid, S. Anwar, Methods for preventing distributed denial of service attacks in cloud computing, Adv. Sci. Lett. 23 (2017) 5282–5285.

[145] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1) (1987) 37–52.

[146] J.W. Grzymala-Busse, LERS-a system for learning from examples based on rough sets, in: Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, Springer Netherlands, Dordrecht, 1992, pp. 3–18.

[147] S. Hettich, S.D. Bay, KDD Cup 1999 Data, University of California, 1999, [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. (Accessed in: 06/2022).

[148] D. Kwon, H. Kim, D. An, H. Ju, DDoS attack volume forecasting using a statistical approach, in: IM, IEEE, Portugal, 2017, pp. 1083–1086.

[149] S. Ray, A quick review of machine learning algorithms, in: COMITCon, IEEE, India, 2019, pp. 35–39.

[150] P. Machaka, O. Ajayi, H. Maluleke, F. Kahenga, A. Bagula, K. Kyamakya, Modelling DDoS attacks in IoT networks using machine learning, 2021.

[151] V. Kecman, Support vector machines – an introduction, in: Support Vector Machines: Theory and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 1–47.

[152] L. Laboratory, 1999 DARPA intrusion detection evaluation dataset, Massachusetts Institute of Technology, 1999, [Online]. Available: https://www.ll.mit.edu/r-d/datasets/1999-darpa-intrusion-detection-evaluation-dataset. (Accessed in: 06/2022).

[153] Y. SU, X. MENG, Q. MENG, X. HAN, DDoS attack detection algorithm based on hybrid traffic prediction model, in: ICSPCC, IEEE, China, 2018, pp. 1–5.

[154] J. Liu, RBF neural network design and simulation, in: Radial Basis Function (RBF) Neural Network Control for Mechanical Systems: Design, Analysis and Matlab Simulation, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 19–53.

[155] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[156] D.G. Kleinbaum, M. Klein, Introduction to logistic regression, in: Logistic Regression: A Self-Learning Text, Springer, USA, 2010, pp. 1–39.

[157] H. Taud, J. Mas, Multilayer perceptron (MLP), in: Geomatic Approaches for Modeling Land Change Scenarios, Springer, Cham, 2018, pp. 451–455.

[158] Z. Abaid, D. Sarkar, M.A. Kaafar, S. Jha, The early bird gets the botnet: A Markov chain based early warning system for botnet attacks, in: LCN, IEEE, UAE, 2016, pp. 61–68.

[159] P. Holgado, V.A. Villagrá, L. Vázquez, Real-time multistep attack prediction based on hidden Markov models, IEEE Trans. Dependable Secure Comput. 17 (1) (2020) 134–147.

[160] L. Laboratory, 2000 DARPA Intrusion Detection Scenario Specific Datasets, Massachusetts Institute of Technology, 2000, [Online]. Available: https://www.ll.mit.edu/r-d/datasets/2000-darpa-intrusion-detection-scenario-specific-datasets. (Accessed in: 06/2021).

[161] S. Shin, S. Lee, H. Kim, S. Kim, Advanced probabilistic approach for network intrusion forecasting and detection, Expert Syst. Appl. 40 (1) (2013) 315–322.

[162] A. Adegboyega, An adaptive score model for effective bandwidth prediction and provisioning in the cloud network, in: GLOBECOM, IEEE, USA, 2015, pp. 1–7.

[163] Z. Fan, Z. Tan, C. Tan, X. Li, An improved integrated prediction method of cyber security situation based on spatial-time analysis, JIT 19 (2018) 1789–1800.

[164] A.P. Leros, A.S. Andreatos, Network traffic analytics for Internet service providers—Application in early prediction of DDoS attacks, in: Machine Learning Paradigms: Advances in Data Analytics, Springer, Cham, 2019, pp. 233–267.

[165] S. Garcia, M. Grill, J. Stiborek, A. Zunino, An empirical comparison of botnet detection methods, C&S 45 (2014) 100–123.

[166] V. Savchenko, O. Ilin, N. Hnidenko, O. Tkachenko, O. Laptiev, S. Lehominova, Detection of slow DDoS attacks based on user's behavior forecasting, IJETER 8 (5) (2020).

[167] K. Yin, T. Nianqing, Study on the risk detection about network security based on grey theory, in: IFITA, 1, IEEE, China, 2009, pp. 411–413.

[168] S. Kivalov, I. Strelkovskaya, Detection and prediction of DDoS cyber attacks using spline functions, in: TCSET, Ukraine, 2022, pp. 710–713.

[169] I. Strelkovskaya, I. Solovskaya, Using spline-extrapolation in the research of self-similar traffic characteristics, JEE 70 (4) (2019) 310–316.

[170] Y. Liu, J. Zhang, A. Sarabi, M. Liu, M. Karir, M. Bailey, Predicting cyber security incidents using feature-based characterization of network-level malicious activities, in: IWSPA, ACM, USA, 2015, pp. 3–9.

[171] P. Passeri, Hackmageddon, Hackmageddon, 2022, [Online]. Available: www.hackmageddon.com.

[172] Z. Wang, Y. Zhang, DDoS event forecasting using Twitter data, in: IJCAI, AAAI Press, Australia, 2017, pp. 4151–4157.

[173] S. Anuar, N.A. Ahmad, S. Sahibuddin, A. Ariffin, A. Saupi, N.A. Zamani, Y. Jeffry, F. Efendy, Modeling malware prediction using artificial neural network, in: H. Fujita, E. Herrera-Viedma (Eds.), SoMeT, 303, IOS Press, SPAIN, 2018, pp. 240–248.

[174] S. Shanmuganathan, Artificial neural network modelling: An introduction, in: Artificial Neural Network Modelling, Springer, Cham, 2016, pp. 1–14.

[175] A. Sapienza, S.K. Ernala, A. Bessi, K. Lerman, E. Ferrara, DISCOVER: Mining online chatter for emerging cyber threats, in: WWW, International World Wide Web Conferences Steering Committee, France, 2018, pp. 983–990.

[176] A. Tse, K.M. Carley, Event-based model simulating the change in DDoS attack trends after P/DIME events, in: D. Lee, Y.-R. Lin, N. Osgood, R. Thomson (Eds.), SSBP-BRiMS, Springer, Cham, 2017, pp. 120–126.

[177] D.E. Sanger, C. Savage, U.S. Says Russia Directed Hacks to Influence Elections, The New York Times, 2016, [Online]. Available: http://www.nytimes.com/2016/10/08/us/politics/us-formally-accuses-russia-of-stealing-dnc-emails.html. (Accessed in: 06/2022).

[178] A.E. Kramer, Vladimir Putin Exits Nuclear Security Pact, Citing 'Hostile Actions' by U.S., The New York Times, 2016, [Online]. Available: http://www.nytimes.com/2016/10/04/world/europe/russia-plutonium-nuclear-treaty.html. (Accessed in: 06/2022).

[179] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J.A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, Y. Zhou, Understanding the Mirai botnet, in: USENIX, USENIX Association, Canada, 2017, pp. 1093–1110.

[180] H. Kaur, H.S. Pannu, A.K. Malhi, A systematic review on imbalanced data challenges in machine learning: Applications and solutions, ACM Comput. Surv. 52 (4) (2019).

[181] A. Boukerche, L. Zheng, O. Alfandi, Outlier detection: Methods, models, and classification, ACM Comput. Surv. 53 (3) (2020).

[182] L. Wang, M. Han, X. Li, N. Zhang, H. Cheng, Review of classification methods on unbalanced data sets, IEEE Access 9 (2021) 64606–64628, http://dx.doi.org/10.1109/ACCESS.2021.3074243.

[183] L. Zi, J. Yearwood, X.-W. Wu, Adaptive clustering with feature ranking for DDoS attacks detection, in: NSS, 2010, pp. 281–286.

[184] I. Sharafaldin, A.H. Lashkari, A.A. Ghorbani, Toward generating a new intrusion detection dataset and intrusion traffic characterization, in: ICISSP, SCITEPRESS, 2018.

[185] A. Mahfouz, A. Abuhussein, D. Venugopal, S. Shiva, Ensemble classifiers for network intrusion detection using a novel network attack dataset, Future Internet 12 (11) (2020) 180.

[186] X. Zhou, M. Belkin, Chapter 22 - semi-supervised learning, in: P.S. Diniz, J.A. Suykens, R. Chellappa, S. Theodoridis (Eds.), Academic Press Library in Signal Processing: Volume 1, in: Academic Press Library in Signal Processing, vol. 1, Elsevier, 2014, pp. 1239–1269.

[187] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: KDD, AAAI Press, USA, 1996, pp. 226–231.

[188] T. Kohonen, The basic SOM, in: Self-Organizing Maps, Springer, Berlin, Heidelberg, 1997, pp. 85–144.

[189] L. Rokach, Ensemble methods for classifiers, in: Data Mining and Knowledge Discovery Handbook, Springer, USA, 2005, pp. 957–980.

[190] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, JCSS 55 (1) (1997) 119–139.

[191] J.H. Friedman, Stochastic gradient boosting, CSDA 38 (4) (2002) 367–378, Nonlinear Methods and Data Mining.

[192] N. Mahmoud, Y. Essam, R. Elshawi, S. Sakr, DLBench: An experimental evaluation of deep learning frameworks, in: IEEE BigDataCongress, IEEE, Italy, 2019, pp. 149–156.

[193] M. Pan, Y. Liu, J. Cao, Y. Li, C. Li, C.-H. Chen, Visual recognition based on deep learning for navigation mark classification, IEEE Access 8 (2020) 32767–32775.

[194] S. Amarjyoti, Deep reinforcement learning for robotic manipulation - the state of the art, 2017, CoRR arXiv:1701.08878.

[195] R. Dey, F.M. Salem, Gate-variants of Gated Recurrent Unit (GRU) neural networks, in: MWSCAS, IEEE, USA, 2017, pp. 1597–1600.

[196] Y. Chen, X. Zhao, X. Jia, Spectral–spatial classification of hyperspectral data based on deep belief network, IEEE J-STARS 8 (6) (2015) 2381–2392.

[197] X.-L. Zhang, J. Wu, Deep belief networks based voice activity detection, IEEE Trans. Audio Speech Lang. Process. 21 (4) (2013) 697–710, http://dx.doi.org/10.1109/TASL.2012.2229986.

[198] D. Jha, L. Ward, A. Paul, W. keng Liao, A. Choudhary, C. Wolverton, A. Agrawal, ElemNet: Deep learning the chemistry of materials from only elemental composition, Sci. Rep. 8 (1) (2018).

[199] F.K. Došilović, M. Brčić, N. Hlupić, Explainable artificial intelligence: A survey, in: MIPRO, IEEE, Croatia, 2018, pp. 0210–0215.

[200] X. He, K. Zhao, X. Chu, AutoML: A survey of the state-of-the-art, Knowl.-Based Syst. 212 (2021) 106622.

[201] D.J. Kedziora, K. Musial, B. Gabrys, AutonoML: Towards an integrated framework for autonomous machine learning, 2020, pp. 1–77, CoRR arXiv:2012.12600.

[202] A.B. de Neira, A.M. Araujo, M. Nogueira, Early botnet detection for the Internet and the Internet of Things by autonomous machine learning, in: MSN, Japan, 2020, pp. 516–523.

[203] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, H. Yu, Federated learning, Synth. Lect. Artif. Intell. Mach. Learn. 13 (3) (2019) 1–207.

[204] Y. Feng, J. Li, T. Nguyen, Application-layer DDoS defense with reinforcement learning, in: IWQoS, 2020, pp. 1–10.

[205] D.K. Dake, J.D. Gadze, G.S. Klogo, DDoS and flash event detection in higher bandwidth SDN-IoT using multiagent reinforcement learning, in: ICCMA, 2021, pp. 16–20.

[206] L. Huang, A.D. Joseph, B. Nelson, B.I. Rubinstein, J.D. Tygar, Adversarial machine learning, in: AISec, ACM, USA, 2011, pp. 43–58.

[207] M. Stamp, Introduction To Machine Learning with Applications in Information Security, CRC Press, USA, 2017.

[208] M. Barreno, B. Nelson, A.D. Joseph, J.D. Tygar, The security of machine learning, Mach. Learn. 81 (2) (2010) 121–148.

[209] M. Kianpour, S.-F. Wen, Timing attacks on machine learning: State of the art, in: Y. Bi, R. Bhatia, S. Kapoor (Eds.), ISWA, Springer, Cham, 2020, pp. 111–125.

[210] M.L. Puterman, Chapter 8 Markov decision processes, in: Stochastic Models, in: Handbooks in Operations Research and Management Science, vol. 2, Elsevier, North-Holland, 1990, pp. 331–434.

[211] M.T.J. Spaan, Partially observable Markov decision processes, in: Reinforcement Learning: State-of-the-Art, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 387–414.

[212] A. Hanbanchong, K. Piromsopa, SARIMA based network bandwidth anomaly detection, in: JCSSE, 2012, pp. 104–108, http://dx.doi.org/10.1109/JCSSE.2012.6261934.

[213] E. Erdogdu, Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey, Energy Policy 35 (2) (2007) 1129–1146.

[214] N. Elamin, M. Fukushige, Modeling and forecasting hourly electricity demand by SARIMAX with interactions, Energy 165 (2018) 257–268.

[215] P. Manigandan, M.S. Alam, M. Alharthi, U. Khan, K. Alagirisamy, D. Pachiyappan, A. Rehman, Forecasting natural gas production and consumption in United States-Evidence from SARIMA and SARIMAX models, Energies 14 (19) (2021).

[216] A.K. Kar, Bio inspired computing – A review of algorithms and scope of applications, Expert Syst. Appl. 59 (2016) 20–32.

[217] U. Rauf, A taxonomy of bio-inspired cyber security approaches: Existing techniques and future directions, Arab. J. Sci. Eng. 43 (12) (2018) 6693–6708.

[218] D.J. Prathyusha, S. Naseera, D.J. Anusha, K. Alisha, A review of biologically inspired algorithms in a cloud environment to combat DDoS attacks, in: Smart Intelligent Computing and Applications, Springer, 2019, pp. 59–68.

[219] Q. Tian, D. Han, Z. Du, DDoS attack detection based on global unbiased search strategy bee colony algorithm and artificial neural network, Int. J. Embed. Syst. 11 (5) (2019) 584–593.

[220] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, JMLR 12 (2011) 2825–2830.

[221] U. Gov, £1.2 Billion for the World's Most Powerful Weather and Climate Supercomputer, UK Gov, 2020, https://www.gov.uk/government/news/12-billion-for-the-worlds-most-powerful-weather-and-climate-supercomputer. (Accessed in: 08/2021).

[222] P.J. Webster, Improve weather forecasts for the developing world, Nature 493 (7430) (2013) 17–19.

[223] R. Bikmukhamedov, A. Nadeev, Lightweight machine learning classifiers of IoT traffic flows, in: SYNCHROINFO, Russia, 2019, pp. 1–5.

[224] D. Kshirsagar, S. Kumar, A feature reduction based reflected and exploited DDoS attacks detection system, JAIHC 13 (1) (2021) 393–405.

[225] R.K. Batchu, H. Seetha, A generalized machine learning model for DDoS attacks detection using hybrid feature selection and hyperparameter tuning, Comput. Netw. 200 (2021) 108498.

[226] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (1) (2014) 16–28.

[227] J. Miao, L. Niu, A survey on feature selection, Procedia Comput. Sci. 91 (2016) 919–926.

[228] B. Venkatesh, J. Anuradha, A review of feature selection and its methods, Cybernet. Inf. Technol. 19 (1) (2019) 3–26.

[229] Y. Xuan, W. Si, J. Zhu, Z. Sun, J. Zhao, M. Xu, S. Xu, Multi-model fusion short-term load forecasting based on Random Forest feature selection and hybrid Neural Network, IEEE Access 9 (2021) 69002–69009.

[230] M. Rostami, K. Berahmand, S. Forouzandeh, A novel community detection based genetic algorithm for feature selection, J. Big Data 8 (1) (2021).

[231] H. Nasiri, S.A. Alavi, A novel framework based on deep learning and ANOVA feature selection method for diagnosis of COVID-19 cases from chest X-ray images, in: T.R. G. (Ed.), Comput. Intell. Neurosci. 2022 (2022) 1–11.

[232] Y. Feng, H. Akiyama, L. Lu, K. Sakurai, Feature selection for machine learning-based early detection of distributed cyber attacks, in: DASC, IEEE, 2018, pp. 173–180.

[233] P.M. Gonçalves, S.G. de Carvalho Santos, R.S. Barros, D.C. Vieira, A comparative study on concept drift detectors, Expert Syst. Appl. 41 (18) (2014) 8144–8156.

[234] G. Andresini, F. Pendlebury, F. Pierazzi, C. Loglisci, A. Appice, L. Cavallaro, INSOMNIA: Towards concept-drift robustness in network intrusion detection, in: AISec, AISec '21, ACM, USA, 2021, pp. 111–122.

[235] B.H. Schwengber, A. Vergütz, N.G. Prates, M. Nogueira, Learning from network data changes for unsupervised botnet detection, IEEE TNSM 19 (1) (2022) 601–613, http://dx.doi.org/10.1109/TNSM.2021.3109076.

[236] A.M. de Araújo, A.B. de Neira, M. Nogueira, Lifelong autonomous botnet detection. (to appear), in: GLOBECOM, IEEE, Brazil, 2022, pp. 1–6.

[237] H. Mrabet, S. Belguith, A. Alhomoud, A. Jemai, A survey of IoT security based on a layered architecture of sensing and data analysis, Sensors 20 (13) (2020) 1–19.

[238] S. Riazul Islam, M.N. Uddin, K.S. Kwak, The IoT: Exciting possibilities for bettering lives: Special application scenarios, IEEE MCE 5 (2) (2016) 49–57.

[239] H. Xu, W. Yu, D. Griffith, N. Golmie, A survey on industrial Internet of Things: A cyber-physical systems perspective, IEEE Access 6 (2018) 78238–78259.

[240] M.B. Yassein, I. Hmeidi, M. Al-Harbi, L. Mrayan, W. Mardini, Y. Khamayseh, IoT-based healthcare systems: A survey, in: DATA, ACM, USA, 2019.

[241] Y. Zou, L. Quan, A new service-oriented grid-based method for AIoT application and implementation, MPLB 31 (19–21) (2017) 1740064.

[242] X. Yu, H. Guo, A survey on IIoT security, in: APWCS, IEEE, Singapore, 2019, pp. 1–5.

[243] J. Li, M. Liu, Z. Xue, X. Fan, X. He, RTVD: A real-time volumetric detection scheme for DDoS in the Internet of Things, IEEE Access 8 (2020) 36191–36201.

[244] M. Alhanahnah, Q. Lin, Q. Yan, N. Zhang, Z. Chen, Efficient signature generation for classifying cross-architecture IoT malware, in: IEEE CNS, IEEE, Beijing, China, 2018, pp. 1–9.

[245] K. Benzekki, A. El Fergougui, A. Elbelrhiti Elalaoui, Software-defined networking (SDN): a survey, Secur. Comm. Netw. 9 (18) (2016) 5803–5833.

[246] B. Rodrigues, T. Bocek, A. Lareida, D. Hausheer, S. Rafati, B. Stiller, A blockchain-based architecture for collaborative DDoS mitigation with smart contracts, in: D. Tuncer, R. Koch, R. Badonnel, B. Stiller (Eds.), LNCCN, Springer, Cham, 2017, pp. 16–29.

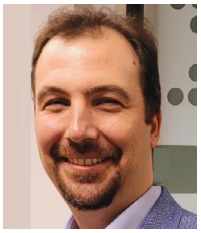
[247] O.E. Tayfour, M.N. Marsono, Collaborative detection and mitigation of distributed denial-of-service attacks on software-defined network, Mobile Netw. Appl. 25 (4) (2020) 1338–1347.

[248] M. Sachdeva, G. Singh, K. Kumar, K. Singh, A comprehensive survey of distributed defense techniques against DDoS attacks, IJCSNS 9 (12) (2009) 7–15.

[249] C.V. Zhou, C. Leckie, S. Karunasekera, A survey of coordinated attacks and collaborative intrusion detection, C&S 29 (1) (2010) 124–140.

[250] L.R. Foulds, Graph Theory Applications, Springer Science & Business Media, USA, 2012.
[251] J.A. Bondy, U.S.R. Murty, et al., Graph Theory with Applications, vol. 290, The Macmillan Press Ltd, Great Britain, 1976.
[252] X. Xiong, K. Ozbay, L. Jin, C. Feng, Dynamic prediction of origin-destination flows using fusion line graph convolutional networks, 2019, CoRR arXiv:1905.00406.
[253] Q. Xie, T. Guo, Y. Chen, Y. Xiao, X. Wang, B.Y. Zhao, Deep graph convolutional networks for incident-driven traffic speed prediction, ACM, USA, 2020, pp. 1665–1674.
[254] X. Li, Z. Fan, Y. Xiao, Q. Xu, W. Zhu, Improved automated graph and FCM based DDoS attack detection mechanism in software defined networks, JIT 20 (7) (2019).
[255] H. Jing, J. Wang, DDoS detection based on graph structure features and non-negative matrix factorization, CCPE (2020) 13.
[256] E. Parliament, C. of the European Union, General data protection regulation (accessed in: 06/2021), 2016, [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679,
[257] Governor of California, California Consumer Privacy Act, California State, 2018, [Online]. Available: https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5. (Accessed in: 06/2021),
[258] P. of the Republic of Brazil, Lei geral de proteção de dados pessoais (LGPD), 2018, [Online]. Available: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. (Accessed in: 06/2021),
[259] F. Klement, H.C. Pöhls, K. Spielvogel, Towards privacy-preserving local monitoring and evaluation of network traffic from IoT devices and corresponding mobile phone applications, in: GIoTS, IEEE, Dublin, Ireland, 2020, pp. 1–6.

Anderson Bergamini de Neira is a Ph.D. candidate at Federal University of Paraná (UFPR), Brazil. His main research interest includes security in computer networks, especially in solutions that use machine learning to reduce the impacts of DDoS attacks. He is a member of the Center for Computational Security Science research team.



Burak Kantarci is an Associate Professor and the Founding Director of the Next Generation Communications and Computing Networks (NEXTCON) Research Lab at the University of Ottawa. Dr. Kantarci is the co-author of over 200 publications in established journals and conferences, and 13 book chapters; and a recent recipient of the Canada Foundation for Innovation (CFI)-John R. Evans Leaders Fund for AI-Backed Internet of Vehicles Laboratory. Dr. Kantarci is well-known for his contributions to the quantification of data trustworthiness in mobile crowd-sensing (MCS) systems, and game theoretic incentives to promote user participation in MCS campaigns with high value data; as well as AI-backed access control, authentication and intrusion detection solutions in sensing environments. He served as the Chair of IEEE Communications Systems Integration and Modeling Technical Committee, and has served as the Technical Program Co-Chair/Symposium Co-chair of more than fifteen international conferences/symposia/workshops including IEEE Global Communications Conference (GLOBECOM) - Communications Systems QoS, Reliability and Modeling (CQRM) symposium. An Editor of the IEEE Communications Surveys and Tutorials, Elsevier Vehicular Communications, an associate editor for IEEE Networking Letters, an area editor for IEEE Transactions on Green Communications and Networking, and an associate editor for IEEE Access. Dr. Kantarci is a Distinguished Speaker of the Association for Computing Machinery (ACM), senior member of ACM, and senior member of the IEEE.

Michele Nogueira received the Ph.D. degree in Computer Science from Sorbonne University–UPMC, LIP6, France (2009). She is currently an Associate Professor of the Computer Science Department at Federal University of Minas Gerais. Her research interests include wireless networks, network security and dependability. Her works search to provide resilience to self-organized, cognitive and wireless networks by adaptive and opportunistic approaches. She is the director of the Center for Computational Security sCience (CCSC) research lab. She served as an Associate Technical Editor for the IEEE Communications Magazine and the Journal of Network and Systems Management. She serves as chair for the IEEE ComSoc Internet Technical Committee.