Measurement: Sensors 31 (2024) 100991

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Prevention and detection of DDOS attack in virtual cloud computing environment using Naive Bayes algorithm of machine learning

Yongqiang Shang

Xinyang Agriculture and Forestry University, Department of Information Engineering Department, Xinyang, Henan, 464000, China

ARTICLE INFO

Keywords:
Machine learning
Cyber attack
Virtual cloud computing environment
Cloud computing
Navie bayes

ABSTRACT

The popularity of cloud computing, with its incredible scalability and accessibility, has already welcomed a new era of innovation. Consumers who subscribe to a cloud-based service and use the associated pay-as-you-go features have unlimited access to the applications mentioned above and technologies. In addition to lowering prices, this notion also increased the reliability and accessibility of the offerings. One of the most crucial aspects of cloud technology is the on-demand viewing of personal services, which is also one of its most significant advantages. Apps that are cloud-based are available on demand from anywhere in the world at a reduced cost. Although it causes its users pain with safety concerns, cloud computing can thrive because of its fantastic instantaneous services. There are various violations, but they all accomplish something similar, taking the systems offline. Distributed denial of service attacks are among the most harmful forms of online assault. For fast and accurate DDoS (Distributed Denial of Service, distributed denial of service) attack detection. This research introduced the DDOS attack and a method to defend against it, making the system more resistant to such attacks. In this scenario, numerous hosts are used to carrying out a distributed denial of service assault against cloud-based web pages, sending possibly millions or even trillions of packets. It uses an OS like ParrotSec to pave the way for the attack and make it possible. In the last phase, the most effective algorithms, such as Naive Bayes and Random Forest, are used for detection and mitigation. Another major topic was studying the many cyber attacks that can be launched against cloud computing.

1. Introduction

DDos attack is a distributed type of attack mode in which an attacker controls a large number of attack machines and sends out DoS attack instructions to the machine. In the latest Internet security report, DDoS attacks remain one of the major cybersecurity threats. The inexpensive pricing and "pay-as-you-go" focused accessibility to computational features and amenities on demand make cloud-based services a formidable competitor to the conventional IT solutions available in prior eras. The use of cloud computing is gaining popularity rapidly. Whether entirely or largely governments and companies have moved their IT infrastructures onto the cloud. Cloud-based Infrastructure offers various advantages compared to traditional, on-site conventional infrastructures. The removal of expenses associated with operation and impairment, as well as the accessibility of materials on request, are only a few of the advantages. However, there are many concerns that cloud consumers have, and the research addresses these issues. The majority of these inquiries centre on safeguarding operational concepts and information. Many security-related attacks can be prevented in conventional IT systems that do not use cloud computing. Focused cloud-based crimes are already using their innovations. Many security vulnerabilities in cloud computing are unique compared to their predecessors in non-cloud computing environments because data and business logic are stored on an external cloud server that lacks accessible oversight. The denial-of-service (DoS) assault is one technique that has been in the spotlight recently. Denial-of-service incidents are directed at the server rather than the people it supports. DoS attackers attempt to flood live servers by masquerading genuine users to overload the service's capacity to handle incoming inquiries [1]. Cloud computing is an Internet-based service that enables users to access configurable computing resource sharing pools (including server, storage, application software, services, networks, etc.) to achieve online access to computing resources on demand. As a mixture of emerging technologies and business models, cloud computing has developed rapidly in recent years due to its advantages of super-large scale, virtualization, high reliability, good scalability and on-demand services. To overcome this issue, multiple inquiries are sent to the server simultaneously. The term "distributed denial of service," or DDoS, refers to a variation on the classic

E-mail address: YongqiangShang2@163.com.

https://doi.org/10.1016/j.measen.2023.100991
Received 25 July 2023; Received in revised form 13 November 2023; Accepted 18 December 2023
Available online 20 December 2023

"denial of service" that uses numerous computers to attack and impair one service at a time simultaneously. Among the most important and possibly catastrophic risks, among many others, is the growing number of distributed denial of service attacks observed. A quarter or more of the world's organizations have experienced a distributed denial of service attack. The authors show great foresight in predicting DDoS attacks and will increasingly focus on cloud-based assets and amenities. Multiple assaults in the past two years corroborate the paper's predictions about future attacks. There have been many attacks recently, but only a few have gained widespread notoriety and interest from scientists. In 2015, Lizard Squad hacked Microsoft and Sony's cloud-based gaming systems, causing both firms to shut down their services on Christmas Day. Distributed denial of service attacks hit Rackspace, a cloud computing services provider, hard. Another massive distributed denial of service attack was launched against Amazon EC2 cloud servers, serving as a magnificent example of an attack. Company activities were severely disrupted, money was lost, and there were immediate and long-term effects on the attacked businesses. In recent years, DDoS attacks have become more frequent, and the botnet used by attackers has become larger, and the network traffic usage has reached a height of 1000G. For cloud computing platforms, DDoS attacks from outside are similar to DDoS attacks from traditional networks. According to the basic principle and characteristics of DDoS attack, the defense is mainly divided into four stages: detection (Detecting), analysis (Analyz-ing), defense (Resisting) and counterattack (Counterattack). The detection and analysis technology is the key to the successful defense against DDoS attacks.

According to research published by Verisign iDefense Security Intelligence Solutions, distributed denial of service (DDoS) assaults have been particularly damaging to the internet and SaaS (Software as a Service) business throughout the past few quarters. More than 75 % of known countermeasures against DDoS assaults utilized services provided by the cloud [2]. "financial damages" refers to one of the worst possible results of a Distributed Denial of Service attack in the cloud. The median price of a distributed denial of service assault is put at $482,000, according to some estimates. Some of the financial losses suffered in Q1 2015 have been detailed in new disclosures from Neustar. Studies show that, on average, more than $72K is stolen in a single hour. Distributed denial of service (DDoS) attacks take on new significance in cloud computing. This variation directly results from the operational difficulties introduced by an assault on the victim network [3]. In the environment of cloud computing, DoS attack technology is also undergoing new changes, and is manifested in a variety of forms. The attack may come from outside the server cluster or from inside the server cluster. At present, the more popular attack method is for the attacker to attack a specific cloud computing platform or server cluster. This attack method causes great harm and various methods, and it is difficult to quickly carry out fault positioning and troubleshooting.

Clouds that provide Infrastructure as a Service (IaaS) to their clients contain virtual machines (VMs) that host the amenities for the clients. The flexibility and on-demand nature of the cloud is made possible by the abstraction of servers. It allows virtual machines to acquire and distribute capabilities on the fly as needed. The advantages of cloud computing, such as upon-request processing and easily accessible assets, have contributed significantly to its recent meteoric rise. As a result, the cloud can now support a more significant number of virtual machines (VMs) with a far greater capacity to meet their resource requirements. This is because a cloud-based virtual machine can access infinite resources. A Distributed Denial of Service (DDoS) assault, also known as an Economic Denial of Sustainability (EDoS) attack or a Fraudulent Resource Consumption (FRC) attack, is the result of this "adaptability" or "auto-scaling," which results in financial losses. Data center-distributed denial of service attacks is the focus of this study. We also define these assaults, compared to more conventional DDoS attacks, and analyze and classify the numerous developments in this field. We will provide a comprehensive taxonomy of these functions to make this analysis more approachable. The popularization of computer network has changed the way we work and live, and the promotion of cloud computing in recent years has provided a more convenient platform for network resource sharing. Cloud computing platform organically integrates computer infrastructure (including server resources, storage resources, network resources, etc.) through virtualization technology means, so as to realize resource sharing among multiple users, and greatly reduce the cost of using resources, making it possible to provide cheap and high-performance services for users.

1.1. DDoS attack and cloud features

Denial of service (DoS, Denial of Service) attack is a destructive attack on the target server through abnormal methods, resulting in its inability to provide services to normal network users. Currently, distributed denial of service (DDoS) assaults have achieved much accomplishment in cloud computing, where hackers make use of the "pay-as-you-go" model. Many factors contribute to cloud computing's meteoric rise in popularity, but these three features stand out as particularly crucial. On the other hand, DDoS attackers have found that the same set of features dramatically aids them in achieving the objectives of their cyber attacks. In the sections that follow, we will examine each of these features more closely. Fig. 1 describes the cloud architecture which was affected by DDoS attack [5]. There are many methods and forms of DoS attack, which are summarized in the following situations: illegal occupation and consumption of computer resources such as CPU, network bandwidth and storage space; changing or even destroying the configuration information of the target server; changing or even destroying the key node equipment in the physical network; and accessing the services by programming.

1.1.1. Automatic sizing

Physical virtualization provides the capacity to scale down, up, and re-resource a live VM. A VM's processing power, primary memory, storage area, and data transfer capacity can all be increased as needed, thanks to these features. When some of the assigned resources are not being used or needed, this can be utilized to free up some of those capabilities. Multiple vendors of services employ this method of resource distribution, which is made practical by automatic scaling and web-based tools. This allows those who use the cloud to calculate their needed facilities using utilization rates or similar matrices. It is possible to extend this functionality to automatically deploy new virtual machines (VMs) on top of existing physical servers and remove them when they are no longer needed. Upward scaling, which refers to adding more machines, and horizontal scaling, which refers to adding more data centres or clouds, are two of the most crucial computing features for

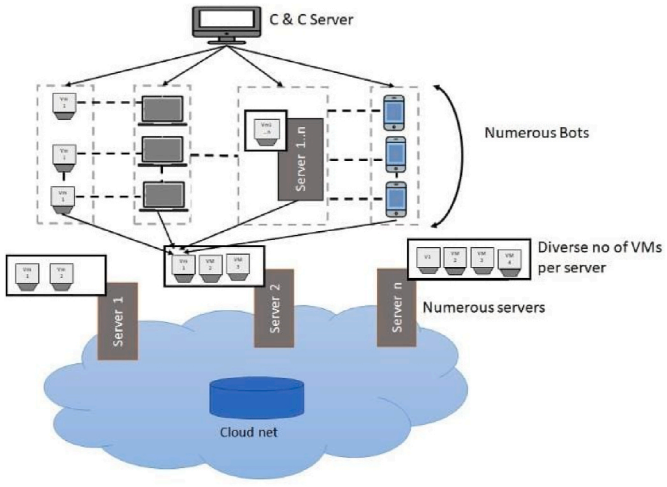

Fig. 1. Cloud architecture DDoS attacks.

utility purposes. Distributing an application across multiple cloud-hosted physical servers is one way to increase its capacity. High-speed connections and ample storage space are the two most essential factors in determining adaptability. The virtualization of OSes is crucial when contemplating the scalability of virtual machines (VMs). The process of replicating a virtual machine and then releasing it is quick. To alleviate strain, duplicate virtual machines might be launched on different servers [4]. This action can be taken at any time when it is required. Streaming virtual machine deployments are an additional significant expansion accelerator because they allow migrating an active virtual server to a different nation's more comprehensive hardware server with practically no interruption. This guarantees ongoing adaptability, which is further strengthened in this manner.

1.1.2. Pay-as-you-go reporting

Upon request, utility services have grown in popularity due to their convenience and the simplified resources reporting and invoicing they provide. Customers of cloud computing services can take advantage of the "Pay-as-you-go" model without making any upfront financial commitments for resources. The administrator of a virtual machine (VM) may want to dynamically adjust the number of resources that are accessible, either by adding more or taking them away [23]. Another perk of adopting a cloud-based system is that you can get more use out of your hardware without worrying about things like electricity, space, cooling, and maintenance. DDoS attacks in the cloud are only possible to comprehend with a firm grasp of the financial aspects of doing so. Since most cloud instances are billed hourly, the minimum possible time frame for accounting is typically 70 min. Funds could be allocated in three ways: a predetermined amount, a pay-as-you-go system, or auctions. The size and volume of data transferred in and out of a computer network also determine its usefulness. The "pay as you go" models are experimental and still in the prototype phase [6].

1.1.3. Multi-tenancy

Multi-tenancy allows several Virtual Machines (VMs) belonging to different VM proprietors to coexist on just one hardware system. Increasing hardware utilization and, consequently, one's return on investment (ROI) can be accomplished through multi-tenancy. On the same physical server, a single user can want to run multiple instances of the same program or entirely distinct ones using different virtual machines.

1.2. Cloud-based DDoS attack situation

The attack depicted in Fig. 1 is very normal. The cloud requires enormous computers that can service multiple users in a standardized setting. An attacker's purpose may not always be limited to a "Denial of Service" but can include reducing the profitability of cloud subscribers. How to prevent assaults like this has been a hot topic since the inception of cloud computing. The term "Fraudulent Resource Consumption" (FRC) attacks have been used in many other works to characterize this type of attack. Dispersed denial of service attacks targeting web pages and hackers plant bots and trojans on compromised systems all over the Internet. A DDoS attack will be executed as an EDoS attack if the target service is hosted in the cloud. "Booters" are businesses that connect their clients with a botnet to launch distributed denial of service attacks (DDoS) on their rivals' web pages. Attacks like these can be spurred on by everything from commercial competition to political rivalry to ransom to full-out cyber war between nations [22]. In view of the working principle that DDoS consumes system resources and causes the system cannot provide normal services, network managers can optimize and reinforce the system to improve the system's tolerance of DDoS attacks, and even block some DDoS attack packets. Firstly, improve the network planning and design scheme to eliminate the unreasonable factors of network structure; then implement the system security vulnerabilities and hidden dangers in the network system in the last, scan the key network equipment such as firewall and router, to find the bottleneck of network equipment and optimize the performance. The approach to cloud computing provides customers with several opportunities and benefits; nevertheless, DDoS attackers also have access to these features and may find them helpful. In order to accomplish "Denial of Service," an attacker launching a DDoS attack will send out a flood of fake inquiries. Fig. 2 describes the classification, prevention and mitigation of DDoS attacks [7]. Although DDoS attack technology is varied, it has many similarities to the phenomena caused by the system. Therefore, through the implementation of a distributed detection system, we will strive to find the behavior of DDoS attacks in the first time, and accurately locate the source and characteristics of attacks. Through the network abnormal traffic analysis system and DDoS detection tool, timely find the abnormal traffic and DDoS behavior in the network, find problems in time, and improve the overall detection and analysis ability of the system.

However, the targeted system must expend many resources to counter this hack. This "overload" condition would be seen as feedback by the "auto-scaling" function, which would then add more CPUs (or other resources) to the VM's existing amount of readily available assets. First, a virtual machine will enter its "normal load VM" phase. Let us assume the DDoS attack has commenced, and the VM is now overburdened as an immediate consequence of the attack. As soon as the cloud detects an overload, its auto-scaling features will kick in, and it will choose among the many methods described in the literature for allocating resources to virtual machines, migrating them, and relocating them. When a virtual machine (VM) gets overloaded, it can be given more resources, transferred to a server with more available resources, or have a copy of itself launched on a separate server [20]. If there is no countermeasure to halt this procedure, further resources will be added. This can continue until the service provider makes a payment or the cloud service provider exhausts all available resources, whichever comes first. The eventual outcome of this is "Service Denial [8]." The vast majority of DDoS attacks are organized and premeditated destructive acts. Relying solely on a technical department or an enterprise, it is impossible to completely solve the problem of security protection, let alone to quickly track and locate the source of attacks. For the defense of DDoS attacks, cloud computing platform suppliers, communication operators and government departments need to establish a cooperation mechanism to complete security defense.

Consequently, this results in billing for resources only when used, which raises the risk of incurring financial losses over the set limit. To keep things manageable, we might run virtual machines with a static resource profile, in which case SLA will not cover the provisioning of extra resources on demand. A "Denial of Service" (DoS) attack would immediately wipe out the cloud's valuable features in this situation. Fig. 3 is the description of tiers of cloud-based DDoS defense.

2. Related works

DDoS attacks, which target computer systems, are becoming increasingly common. DDoS perpetrators have expanded their reach into practically every area of technology, especially the cloud, the IoT, and the edge. Distributed denial of service (DDoS) attacks flood the targeted machine or host with so much traffic that it crashes or exhausts all available resources (including the network). Multiple strategies for defense have been proposed, but they have yet to be successful due to attackers' ability to educate themselves to employ recently discovered computerized ways of attack. Because of this, we presented a machine-learning-based approach to spotting distributed denial-of-service (DDoS) attacks in the cloud. K Nearest Neighbour, Random Forest, and Naive Bayes are three different categorization machine learning techniques that help the system detect a distributed denial of service attack with a 99.75% success rate. In our study, we offer a machine learning-based approach to identifying and blocking DDoS attacks against servers situated in the cloud. Data mining for relevant statistics

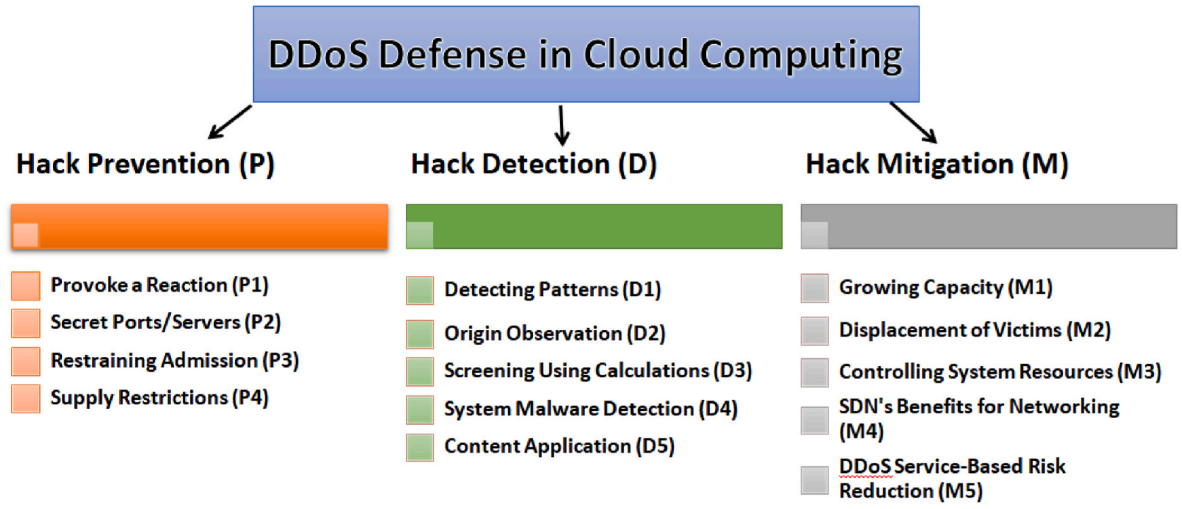

Fig. 2. Cloud DDoS classification, prevention, and mitigation.

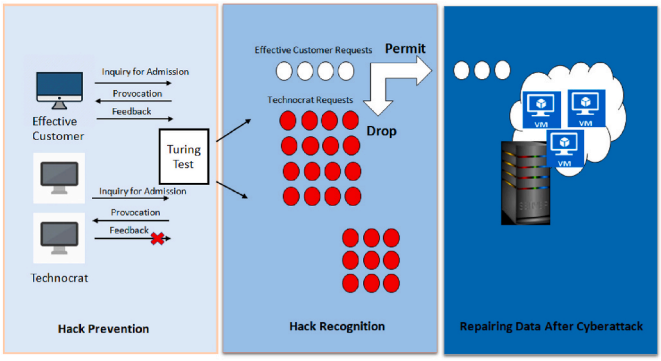

Fig. 3. Several tiers of cloud-based DDoS defence.

largely determined our recommended technique's efficacy. Table 1 illustrate the results, from which it can be deduced that the suggested method has an excellent success rate (about 99.78 %) in identifying DDoS attacks while producing few errors. Since we focused primarily on the supervised learning method in this study, future studies may investigate uncontrolled or reinforcement learning methods [9].

Distributed denial of service (DDoS) attacks are more challenging to execute on the public internet than on a conventional network. There is more than one threat to the cloud, and its surroundings are under attack from several directions. Existing machine learning techniques, such as neural classifiers, can be used to identify DDoS attacks. This research aims to shed light on the results of an investigation into distributed denial of service (DDoS) attacks in cloud settings. The number of false positives rises when artificial intelligence methods are applied for detection. The ANN, SVM, kNN, J48, Feature rank and Feature selection algorithms frequently detect Distributed Denial of Service (DDoS) attacks in a cloud context [10].

The goal of this research was to examine several works associated with the identification of network assaults in both traditional and cloud-based infrastructures. In the following paper, we will examine the wide variety of attacks that could occur in a cloud environment. There is sometimes a conflation between the terms "bandwidth reduction" and "resource reduction" when describing the impacts of distributed denial of service (DDoS) attacks. Most distributed denial of service (DDoS) assaults in the cloud are SYN Flood or Flash Crowd assaults. The analysis found that TCP denial of service low-rate assaults and performance decreases are two of the most prevalent attack categories [11].

To spot distributed denial of service (DDoS) attacks, researchers are trying out various machine-learning algorithms, some of which have shown greater precision than others. In experiments, real-time network logs, KDD, NSL-KDD, and CIDDS datasets were used to identify network attacks. Also used to predict DDoS attacks, linear regression, and logistic regression algorithms have been found to have high false favourable rates when implemented in several databases. To improve precision and recognition rates, however, it is constantly necessary to increase the number of records used for training and testing the dataset, which is a difficult task in and of itself. DDoS assaults cover a wide array of topics. Therefore, researchers can use many machine-learning techniques and classifiers in future studies. Furthermore, regression analysis has received more usage in recently released literature [12]. As a potential research strategy, we can reduce dimensionality and then use the remaining data for regression evaluation.

The number of people using online resources has increased recently due to the COVID-19 outbreak. As a direct consequence, there has been an increase in the number of end users subscribing to various cloud-based applications, which provide various services to the end user. DDoS assaults, on the other hand, are aimed at interrupting cloud computing services' availability and processing power. This has the effect of negatively impacting both the performance and accessibility of cloud computing resources. There is currently no reliable method for detecting or filtering DDoS attacks, so they are a reliable tool for anyone looking to launch cyberattacks. Recently, scientists have started experimenting with machine learning (ML) techniques to develop effective ML-based tactics to detect distributed denial of service (DDoS) attacks

Table 1
Dataset snapshot.

No.	Time	Preliminary Place	Target	Protocol	Length	Information
1	113.6020	11.0.2.20	194.172.8.2	DNS	77	Standard query 0x0aa0 A
2	113.6037	11.0.2.20	11.0.2.20	DNS	174	Standard query response 0x375e AAAA
3	113.6039	91.32.134.1	11.0.2.20	TCP	66	82 > 59,501 (ACK)Seq = 310 Ack = 18 Win = 66,646 Len = 0
4	113.6039	91.32.134.1	11.0.2.20	TCP	66	82 > 59,523 (ACK)Seq = 310 Ack = 18 Win = 66,646 Len = 0
5	113.6039	91.32.134.1	11.0.2.20	TCP	66	82 > 59,547 (ACK)Seq = 310 Ack = 18 Win = 66,646 Len = 0

[13]. In this scenario, we offer a method for detecting distributed denial of service (DDoS) assaults in a cloud computing environment by combining big data with deep learning methods. The proposed method employs big data sparking innovation to examine many incoming packets and a deep learning machine learning algorithm to filter fraudulent transmissions. Both of these technologies are used to make the methodology more effective. The testing and training phases were done with the KDDCUP99 dataset, and the final result attained a precision of 99.82 %. Even if the number of people using smart devices proliferates, the computing power and resources available in these devices still need improvement.

The cloud-based system offers multiple solutions for overcoming the issue of scarce resources by allowing for their cooperative use. The cloud computing platform is periodically targeted by attackers while being susceptible to a wide range of cyber threats. As such, we provide access to a DDoS warning system that is capable to detect the DDoS attack in a timely and accurate fashion. To avoid malicious or undesirable communications from reaching a cloud computing environment, we offer an approach that employs big data and deep learning techniques. This is achieved by employing these methods. We hope to eventually implement our suggested approach along with additional methods to enhance its overall functioning and test its usefulness on a wide range of datasets [14].

Additionally, a more effective DDoS attack avoidance mechanism might be constructed and recommended as a future work of this study in order to manage DDoS attacks in a cloud computing environment in an efficient manner. The examination of various DDoS prevention strategies that have been used in the past, as well as those that are considered state-of-the-art, is the only purpose of this work. The scope of future study may be expanded to include the presentation of a novel and effective DDoS prevention method to deal with the attacks [15].

The term "cloud computing" describes a new and attractive model for administering and distributing offerings over the World Wide Web. Because of this, information retention strategies are changing across the IT environment. Data security must be considered when handling massive amounts of data storage. Intruders pose one of the biggest challenges to data security in the modern Internet environment. The resources, data, and applications stored on the public internet are vulnerable to assault due to the system's connection. Intrusion Detection Systems (IDS) are employed in the cloud to monitor malicious behavior on both the network and the host systems. Because it creates so much illicit information online, detecting a Distributed Denial of Service (DDoS) attack is challenging for Intrusion detection systems (IDS). Cybersecurity analytics can aid in the detection of intrusions through the use of methods for data mining. Many distinct approaches have been developed with machine learning methods as their foundation [16].

Selecting features is another effective method for decreasing the dataset's dimensionality. This research proposes two distinct approaches for utilizing the dataset generated via NSL-KDD. Learning Vector Quantization (LVQ) is a filtering technique that comes first. The second technique is dimensionality reduction by principal component analysis (PCA). Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) categorization were applied to the characteristics chosen from each technique, and the results were compared for their ability to identify DDoS attacks [17]. The results show that the LVQ-based DT method is superior to the alternatives when it comes to spotting attacks. Unauthorized access to confidential data must be detected as the first step in securing that information [18].

The NSL-KDD standard is the foundation for a cloud-based intrusion detection system. In this study, we explore data pertaining to distributed denial of service attacks. LVQ, PCA, and other feature selection methods were used to classify the attacks using machine learning techniques such as neural networks, support vector machines, and decision trees. In order to properly categorize DDoS attacks, it was necessary to look at how well various techniques worked [19–21]. The PCA selected 21 features from a possible 42, while the LVQ selected only 20. The results suggest that LVQ-based feature selection in the DT model may be more accurate than the other methods in identifying attacks. As mentioned earlier, the model also outperforms its predecessors in terms of accuracy, recollection, particularity, and f-score.

3. Materials and methods

3.1. Navie Bayes algorithm

The premise that the most straightforward answers often turn out to be the most enlightening is evident in Naive Bayes and may be demonstrated in practice in daily situations. Machine learning has come a long way in recent years, but its continued development shows that it can still be kept very straightforward without compromising efficiency, accuracy, or dependability. It serves many functions and has particular strength in resolving problems associated with natural language processing (NLP). In machine learning, the naive Bayes technique is a standard statistical methodology used to solve classification problems based on the Bayes Theorem. To clarify any lingering questions, the following paragraphs will thoroughly explain the Naive Bayes algorithm and its core concepts. The speed with which an NB model may be built makes it particularly useful when dealing with vast amounts of data. The Naive Bayes approach has been widely used because of its simplicity and ability to outperform more complex classification techniques. The foundation of a Bayesian classification is the assumption that indicators can be treated separately. A Naive Bayes classifier assumes that the presence of one feature in a class does not influence the presence of any other feature, which simplifies things.

The Naive Bayes classifier is a popular guided machine learning approach in applications like text classification. Since it mimics the distribution of inputs for a given class or category, it belongs to the group of learning algorithms known as generative learning approaches. To be successful, this tactic relies on the assumption that the input data's attributes are conditionally independent given the class. This allows for fast and accurate recommendation generation by the system.

Naive Bayes classifiers, which implement Bayes' statistical theorem, are often thought of as being used for more fundamental probabilistic categorization tasks. This theorem incorporates empirical evidence and supplementary context when determining a hypothesis's credibility. In order to function, the naive Bayes classifier relies on the assumption that the input data's attributes are unrelated to one another. Contrarily, real-world scenarios usually play out differently. Although based on an unduly naive premise, the Naive Bayes classifier sees widespread application. This is because it serves its purpose well and has proven highly efficient in several practical settings.

One of the simplest Bayesian network models, naive Bayes classifiers, can achieve high levels of reliability when used in conjunction with kernel density estimation. Despite their simplicity, they are used less than other Bayesian network models. When the distribution pattern of the input data is not given, using a kernel function to approximate the probability density function of the input data can help the classifier operate better. The purpose of developing this strategy was to raise efficiency. This proves that the naive Bayes classifier is an effective machine learning technique for various purposes, including but not limited to text categorization, spam filtering, and sentiment analysis. Thomas Bayes is credited with developing the method for predicting a probability given a set of known probabilities currently known as Bayes' Theorem. Fig. 4 is the layout of Navie bayes.

3.2. Understanding Naive Bayes and machine learning

Machine learning has two main branches: supervised learning and unsupervised learning. Classification and regression are two subsets of supervised learning that can be distinguished here. Classification is where the Naive Bayes method excels. The naive Bayes method was used for face recognition. People's faces and other features, like their noses,

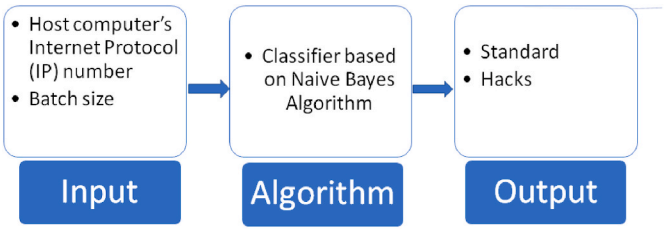

Fig. 4. Procedure for navie bayes.

mouths, eyes, etc., can be recognized using this classification method. In meteorology, it can be used to foretell whether the following weather will be pleasant or unpleasant. Doctors can make accurate diagnoses with the help of the classifier. Doctors can assess a patient's likelihood of developing cancer, cardiovascular disease, or other disorders using the Naive Bayes approach. Using a Naive Bayes classifier, Google News can decide whether a news piece is about politics, the world, or any other topic. The Naive Bayes classifier has the advantages of being simple, easily implemented, and requiring little training data. Both continuous and discrete data types are manageable using this method. It is stable even when exposed to many predictors and data points. It is fast, can be used to make predictions in the here and now, and does not care about trivial details.

4. Proposed method

Gathering relevant data should be the initial step. By collecting relevant data, we can locate and exploit several security holes in the victim's computers in our attack. All available information regarding running services, open and closed ports, and other security holes is compiled during the information-gathering phase. Here, the attacker has a better chance of learning the weak spots of the victim, making further attacks much simpler. The cloud service provider assigns a different port number to each of its services, such as: In most cases, FTP uses port 990, but it can use port 21 as well; HTTP uses port 80. TCP and UDP use ports 20 through 23 for various purposes.

In conclusion, gathering information is a procedure that provides an attacker with all the necessary data to launch a successful attack on any target system. In order to learn more about a network, we can employ the Nmap scanner. It simply needs the target machine's IP address to launch an attack; at this point, it will perform a full system scan, revealing the targeted system's activity, services, open ports, and so on. This implies that when the exposed connection is found, whatever occurring right now may be shown, regardless of what OS the other system is using. We would probably come up with an attack plan, and that plan would involve a Distributed denial of service attack, which would involve methods like the "ping of death." A distributed denial of service (DDOS) assault is one of the most damaging types of cyberattacks since it disrupts the entire system. Due to the flood of packets caused by the DDoS assault, all services are either momentarily or completely inaccessible. ParrotSec, like Kali and Ubuntu, can be managed via command line interface, with the shell or terminal serving as the primary interface for entering these instructions. This feature is shared with ParrotSec. Since ParrotSec handles everything, you can type "PING IP" into the console, and it will be carried out. Since the victim site would receive over 65 thousand packets, all services would be taken down. This is how an assault could be generated. The subsequent stage is detection. In this case, the target is a website hosted in the cloud, and Nmap is used to scan the entire site in order to locate any security flaws. This would lead to the exposure of any underlying problems. After the exposed ports have been made public, a Python script comprising a distributed denial of service attack will be created and run. This implies that when the exposed connection is found, whatever occurring right now may be shown, regardless of what OS the other system is using. We would probably come up with an attack plan, and that plan would involve a Distributed denial of service attack, which would involve methods like the "ping of death." A distributed denial of service (DDOS) assault is one of the most damaging types of cyber-attacks since it disrupts the entire system. Due to the flood of packets caused by the DDoS assault, all services are either momentarily or completely inaccessible. ParrotSec, like Kali and Ubuntu, can be managed via command line interface, with the shell or prompt serving as the main interface for entering these instructions.

Wireshark thoroughly analyzes each incoming packet. After finishing the thorough packet analysis, a large data set was produced, which may indicate the presence of a classifier. The experimental setting demonstrates that both the random forest and the naive Bayes classifier, both of which are well-known, produce excellent results. While various other classifiers may be used for detection (support vector machines, k-nearest neighbors, k-means, etc.), "Naive Bayes" is still the most effective.

In this work, naive Bayes is applied to the problem of predicting application-layer packets during distributed denial-of-service attacks.

Notwithstanding the apparent simplicity, the Naive Bayes algorithm may make precise forecasts using the current data. The data set under consideration was trained with naive Bayes, and then a fresh information set was built using the cross-validation technique with 65 folds. This was done so that we could figure out where the files were coming from and where they were going. The true affirmative level, false alarm rate, fake negative level, and many more are just some of the metrics that may be derived from this fresh information set. Naive Bayes, a technique for making predictions, produces a mix of correct and incorrect results. A fake negative is considered an alarm for the benefit of internet consumers. Naive Bayes and random forest both correctly identified the true positives as ordinary packets, whereas the false negatives were classified as DDoS attacks.

5. Experimentation & results

5.1. Data pre-processing

Regarding data mining, the most efficient method is preliminary data processing. It streamlines complex information into something everyone can understand. Due to its unreliability and lack of granularity, real-time data necessitates transforming pretreatment into valuable information. This is because information in real-time is often unreliable and vague. Weka includes numerous options for preprocessing filters. A single filter, such as normalization, is chosen from the available options. Data standardization, or "making data un-redundantly," refers to removing superfluous or identical information from a dataset.

5.2. Training data set

The procedure for the collection of collecting training information includes the construction of a machine-learning model. Programming a computer algorithm typically requires the use of data to train it. Said training information is a subset of a dataset used for instruction and evaluation alongside the entire dataset. Separating the datasets into training and testing sets is an essential first step when developing a machine learning-based model. However, a model driven by machine learning is necessary to generate further forecasts against the newly acquired dataset.

5.3. Prediction algorithm

Following the development and validation of the information set, various algorithms have been developed through this process to anticipate several of the issues. In this particular scenario, one must consider identifying whether DDoS messages are harmful or not.

5.4. Prediction of naive bayes

The percentages of real positives and fake positives are displayed in this figure.

The percentage of fake positives is seen as an indicator of a distributed denial of service attack (DDoS) or of fake data packets. In contrast, the proportion of actual positives is the standard one. In this case, the average mean of actual packets is 0.973, while the overall mean for fraudulent transmissions is approximately 0.05.

5.5. Proposed formula for naive bayes

$$P(x|y) = P(y|x)\ P(x)\ /\ P(y)$$

Where, The conditional probability of y given x is denoted by P (y|x), The likelihood of a class being P(x) and the conditional likelihood of a predictor is P(y), Probability of occurrence is P (x|y).

5.6. Basic theory

5.6.1. Three-way handshake

The between-machine communication paradigm is depicted in Fig. 2, and it must be adhered to for the communication to succeed. A three-way handshake is the name given to this particular protocol. Within the context of this dialogue, a protocol exchange takes place between the server and the hacker. When establishing a standard TCP relationship, the attacker contacts the client by sending an SYN protocol. This is referred to as the "three-way handshake." A buffer will be allotted to the user by the server as a reaction, and the server will also send back an ACK packet in addition to the SYN packet. At this stage, the connection is in a state that is referred to be "partially accessible," and it is waiting for an ACK response from the adversary in order to complete the link configuration. The process that occurs once it has been determined that a relationship has been successfully established is called the three-way handshake.

On the other hand, instances known as TCP SYN Flood are intended to exploit this three-way handshake by saturating the server with an excessive number of SYN queries. The denial of functionality attack, of which TCP SYN Flood is a prominent example, falls within the DoS category. Employing a prolonged link and monitoring a duplicate of the server's activity is required for a packet capture program to identify a TCP SYN Flood as having occurred. One way to accomplish this is to keep an eye on a copy of the server's traffic. Introducing an incoming IP Address to the server typically corresponds with the manifestation of TCP SYN Flood properties. After being submitted to calculation within a predetermined period, IP Addresses that continually show on the server are utilized to get characteristics in a DDoS attack.

5.6.2. Naive Bayes algorithm

A simple computational approach that can be used to calculate conditional likelihoods is the Naive Bayes Theorem. A probabilistic condition quantifies the likelihood of one event based on the presumption, premise, declaration, or reality that a second event has already occurred. An analogy would be the chance of something happening after something else has happened. The posterior likelihood can be computed using a formula like the one below based on the Naive Bayes theorem.

$$P(A|B) = \frac{P(B|\mathrm{A})P(A)}{P(B)}$$

If A is more likely if B happens to be accurate, then P $(A|B)$ represents the conditional likelihood of B if A is true. In probability theory, P(A) stands for the likelihood of occurrence A, and P(B) stands for the likelihood of occurrence B. We discussed using the packet-capturing software as a computational input to estimate the IP address and packet length obtained. We did the maths using the Naive Bayes method and the Gaussian distribution. After the computation, the outcomes are displayed on a two-dimensional network. The Gaussian Naive Bayes approach, which requires the calculation of the mean and standard deviation for analysis, is applied once the quantitative input has been gathered. Table 1 is about the dataset format sample.

5.6.3. Matlab's Current classification using the Naive Bayes algorithm

Matlab is the application we employ for the method of categorization because it is not only user-friendly but also highly effective in producing aesthetically pleasing outcomes. In the environment of analyzing information, a tool built into Matlab allows users to do Naive Bayes categorization. Using this method, we can also classify network traffic as either K, L, or Q to gain further insight into the type of data transmitted throughout an internet connection. This concept will be challenging to grasp for a significant number of individuals. The Matlab script for the Naive Bayes classification and the parameters that go along with it are displayed in the following figure. The results of categorizing the information obtained from the system are shown in the figure. The nonlinear shape the blue line represents limits the standard class set, of which the green circle is a component. The blue line shows these limitations. The other variety is an array of red squares depicting some threat. Fig. 5 defines the DDoS attack detection using MATLAB.

6. Conclusion

The key goals of this study are to learn how to recognize and prevent attacks involving distributed denial-of-service. The first and most crucial step is determining which ports can be exploited. Nevertheless, this approach is not risk-free because susceptible ports are more likely to be exploited. Given ParrotSec's track record for stability and performance, we decided it would be the ideal choice for our company's computer system. Since a DDoS attack involves sending one million separate packets toward the target, starting with an on-the-internet website would be best. The targeted website was taken offline after it became clear that an assault had happened. Machine learning is constructive in this detecting process as well. Using this data, the most popular and accessible tool, "weka," is being trained. Employing pre-processing techniques and the "discretize" filter to achieve the desired effect. Therefore, the following phase is not only quite intriguing but also rather useful for both forecasting and detecting. We employed both methods and compared the findings on the same platform, and we found that the naive Bayes method provides the most trustworthy conclusions. PCA selected 21 features from the possible 42 features, while LVQ selected only 20 features. The results suggest that LVQ based feature selection in the DT model may be more accurate than other methods in identifying attacks. As mentioned earlier, the model also outperformed the previous models in terms of accuracy, recall, specificity, and f-score. It was shown that the naive Bayes model had significantly better predictive power than the random forest model. There is a chance that a false positive rate warning will be triggered for packet transmissions within a network. Moreover, when compared to the random forest, naive Bayes produces considerably more accurate forecasts. It was demonstrated that the Naive Bayes algorithm outperformed the random forest technique to identify the false and actual rate of transmissions. The result detection is not carried out in real time. Although attacks can be detected, real-time alarm cannot be realized in the environment of high cluster security, so the feasibility of real-time monitoring under Hadoop platform should be studied continuously.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

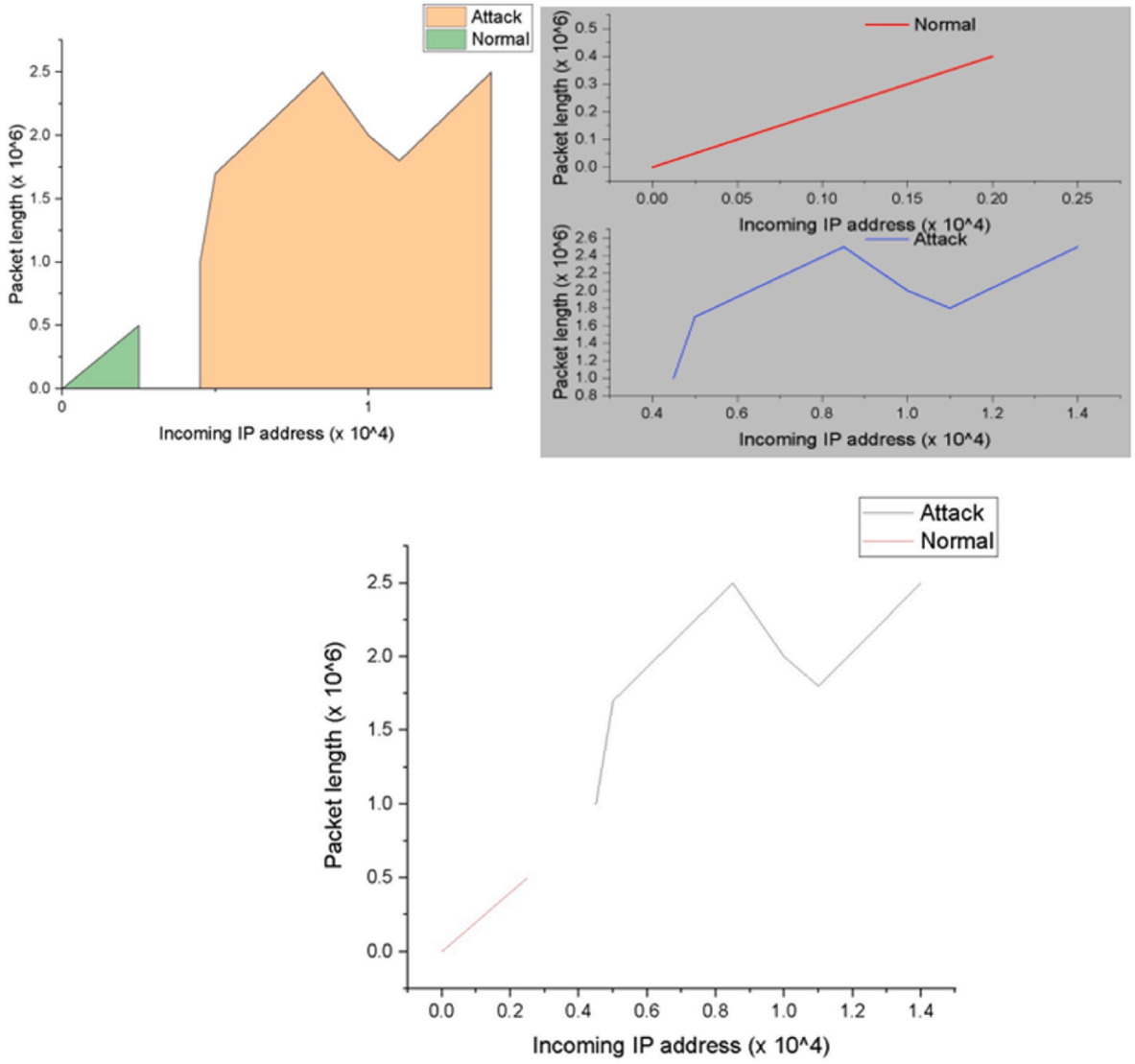

Fig. 5. Categorization outcome using MATLAB module.

Data availability

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

Acknowledgement

The study was supported by Key R&D and Promotion Special Project (Science and Technology Research) in Henan Province (232102210146)"

References

[1] X. Jing, Z. Yan, W. Pedrycz, Security data collection and data analytics in the internet: a survey, IEEE Commun. Surv. Tutorials 21 (1) (2019) 586–618.
[2] K.J. Singh, K. Thongam, T. De, Detection and differentiation of application layer DDoS attack from flash events using fuzzy-GA computation, IET Inf. Secur. 12 (6) (2018) 502–512.
[3] T. Subbulakshmi, S. Mercy Shalinie, C. Suneel Reddy, A. Ramamoorthi, Detection and classification of DDoS attacks using fuzzy inference system, Commun. Comput. Inf. Sci. 89 CCIS (2010) 242–252.
[4] N. Tabassum, M. S. Khan, S. Abbas, T. Alyas, A. Athar, and M. A. Khan, "EAI Endorsed Transactions Intelligent reliability management in hyper- convergence cloud infrastructure using fuzzy inference system,"vol. 4, no. 5, pp. 1–12.
[5] A.K. Soliman, C. Salama, H.K. Mohamed, Detecting DNS reflection amplification DDoS attack originating from the cloud, in: Proc. - 2018 13th Int. Conf. Comput. Eng. Syst. ICCES 2018, 2019, pp. 145–150.
[6] P. Arun Raj Kumar, S. Selvakumar, Detection of distributed denial of service attacks using an ensemble of adaptive and hybrid neuro-fuzzy systems, Comput. Commun. 36 (3) (2013) 303–319.
[7] A.S. Boroujerdi, S. Ayat, A robust ensemble of neuro-fuzzy classifiers for DDoS attack detection, in: Proc. 2013 3rd Int. Conf. Comput. Sci. Netw. Technol. ICCSNT 2013, 2014, pp. 484–487.
[8] L. Kwiat, C.A. Kamhoua, K.A. Kwiat, J. Tang, Risks and benefits: game-theoretical analysis and algorithm for virtual machine security management in the cloud, Assur. Cloud Comput. (2018) 49–80.
[9] H.S. Mondal, M.T. Hasan, M.B. Hossain, M.E. Rahaman, R. Hasan, Enhancing secure cloud computing environment by Detecting DDoS attack using fuzzy logic, in: 3rd Int. Conf. Electr. Inf. Commun. Technol. EICT 2017, 2018-Janua, 2018, pp. 1–4. December.
[10] P. Mishra, E.S. Pilli, V. Varadharajan, U. Tupakula, Intrusion detection techniques in cloud environment: a survey, J. Netw. Comput. Appl. 77 (October 2016) (2017) 18–47.
[11] R. Biswas, J. Wu, Filter assignment policy against distributed denial-of-service attack, Proc. Int. Conf. Parallel Distrib. Syst. - ICPADS 2018– Decem (2019) 537–544.
[12] S. Abbas, T. Alyas, A. Athar, M.A. Khan, A. Fatima, W.A. Khan, EAI Endorsed Transactions Cloud Services Ranking by Measuring Multiple Parameters Using AFIS, 2014, pp. 1–7.
[13] K. Iqbal, M. Adnan, S. Abbas, Z. Hasan, A. Fatima, Intelligent transportation system (ITS) for smart-cities using mamdani fuzzy inference system, Int. J. Adv. Comput. Sci. Appl. 9 (2) (2018) 94–105.
[14] R.L. Neupane, T. Neely, P. Calyam, N. Chettri, M. Vassell, R. Durairajan, Intelligent defense using pretense against targeted attacks in cloud platforms, Future Generat. Comput. Syst. 93 (2019) 609–626.
[15] T. Alyas, M.S. Khan, Intelligent reliability management in software based cloud ecosystem using AGI 17 (12) (2017) 134–139.

[16] N.S. Naz, S. Abbas, M. Adnan, B. Abid, N. Tariq, M. Farrukh, Efficient load balancing in cloud computing using multi-layered mamdani fuzzy inference expert system, Int. J. Adv. Comput. Sci. Appl. 10 (3) (2019) 569–577.
[17] Rudol, Implementasi keamanan jaringan komputer pada virtual private network (vpn) menggungakan, Implementasi Keamanan Jar. Komput. Pada Virtual Priv. Netw. Menggungakan Ipsec 2 (1) (2017) 65–68.
[18] W. Alosaimi, M. Alshamrani, K. Al-Begain, Simulation-based study of distributed denial of service attacks prevention in the cloud, Proc. - NGMAST 2015 9th Int. Conf. Next Gener. Mob. Appl. Serv. Technol. (2016) 60–65.
[19] N.C.S.N. Iyengar, G. Ganapathy, Chaotic theory based defensive mechanism against distributed Denial of Service Attack in cloud computing environment, Int. J. Secur. its Appl. 9 (9) (2015) 197–212.
[20] S.A. Miller, O. Behalf, C. America, CASE STUDY HYPERCONVERGENCE VS CLOUD, 2017, pp. 134–139.
[21] T. Alyas, M.S. Khan, Intelligent reliability management in software based cloud ecosystem using AGI 17 (12) (2017) 134–139.
[22] R.E. Spiridonov, V.D. Cvetkov, O.M. Yurchik, Data Mining for Social Networks Open Data Analysis, 2017, pp. 395–396.
[23] L. Wang, Y. Ma, J. Yan, V. Chang, A.Y. Zomaya, pipsCloud: high performance cloud computing for remote sensing big data management and processing, Future Generat. Comput. Syst. 78 (2018) 353–368.