Technological Forecasting & Social Change 177 (2022) 121554

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

A novel approach for DDoS attacks detection in COVID-19 scenario for small entrepreneurs

Akshat Gaurav [a], Brij B. Gupta [b,c,d,*], Prabin Kumar Panigrahi [e]

[a] Ronin Institute, Montclair, NJ, USA
[b] Department of Computer Science and Information Engineering, Asia University, Taichung 413, Taiwan
[c] King Abdulaziz University, Jeddah, Saudi Arabia
[d] Department of Computer Engineering, National Institute of Technology Kurukshetra, Kurukshetra, Haryana 136119, India
[e] Indian Institute of Management Indore, India

ARTICLE INFO

Keywords:
DDoS
Flash crowd
Entropy
Machine learning
Small entrepreneurs

ABSTRACT

The current COVID-19 issue has altered the way of doing business. Now that most customers prefer to do business online, many companies are shifting their business models, which attracts cyber attackers to launch several kinds of cyberattacks against commercial companies simultaneously. The most common and lethal DDoS attack disables the victim's online resources. While large businesses can afford defensive measures against DDoS assaults, the situation is different for new entrepreneurs. Their lack of security resources restricts their ability to ward off DDoS attacks. Here, we aim to highlight the problems that prospective entrepreneurs should be aware of before joining the business, followed by a filtering mechanism that efficiently identifies DDoS assaults in the COVID-19 scenario, which is the subject of our research. The suggested approach employs statistical and machine learning techniques to discriminate between DDoS attack data and regular communication. Our suggested framework is cost-effective and identifies DDoS attack traffic with a 92.8% accuracy rate.

1. Introduction

The European Commission, the EU's executive arm, defines micro, small, and medium-sized entrepreneurs (Che and Zhang, 2019; Gupta et al., 2013; Maroufkhani et al., 2020). This term applies outside the geographical area covered by the EU's authority (Commission, 1996). In response to the Council of Industry request, the European Commission suggested in 1992 that it restrict the definition of small and medium-sized entrepreneurs the Commission used in its expansion (Berisha and Pula, 2015). April 1996 saw the first proposal that laid the groundwork for a distinct definition of SMEs. The community and national level definitions might result in discrepency (Berisha and Pula, 2015).Table 1 summarizes the classification of small and medium-sized entrepreneurs according to European Union and World Bank.

COVID-19 is one of the world's most serious disasters, having claimed about 4,498,451 lives and displaced millions of people from many nations (Alowibdi et al., 2021; COVID-19, 0000; Sedik et al., 2021). COVID-19 has a significant impact on human life and requires company owners to migrate to an online platform (Alowibdi et al., 2021; Papadopoulos et al., 2020; Pashchenko, 2021; Rahman et al., 2021). Most large business owners have the knowledge and resources necessary to implement the proposed defense techniques against various cyber-attacks. On the other hand, small and medium-sized entrepreneurs lack the proposed knowledge and tools required to protect their online business platform from any type of cyber-attack (Carías et al., 2021; Millaire et al., 2017; Yeniman Yildirim et al., 2011).

DDoS attacks (Bhushan and Gupta, 2019; Chhabra et al., 2013; Dahiya and Gupta, 2021) are the deadliest of all kinds of cyber-attacks since they render the online business platform inaccessible to its consumers, making it impossible for them to do business. Small and medium-sized entrepreneurs suffer economic losses as a result of DDoS attacks, which raises the small and medium-sized companies' overall losses. As represented in Fig. 1, in a DDoS attack, the hacker utilizes a network of several number of hacked computers to create a large amount of fictitious traffic, depleting the victim's resources. In 2021, DDoS attacks were increased by 25% compared to 2020. Also, there were approximately 1392 DDoS attacks per day in 2021 (Fig. 2). India is the most affected country in the world due tovarious DDoS attack (Azure, 2021). Due to this, small commercial businesses find it very difficult to identify and prevent DDoS attacks due to their limited

* Corresponding author.
E-mail addresses: akshat.gaurav@ronininstitute.org (A. Gaurav), bbgupta.nitkkr@gmail.com (B.B. Gupta), prabin@iimidr.ac.in (P.K. Panigrahi).

https://doi.org/10.1016/j.techfore.2022.121554
Received 2 November 2021; Received in revised form 29 January 2022; Accepted 1 February 2022
Available online 3 February 2022

budgets.

Apart from DDoS attacks, small and medium-sized entrepreneurs must contend with the flash crowd. The crowded flash crowd is an instance in which many people simultaneously attempt to visit one specific website, which subsequently hampers the performance of that portal. Thus, if small and medium-sized businesses inadvertently filter out flash crowd traffic , this undermines their reputation and escalates their losses. An online shopping business in Australia recently had to pay a significant amount of refunds because it could not manage a large amount of traffic produced by consumers. Because the fundamental features of DDoS attacks and flash crowds are almost identical, it is difficult to distinguish between the two types of scenarios.

1.1. Challenges in the detection of DDoS attacks

There are many defensive techniques available against DDoS attacks (Gupta et al., 2020), but none of them fully resolves the issue. There are many reasons for this, including the following:

- Due to the abundance of open-source tools accessible on the internet, anybody may use them to launch an assault.
- DDoS attacks almost always include faked IP packets, making it perticularly impossible to pinpoint the source of the attack. Additionally, the length of an attack has decreased to approximately 4 minutes in recent years. Thus, the affected machine crashes before any protection solution can detect the attack. As a result, obtaining comprehensive information on DDoS attacks is very challenging.
- The lack of a standard benchmark for DDoS defense filters in the computer industry makes it impossible to compare defensive products with their counterparts on the market directly.
- Numerous new technologies are entering the market, including cloud computing, fog computing, industrial computing systems, and the Internet of Things. Thus, it is a difficult job to enhance traditional defensive techniques in such a manner that they can be used in these situations.

1.2. Contribution

The previous subsection explains the details about the small entrepreneurs and the effect of DDoS attacks on them. In this context, we proposed a DDoS detection approach for small entrepreneurs. The main contribution of our proposed approach is as follows: Our proposed approach is reactive; hence, it is more accurate and its response time is less than other recent DDoS detection techniques. Our proposed approach uses statistical and machine learning approaches to detect various DDoS attacks. Our proposed approach is economical; hence, small entrepreneurs efficiently use it.

1.3. Organization

The rest of the paper is organized as follows: Section 2 reviews the latest work in the field of DDoS and flash crowd detection. Section 3 gives the motivation of our proposed approach. Section 4 explains the components of our proposed approach, and Section 5 represents the simulation results. Finally, Section 6 concludes the paper.

2. Related work

Researchers proposed different security techniques (Gupta et al., 2021b; Masud et al., 2020; Nguyen et al., 2021; Rahman et al., 2021; Wang et al., 2020) for the identification and detection of DDoS attacks. In this section, we review some techniques proposed by the researchers for the detection of DDoS attacks.

The authors in (Mishra et al., 2021) suggested a low-cost defensive method against DDoS attacks based on differences in entropy between DDoS attacks and normal traffic. Additionally, the authors suggested a method for mitigating the attack's intensity. The suggested approach has three benefits over other current techniques: i) it has a high detection rate, (ii) it has a low false-positive rate, and (iii) it has the capacity to mitigate.

The authors (Gupta et al., 2021a) provide a method for addressing authentication and security problems associated with smart vessels in maritime transport. By authenticating devices in maritime transport and detecting different cyberattacks such as DDoS. The suggested method employs an identity-based technique to authenticate smart vessel access. However, this approach is applies to maritime transport only.

In this paper (Khan and Quadri, 2020), Khan et. al attempts to outline the issues that a potential entrepreneur should consider before entering the autonomous vehicle business, followed by innovative ideas that may assist in overcoming the essay's hurdles. The authors developed the proposals and guidelines using the cybersecurity principles of confidentiality, integrity, and availability.

The authors (Zhou et al., 2021a) present the construction of a secure data sharing method and a cyber-attack detection approach utilising

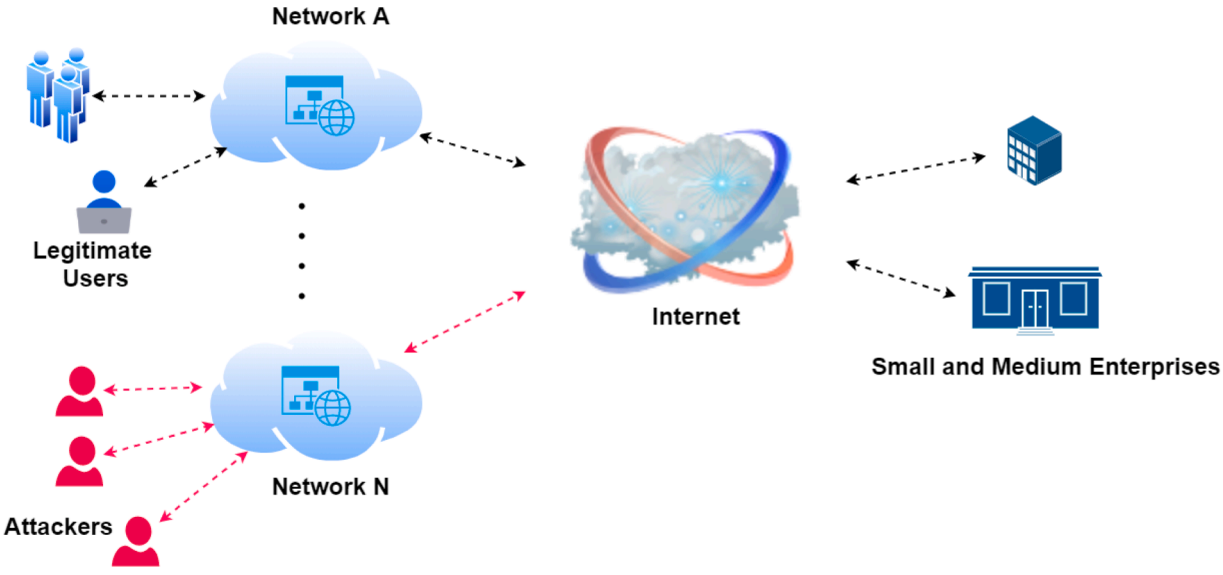


Fig. 1. DDoS attack in small and medium enterprises.

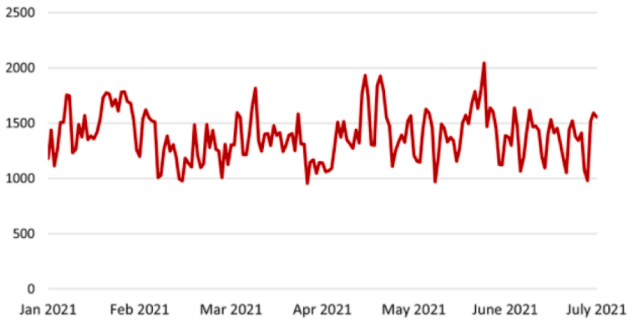
identity-based encryption (IBE) and deep learning algorithms in their study. The suggested system utilises identity-based encryption to handle access control to the VANET's smart cars, guaranteeing that no personally identifiable information is released. Deep learning technology evaluates network abnormalities and block malicious packets.

Attacks on a virtual machine monitor (VMM) that regulates the VMs may be identified by examining packet transmission data. Thus, the authors (Shidaganti et al., 2020) suggest a method in this article to halt such identified assaults at their source and analyze the proposed solutions for a few distinct kinds of such attacks. The authors suggest selective cloud egress filtering (SCEF) that includes modules for dealing with identified threats. If the process determines the attack, the SCEF notifies the VMM of which VMs are involved, allowing for targeted remediation.

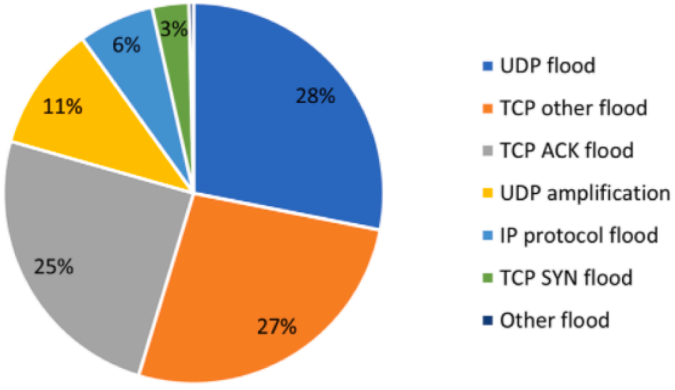
In this study, the authors' (Zhang et al., 2020) goal is to provide a broad framework for understanding the features of vulnerabilities in information systems, such as which category a particular vulnerability belongs to, the possible dangers it presents, and the critical indications for resolving it. Additionally, the authors gather data on actual vulnerabilities discovered in companies' information systems through a leading vulnerability report site. The system extracts four layers of features : word, phrase, subject, and record. The experimental findings demonstrate that the broad framework assists in characterizing the modes and patterns associated with different kinds of vulnerabilities. Based on boosting, authors in (Cvitić et al., 2021) proposed a DDoS detection technique in IoT devices . Authors () proposed RFID tags for mutual authentication in IoT devices. Author (A. Dahiya, 2021) proposed a game theory-based security mechanism during the COVID-19 scenario.

The authors (Dahiya and Gupta, 2020) suggested a method for mitigating DDoS assaults via a multi-attribute-based auction. Here, the authors proposed a reputation-based detection method in which the marginal utility of a user determines his/her reputation. Two distinct payment schemes for normal and fraudulent users have been suggested along with the identification method. In this method, a greedy resource allocation strategy distributes resources properly among authorised users. The differential payment system penalises malicious users who manipulate their offer to get the greatest share of restricted resources .

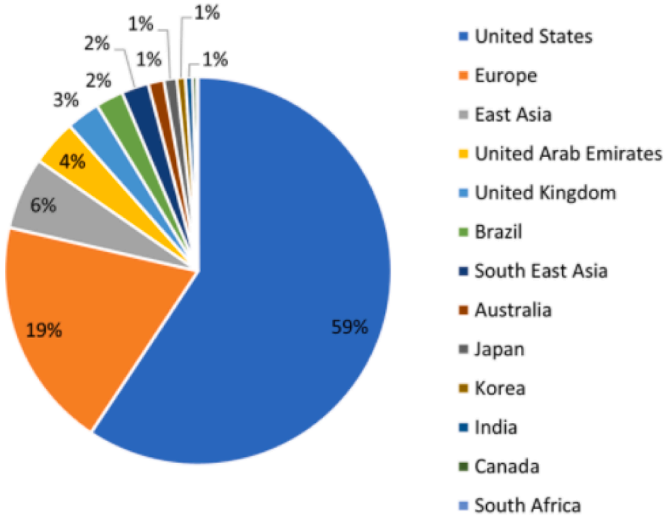
The authors (Dahiya and Gupta, 2021) offer a DDoS detection method based on Bayesian game theory. The service provider and legitimate users are supposed to monitor the network for an extended period and accumulate probabilistic information about whether or not another user is malevolent. The service provider and legal users use this probabilistic information to adjust their behaviour in response to malicious users on the network. Taking these assumptions and facts into account, the authors offer a Bayesian pricing and auction method for achieving Bayesian Nash Equilibrium points in various situations in which probabilistic knowledge benefits genuine consumers and service providers. Additionally, we offer a reputation evaluation and updating



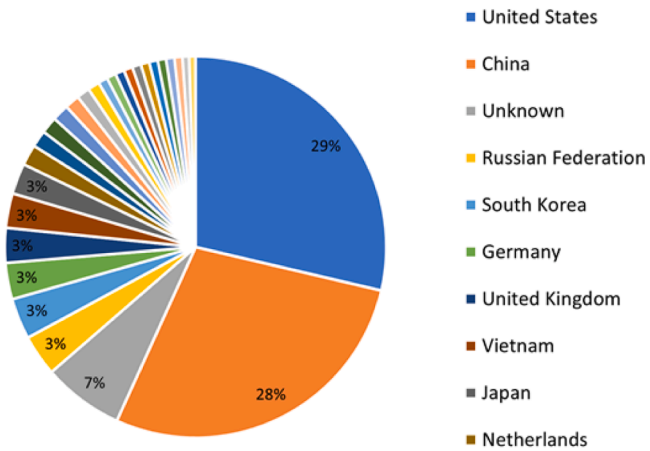
(a) Number of DDoS attacks



(b) DDoS attack vectors



(c) Affected countries by DDoS attack



(d) DDoS attack source

Fig. 2. DDoS statics by azure (Azure, 2021).

system that considers payment and participation characteristics when determining a user's trustworthiness.

Apart from the above-explained approaches, the researchers have propsoed many other techniques and methods for increasing security and privacy (Zhou et al., 2021b; 2019; 2021c; 2020). However, these approaches are not efficient in small and medium enterprises. Table 2 represents the comparison of some important approaches .

3. Motivation

As a result of the preceding explanation, we can see the potential damage caused by a DDoS attack. Two distinct attack techniques damage on small entrepreneurs: a bandwidth depletion attack and a resource depletion attack. In a bandwidth depletion assault, the attacker floods the small and medium entrepreneur network's available bandwidth with malicious packets; in a resource depletion attack, the attacker attempts to use all available resources on the small and medium enterprise network. Researchers have spent significant efforts on methods for detecting and mitigating DDoS attacks.

As mentioned before, the features of the flash crowd are similar to those of DDoS assaults; however, since the flash crowd traffic is produced by legal users, blocking this traffic may result in economic loss or diminished reputation for small and medium-sized entrepreneurs. Additionally, attackers may attempt to mimic the features of the flash mob to avoid detection filters. Numerous methods developed distinguish DDoS attacks from other types of attacks. These approaches identify flash crowd situations through the use of information theory-based methodologies. We discovered that when information theory-based techniques and statistical methods such as packet scoring are combined, the efficacy of information theory-based methods is enhanced since individual packet analysis is feasible. As a result, we focu our efforts on creating a filtering approach that uses information theory and static techniques. We used machine learning techniques to quickly and accurately identify different attack traffic.

4. Proposed methodology

This article provides a detection technique that uses entropy to find DDoS attacks for small and medium entrepreneurs. The two phases of our proposed method are as follows: the entropy calculation phase and the machine learning stage. The entropy calculation step determines the entropy of incoming packets.The second phase uses machine learning models to determine whether a packet is malicious or legal based on its entropy value.

4.1. Entropy calculation phase

This section explains the entropy calculation phase of our proposed approach.

Definition 1. Suppose there is a discrete random variable X with a range of n possible values; the probability of the occurrence of that discrete random variable is given by Eq. 1.

$$Probability = P(x) = \frac{P(xn_i)}{\sum_{i=1}^{n} xn_i} \tag{1}$$

Table 1
Classification of small and medium entrepreneurs (Berisha andPula, 2015).

Category	European Union		World Bank	
	Head Count	Turnover (million)	Total Employe	Total Assets (million)
Medium	<250	<€ 50	>50; ≤300	>$3; ≤ $15
Small	<50	<€ 10	>10; ≤ 50	>$0.1; ≤ $3
Micro	<10	<€ 2	< 10	≤ $0.1

Table 2
Comparison of different approaches .

Approach	Complexity	Detection rate	Remarks
(Mishra et al., 2021)	Low	High	SDN based
(Bhushan and Gupta, 2018)	Low	High	SDN based
(Zhou et al., 2021a)	Moderate	High	IBE based
(Gupta et al., 2021a)	Moderate	High	IBS signature based
(Tewari and Gupta, 2020)	Low	Moderate	RFID tags
(Cvitić et al., 2021)	Low	Moderate	Boosting based

Definition 2. Shannon provides a formal definition of the notion of entropy. An uncertainty of a random variable is measured in terms of the uncertainty of the variable. In the case of random variables, the entropy $\widehat{E}$(X) is suppressed in Eq. 2

$$\widehat{E} = -\sum_{i=1}^{n} P_i \times \log(P_i) \tag{2}$$

Theorem 1. *The entropy value, $\widehat{E}(X)$, is stationary, which means that it has the same value over two distinct time periods.*

Proof. Most of DDoS attackers use tools to launch their assaults to target small and medium enterprises. For faking the destination address of attack packets, these tools make use of a preset software that runs in the background. Consequently, we can show that there is a linear function that represents all of the faked addresses.

$$X_i = z(x) = lX_i + m$$

Therefore, the cluster probability is represented as:

$$p(x_i) = p(f(x_i)); i \in i, 2, \ldots N$$

from Eq. 2,

$$\widehat{E} = -\sum_{i=1}^{N} p(x_i)$$
$$\widehat{E} = -\sum_{i=1}^{N} p(f(x_i))$$
$$\widehat{E} = H(Y)$$

Where $Y \in \{(x_1), f(x_2), \ldots f(x_N)\}$, Thus, the cluster entropy associated with DDoS attack traffic may be described by a stable stochastic process.□

Theorem 2. *The graph of Entropy, $\widehat{E}$, variation is represented by a concave function.*

Proof. Let a function m(v) = - $j$log$j$ then

$$m' = -\log v - v \times \log e \times \frac{1}{j}$$
$$= -\log v - \log e$$

and

$$m'' = -\frac{1}{v} \times \log v < 0$$

The above equation is reduced as:

$$\widehat{E}(x) = \sum_{i=1}^{N} v(P(x))$$

Thus, we can say that entropy is a concave function□

Theorem 3. *Entropy of flash crowd is less than the DDoS attack scenario.*

Proof. Using Jensen's inequality, we can represent the DDoS attack

traffic by a monotonically increasing function.

$$Ef(x) \geq f(Ex) \tag{3}$$

The flash crowd probability distribution, on the other hand, is modeled as a monotonically growing concave function, using Jensen's inequality

$$Ef^{'}(x) \leq f^{'}(Ex) \tag{4}$$

If $P^1 = \{p_1^1, p_2^1 \ldots p_n^1\}$ represents the probability of different clusters during flash crow scenario and $P^2 = \{p_1^2, p_2^2 \ldots p_n^2\}$ represents cluster entropy. Packets are more randomly distributed during DDoS attack, so we can say that

$$P_i^1 < P_i^2, 1 \leq i \leq n \tag{5}$$

Therefore, from the above equations and the definition of entropy, we can say that

$$\widehat{E}^2(X) > \widehat{E}^1(X) \tag{6}$$

where $H^2(X)$ represents the entropy of the DDoS attack and $H^1(x)$ represents the entropy of flash crowd scenario.□

4.2. Machine learning phase

This phase consists of computing the entropy of the incoming traffic and organizing the data set for further analysis and usage in the next step. To determine which machine learning technique is the most successful in differentiating DDoS attack traffic from normal traffic, we will analyze our data set during this phase. We employed six of the most frequently used machine learning techniques when analyzing the data set, which are detailed in the following section.

4.2.1. Support vector machine

Vapnik et al.(Cortes and Vapnik, 1995; Vapnik, 1995) developed the support vector machine (SVM) (Zhang et al., 2004), a machine learning approach that is used in regression (Schölkopf et al., 2000; Smola and Schölkopf, 2004) and pattern recognition techniques'(Burges, 1998; Schölkopf et al., 2000). An SVM maps the data in the input space to a linear-separable high-dimensional feature space using the kernel mapping method (Scholkopf et al., 1999). The decision function of an SVM is proportional to the number of SVs and their weights, and it also links kernels chosen a priori, such as Gaussian and polynomial (Scholkopf et al., 1999; Smola et al., 1998). Suppose a and $b \in \mathscr{R}$ represents the input variable and output variable. Then, a linear estimate function transforms the input variable into the output variable .

$$b = f(a, w) = w^T \varphi(a) + c \tag{7}$$

where 'W' is the weight, φ is the non-linear mapping, and 'c' is constant. We have to find the optimal 'f' with the lowest error. Authors (Cortes and Vapnik, 1995; Vapnik, 1995; Zhang et al., 2004) suggest different ways to find optimal value of 'f' with the least amount of error.

$$Error = \frac{1}{2}\| w \|^2 + \frac{K}{z}\sum_{x=0}^{z} |b_x - f(a_x, w)|_{\epsilon} \tag{8}$$

Where 'K' is a constant, and ϵ is a small positive number. The last term in Eq. 8 is reduces to (Zhang et al., 2004)

$$|b - f(a, w)|_{\epsilon} = \begin{cases} 0, & if\ b - f(a, w) < \epsilon \\ |b - f(a, w)| - \epsilon & otherwisw \end{cases} \tag{9}$$

By using the Lagrange multiplier technique above defined two equations reduced to the following equation

$$f(x) = \sum_{i=1}^{j} (\beta_i^* - \beta_i) K(x - x_i) + c \tag{10}$$

Where K is SV kernel.

4.2.2. Logistic regression

LR used a linear (Eq. 11) to classify the data points. We used the sigmoid function (Eq. 12) to limit the linear equations' output (Eq. 13) .

$$Linerar\ equation\ (z) = \theta_0 + \theta_1.i_1 + \theta_2.i_2 + \ldots \tag{11}$$

$$sigmoid\ functiong(m) = \frac{1}{1 + e^m} \tag{12}$$

$$Output(y) = \frac{1}{1 + e^z} \tag{13}$$

4.2.3. Decision Tree Classifier (DTC)

This method divides the data set repeatedly according to a criterion that optimizes data separation, producing a tree-like structure. A decision tree structure is comprised of just two components: the Decision Node and the Leaf Node, as represented in Fig. 3.

A decision node has many branches, whereas a leaf node represents the decision's final output. When developing a Decision Tree, the most difficult problem is determining which attributes should be used for the tree's leaf node and decision node. In this case, the information gain method is used to make the selection, which involves partitioning the dataset and selecting the most appropriate choices for decision nodes and leaf nodes. To distinguish between leaf nodes and decision nodes in the information gain method, entropy is utilized as a distinguishing factor (algorithm 1).

$$IG = Entropt(D) - [w \times E(F)] \tag{14}$$

where E(D) is the entropy of the dataset, 'w' is the weight, and E(f) is the entropy of the feature.

4.2.4. Random forest

Random forest (RF) is a classification and regression technique based on ensemble learning that is particularly well suited for issues requiring data sorting into classes. Breiman and Cutler were the ones that came up with the algorithm (Breiman and Cutler, 2007). In RF, prediction is accomplished via the use of decision trees. During the training phase, multiple decision trees are built and then utilised for class prediction; this is accomplished by taking into account the voted classes of all individual trees during the training phase (Fig. 4). The classwith the largest margin function (MF) (Liu et al., 2012) is considered the output. The margin function is calculated by Eq. 15.

$$MF(a, b) = av_k I(h_k(X) = Y) - max_{j \neq y} av_k I(h_k(Xj) = j \tag{15}$$

4.2.5. Gradient boosting

One of the most effective machine learning algorithms is gradient boosting. When we speak about machine learning algorithms, we often consider them to have two basic kinds of errors: bias and variance. As one of the boosting methods, gradient boosting is used to reduce the model's bias error (Natekin and Knoll, 2013). The gradient boosting

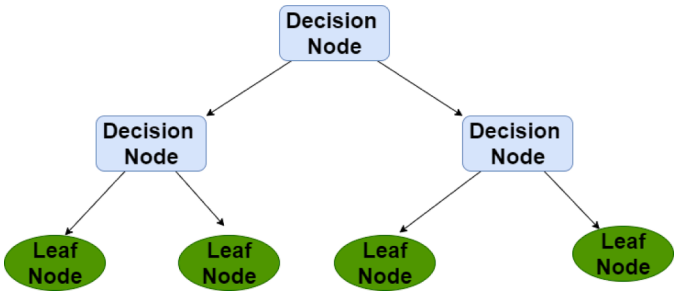


Fig. 3. Decision tree classifier.

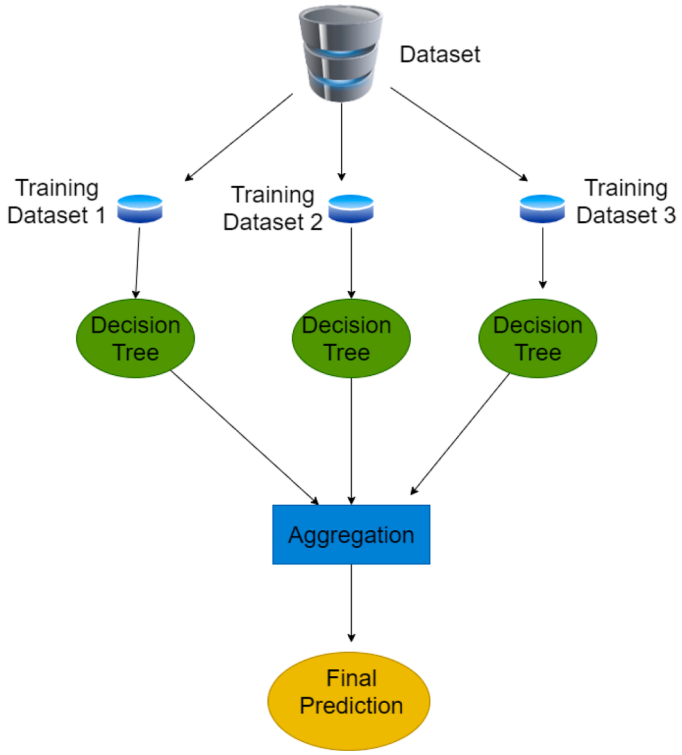


Fig. 4. Random forest technique.

technique may be thought of as a mathematical optimization procedure whose objective is to create an additive model with the smallest loss function. In this way, the gradient boosting method continuously increases the number of decision trees that decrease the loss function during each step. The gradient boosting method performs better if the contribution of the new decision tree is reduced at each iterative step using a shrinkage parameter 'β', referred to as the learning rate (Touzani et al., 2018). The shrinking method in gradient boosting is based on the principle that a greater number of tiny steps results in more accuracy than a smaller number of major activities. Algorithm 2 explains this procedure.

4.2.6. Multinomial naive bayes

NLP often employs the Multinomial Naive Bayes method for probabilistic learning, which is a strategy for probabilistic learning. For the independent variable, it is necessary to predict the tag using the Bayes theorem. It calculates the likelihood of each title in a sample and provides the label with the greatest chance of being correct.

4.3. Proposed algorithm

This subsection explains details of the algorithm applied at the gateway router of small and medium enterprises.

Because our suggested method is applied to the router, it can rapidly and efficiently detect the DDoS assault. Our suggested method is reactive, in the sense that it is capable of detecting and mitigating attack traffic in real time. The following are the stages in our suggested strategy:

- The method we offer examines the incoming traffic for each time frame δt.
- The attribute values of the packets are retrieved and aggregated for each time frame.
- Following that, the entropy value for the packets in the time frame is computed.
- The entropy value is then input into the trained machine learning model, which predicts whether a DDoS assault or normal traffic causes the incoming data.

Algorithm 3 gives the details of our proposed approach. The attributes used in the algorithm are explained in Table 3

5. Results and analysis

5.1. Dataset preprocessing

The simulation is conducted out using OMNET++. The attack packets overwhelm the victim with a huge amount of erroneous traffic in this scenario. In this instance, the attacker node produces packets every one second, while the genuine nodes emit packets every five seconds. The whole simulation takes 100 seconds to complete, and during that time, all log data is gathered. Because our suggested method is not protocol-specific, we simulate it using a generic routing protocol. Null values are removed from the dataset during the preparation step, since the dataset contains a large number of them. Fig. 5(b) and Fig. 5(a) illustrate the change in entropy between DDoS attack period and non-attack time. Finally, the dataset has been partitioned into a training and a testing set, allowing for an exhaustive evaluation of the proposed technique on both sets.

5.2. Machine learning techniques

This subsection compares six different machine learning techniques used to analyze the dataset prepared in the previous subsection. We calculate the following statistical parameters for the comparison.

- **Precision**- It determines the proportion of genuine packets in the overall number of forwarded packets.

$$Precision(P) = \frac{\delta T}{\delta T + \delta F} \tag{16}$$

- **Recall**- It quantifies the proportion of valid packets that are not rejected as a result of the suggested method.

$$Recall(R) = \frac{\delta T}{\delta T + \delta \widehat{F}} \tag{17}$$

- **Accuracy**- It assesses the adequacy of our suggested strategy.

$$Accuracy = \frac{\delta T + \delta \widehat{T}}{TotalPackets} \tag{18}$$

- **F-1 Score**- It assesses the effectiveness of the suggested strategy.

$$F-1score = 2 \times \frac{P \times R}{P + R} \tag{19}$$

where δT si True positive, δF is false positive, $\delta \widehat{T}$ is true negative, and $\delta \widehat{F}$ si false negative.

5.3. Confusion matrix calculation

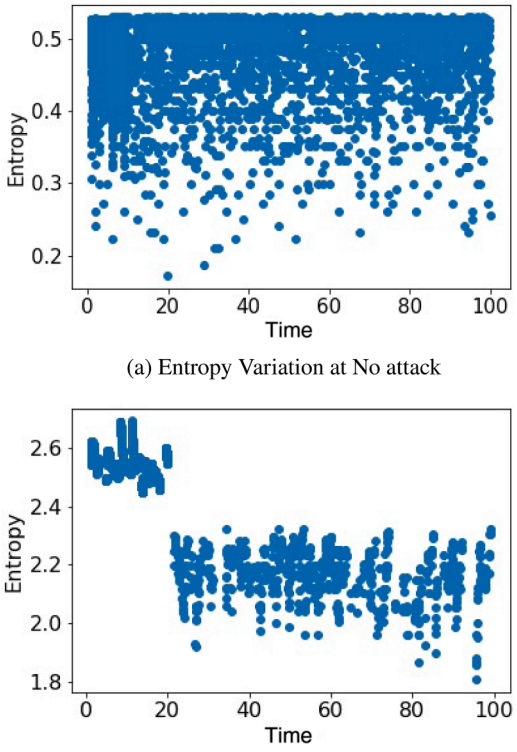
A confusion matrix (Fig. 6) is often used to assess the performance of the proposed model. In this matrix, the actual goal values are compared to the predictions produced by the machine learning model. This information provides us with a comprehensive picture of how well our classification model is doing as well as the kind of mistakes it is committing on a consistent basis. We can compute the precision, accuracy, recall, and f-1 score using a confusion matrix.

Input: Packet P_k
Output:DDoS attack or Flash Crowd
Start if $D_c > D_n$ **then**
  for *Packet* P_k *in time window* δt **do**
    $\hat{A}_i[k] \Leftarrow Pk$ packet score $PS_i(k)$
    entropy $\hat{E}_i(k)$;
    Pass the entropy value to Machine learning model
    if $\hat{E}_i[k]$ *is malacious* **then**
      return DDoS attack detected
    end
    else
      Flash crowd detected
    end
  end
end
End

Algorithm 3. Filtering.

Table 3
Attributes used in algorithm .

Term	Explanation
Pk	K^{th} incoming packet
Ai[k]	Kth attribute of i^{th} packet
H[k]	Entropy Value for k^{th} packet
D_n	Normal Data rate
D_c	Current Data rate



(a) Entropy Variation at No attack

(b) Entropy Variation at Attack

Fig. 5. Dataset representation.

5.4. State-of-art comparison

This part analyses several machine learning methods and determines which one is the most suitable for our suggested strategy. First, we compute the accuracy, precision, recall, and f-1 score for each of the six machine learning methods. The value of these statistical characteristics for machine learning methods is shown in Table 4. To better understand the results, we represent the statistical parameters in the graphical formation in Figure 7. Fig. 7 and Table 4 clearly show that the LR-based technique has the highest recall rate and accuracy. Hence, we can say that the LR-based technique can be used with our proposed approach to identifying DDoS attacks.

6. Conclusion

Following the COVID-19 pandemic, the way people work has changed dramatically. Because the majority of consumers now work online and prefer online sales and purchases, company owners are migrating to online platforms. This new transaction makes it simpler for hackers to steal users' sensitive information or disrupt the web platform's regular operations. Cyberattackers often use a DDoS attack due to its simplicity of deployment and ability to fully consume the target system's resources. The DDoS attack's objective is to bring the victim's

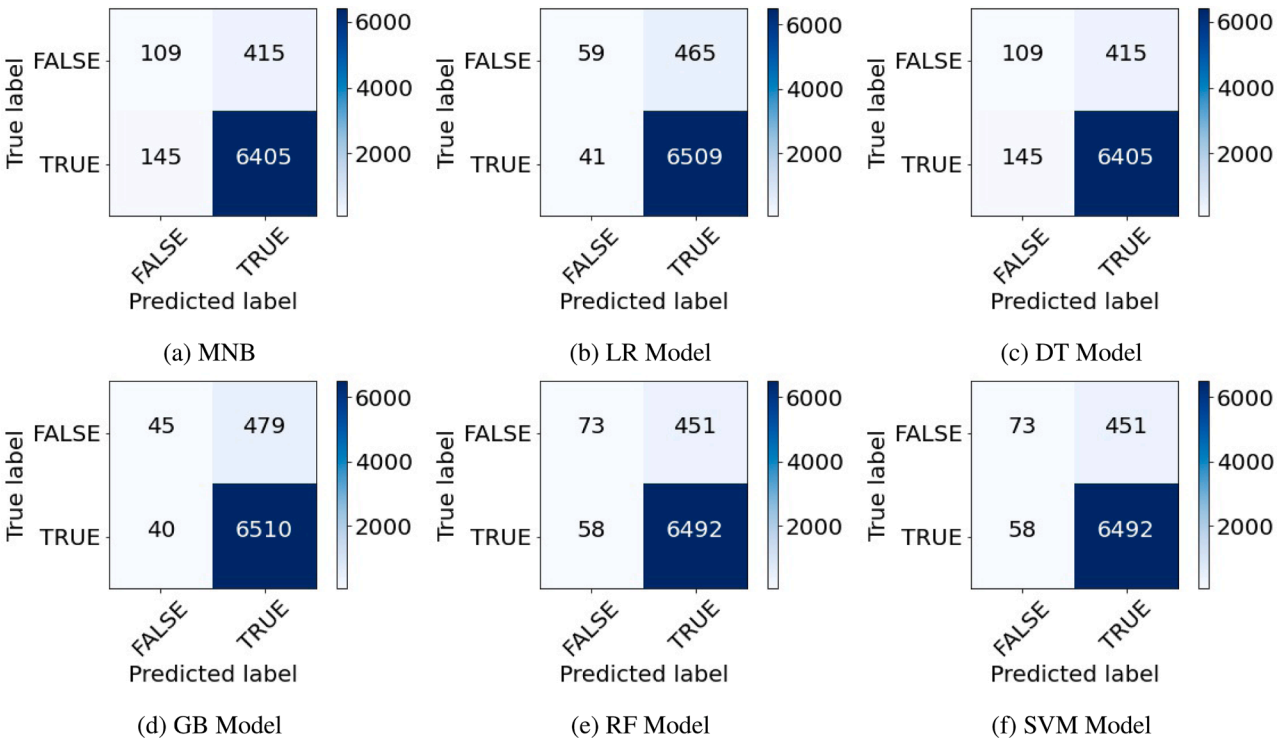


Fig. 6. Confusion matrix for different machine learning models.

Table 4
Comparison of different Machine Learning techniques .

Parameter	MNB	LR	DTC	GBC	RF	SVM
Accuracy	0.921	0.9284	0.92	0.9266	0.921	0.9280
Precision	0.94	0.93	0.94	0.93	0.94	0.94
Recall	0.98	0.99	0.98	0.99	0.98	0.99
F1 score	0.96	0.96	0.96	0.96	0.96	0.96

system to halt or to deplete its processing capacity. When a flash crowd is present, which is when real people generate large quantities of bandwidth, the DDoS attack becomes more difficult to detect. Given this, identifying DDoS attacks efficiently and accurately has long been a major research challenge. Due to the similarities between DDoS attacks and flash crowd, it is almost difficult to distinguish them. In this context, we present a method in this article that identifies DDoS assaults effectively and distinguishes them from the flash crowd for small and medium-sized entrepreneurs using entropy and machine learning. The

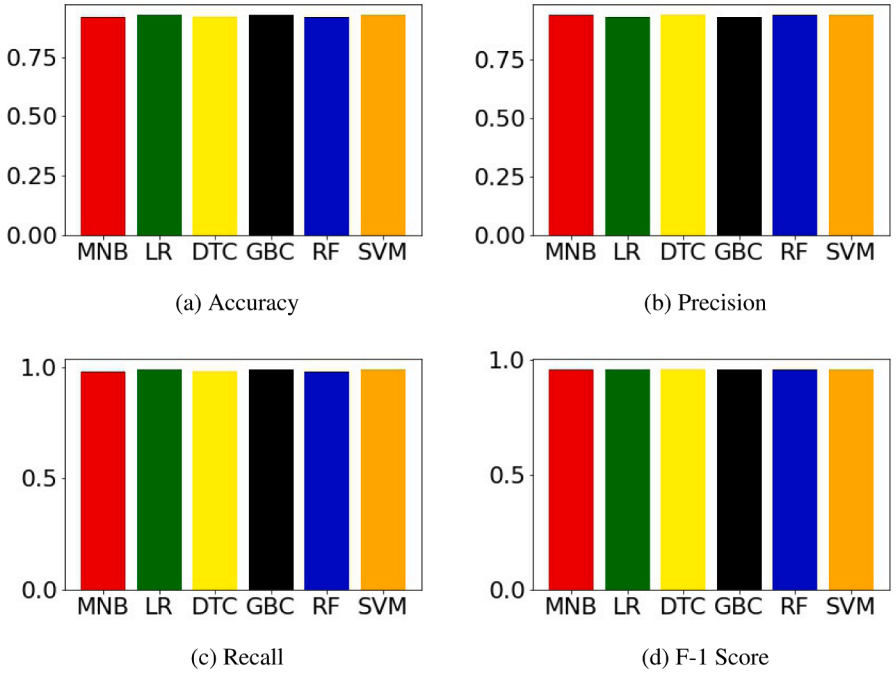


Fig. 7. Statistical parameters calculation.

Input: Dataset (D)
Output: Decision Tree creation
Start
Select the root node from dataset D
Create a queue containing the root element (A)
while *All elements of D are not analysised* **do**
  for *Each element in D* **do**
    Use Information gain to select whether a node is a decision node or leaf node
  end
end
End

Algorithm 1. Decision Tree Classifier.

Input: Dataset (D)
Output: Gradient Boosting
Start
Set initial estimation $\hat{f}$ as constant
Set number of iterations to k
Set the loss function $\phi(y, f(x))$
Set shrinking parameter α
Set base learner model h(x,θ)
for i=*1 to* K **do**
  Calculate gradient $(g_i(x))= \left[\frac{E\partial\phi(y, f(x)}{\partial f(x)}|x\right]_{f(x)=\hat{f}^{i-1}(x)}$
  Calculate optimal $\alpha_i = arg\ min_\alpha \sum_{j=1}^{P} \phi\left[y_j f_{i-1}(x_j)\alpha h(x_i, \theta_i)\right]$
  $\hat{f}_i \leftarrow \hat{f}_{i-1} + \alpha_i\ h(x, \theta_i)$
end
End

Algorithm 2. Gradient Boosting (Natekin and Knoll, 2013).

dataset was generated using the OMNET++ discrete event simulator and used to train six machine learning algorithms. The accuracy, precision, recall, and f1 score are used to determine the efficacy of machine learning techniques. On the datasets, certain models, such as LR, outperformed others, including DT, SVM, LR, MNB, RF, and GB, in terms of accuracy. We want to do further testing on a variety of data sets in the future.

CRediT authorship contribution statement

Akshat Gaurav: Conceptualization, Formal analysis, Writing – original draft. **Brij B. Gupta:** Conceptualization, Formal analysis, Writing – original draft, Supervision. **Prabin Kumar Panigrahi:** Conceptualization, Writing – original draft, Supervision.

References

A. Dahiya, 2021. Game Theory for Cyber Security during COVID-19 Pandemic: A Holistic Approach. https://insights2techinfo.com/game-theory-for-cyber-security-during-covid-19-pandemic-a-holistic-approach/. Online; accessed 29 November 2021.

Alowibdi, J.S., Alshdadi, A.A., Daud, A., Dessouky, M.M., Alhazmi, E.A., 2021. Coronavirus pandemic (covid-19): emotional toll analysis on twitter. International Journal on Semantic Web and Information Systems (IJSWIS) 17 (2), 1–21.

Azure, 2021. Azure ddos protection-2021 q1 and q2 ddos attack trends. https://azure.microsoft.com/en-us/blog/azure-ddos-protection-2021-q1-and-q2-ddos-attack-trends/. Online; accessed 29 November 2021.

Berisha, G., Pula, J.S., 2015. Defining small and medium enterprises: a critical review. Academic Journal of Business, Administration, Law and Social Sciences 1 (1), 17–28.

Bhushan, K., Gupta, B., 2018. Detecting ddos attack using software defined network (SDN) in cloud computing environment. 2018 5th international conference on signal processing and integrated networks (SPIN). IEEE, pp. 872–877.

Bhushan, K., Gupta, B.B., 2019. Distributed denial of service (ddos) attack mitigation in software defined network (sdn)-based cloud computing environment. Journal of Ambient Intelligence and Humanized Computing 10 (5), 1985–1997.

Breiman, L., Cutler, A., 2007. Random forests-classification description. Department of Statistics, Berkeley 2.

Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery 2 (2), 121–167.

Carías, J.F., Arrizabalaga, S., Labaka, L., Hernantes, J., 2021. Cyber resilience self-assessment tool (cr-sat) for smes. IEEE Access.

Che, Y., Zhang, B., 2019. Contextual determinants of e-entrepreneurship: opportunities and challenges. International Journal on Semantic Web and Information Systems (IJSWIS) 15 (3), 1–15.

Chhabra, M., Gupta, B., Almomani, A., 2013. A novel solution to handle ddos attack in manet.

Commission, E., 1996. Commission recommendation of 3 april 1996 concerning the definition of small and medium-sized enterprises. Official Journal 107, 0004–0009.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20 (3), 273–297.

COVID-19, W.,. Coronavirus disease (COVID-19) - World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

Cvitić, I., Peraković, D., Gupta, B., Choo, K.-K.R., 2021. Boosting-based DDoS detection in internet of things systems. IEEE Internet of Things Journal.

Dahiya, A., Gupta, B.B., 2020. Multi attribute auction based incentivized solution against ddos attacks. Computers & Security 92, 101763.

Dahiya, A., Gupta, B.B., 2021. A reputation score policy and Bayesian game theory based incentivized mechanism for DDoS attacks mitigation and cyber defense. Future Generation Computer Systems 117, 193–204.

Gupta, B.B., Dahiya, A., Upneja, C., Garg, A., Choudhary, R., 2020. A comprehensive survey on DDoS attacks and recent defense mechanisms. Handbook of Research on Intrusion Detection Systems 186–218.

Gupta, B.B., Gaurav, A., Hsu, C.-H., Jiao, B., 2021. Identity-based authentication mechanism for secure information sharing in the maritime transport system. IEEE Transactions on Intelligent Transportation Systems.

Gupta, B.B., Li, K.-C., Leung, V.C., Psannis, K.E., Yamaguchi, S., et al., 2021. Blockchain-assisted secure fine-grained searchable encryption for a cloud-based healthcare cyber-physical system. IEEE/CAA Journal of Automatica Sinica.

Gupta, P., Seetharaman, A., Raj, J.R., 2013. The usage and adoption of cloud computing by small and medium businesses. International Journal of Information Management 33 (5), 861–874.

Khan, M.K., Quadri, A., 2020. Augmenting cybersecurity in autonomous vehicles: innovative recommendations for aspiring entrepreneurs. IEEE Consumer Electronics Magazine 10 (3), 111–116.

Liu, Y., Wang, Y., Zhang, J., 2012. New machine learning algorithm: Random forest. International Conference on Information Computing and Applications. Springer, pp. 246–252.

Maroufkhani, P., Tseng, M.-L., Iranmanesh, M., Ismail, W.K.W., Khalid, H., 2020. Big data analytics adoption: determinants and performances among small to medium-sized enterprises. International Journal of Information Management 54, 102190. https://doi.org/10.1016/j.ijinfomgt.2020.102190.

Masud, M., Gaba, G.S., Alqahtani, S., Muhammad, G., Gupta, B., Kumar, P., Ghoneim, A., 2020. A lightweight and robust secure key establishment protocol for internet of medical things in covid-19 patients care. IEEE Internet of Things Journal.

Millaire, P., Sathe, A., Thielen, P., 2017. What all cyber criminals know: small & midsize businesses with little or no cybersecurity are ideal targets. NJ, USA.

Mishra, A., Gupta, N., Gupta, B., 2021. Defense mechanisms against ddos attack based on entropy in sdn-cloud using pox controller. Telecommunication systems 77 (1), 47–62.

Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Frontiers in neurorobotics 7, 21.

Nguyen, G.N., Le Viet, N.H., Elhoseny, M., Shankar, K., Gupta, B., Abd El-Latif, A.A., 2021. Secure blockchain enabled cyber–physical systems in healthcare using deep belief network with resnet model. Journal of parallel and distributed computing 153, 150–160.

Papadopoulos, T., Baltas, K.N., Balta, M.E., 2020. The use of digital technologies by small and medium enterprises during covid-19: implications for theory and practice. International Journal of Information Management 55, 102192.

Pashchenko, D., 2021. Fully remote software development due to covid factor: results of industry research (2020). International Journal of Software Science and Computational Intelligence (IJSSCI) 13 (3), 64–70.

Rahman, M.A., Hossain, M.S., Alrajeh, N.A., Gupta, B., 2021. A multimodal, multimedia point-of-care deep learning framework for covid-19 diagnosis. ACM Transactions on Multimidia Computing Communications and Applications 17 (1s), 1–24.

Scholkopf, B., Mika, S., Burges, C.J., Knirsch, P., Muller, K.-R., Ratsch, G., Smola, A.J., 1999. Input space versus feature space in kernel-based methods. IEEE Transactions on Neural Networks 10 (5), 1000–1017.

Schölkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L., 2000. New support vector algorithms. Neural computation 12 (5), 1207–1245.

Sedik, A., Hammad, M., Abd El-Samie, F.E., Gupta, B.B., Abd El-Latif, A.A., 2021. Efficient deep learning approach for augmented detection of coronavirus disease. Neural Computing and Applications 1–18.

Shidaganti, G.I., Inamdar, A.S., Rai, S.V., Rajeev, A.M., 2020. Scef: a model for prevention of ddos attacks from the cloud. International Journal of Cloud Applications and Computing (IJCAC) 10 (3), 67–80.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and computing 14 (3), 199–222.

Smola, A.J., Schölkopf, B., Müller, K.-R., 1998. The connection between regularization operators and support vector kernels. Neural networks 11 (4), 637–649.

Tewari, A., Gupta, B.B., 2020. Secure timestamp-based mutual authentication protocol for iot devices using rfid tags. International Journal on Semantic Web and Information Systems (IJSWIS) 16 (3), 20–34.

Touzani, S., Granderson, J., Fernandes, S., 2018. Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy and buildings 158, 1533–1543.

Vapnik, V.N., 1995. The nature of statistical learning. Theory.

Wang, H., Li, Z., Li, Y., Gupta, B.B., Choi, C., 2020. Visual saliency guided complex image retrieval. Pattern recognition letters 130, 64–72.

Yeniman Yildirim, E., Akalp, G., Aytac, S., Bayram, N., 2011. Factors influencing information security management in small- and medium-sized enterprises: a case study from turkey. International Journal of Information Management 31 (4), 360–365. https://doi.org/10.1016/j.ijinfomgt.2010.10.006.

Zhang, L., Zhou, W., Jiao, L., 2004. Wavelet support vector machine. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 34 (1), 34–39.

Zhang, X., Xie, H., Yang, H., Shao, H., Zhu, M., 2020. A general framework to understand vulnerabilities in information systems. IEEE Access 8, 121858–121873. https://doi.org/10.1109/ACCESS.2020.3006361. Conference Name: IEEE Access.

Zhou, Z., Gaurav, A., Gupta, B.B., Lytras, M.D., Razzak, I., 2021. A fine-grained access control and security approach for intelligent vehicular transport in 6g communication system. IEEE Transactions on Intelligent Transportation Systems.

Zhou, Z., Li, Y., Zhang, Y., Yin, Z., Qi, L., Ma, R., 2021. Residual visualization-guided explainable copy-relationship learning for image copy detection in social networks. Knowledge-Based Systems 228, 107287.

Zhou, Z., Mu, Y., Wu, Q.J., 2019. Coverless image steganography using partial-duplicate image retrieval. Soft Computing 23 (13), 4927–4938.

Zhou, Z., Su, Y., Zhang, Y., Xia, Z., Du, S., Gupta, B.B., Qi, L., 2021. Coverless information hiding based on probability graph learning for secure communication in iot environment. IEEE Internet of Things Journal.

Zhou, Z., Wu, Q.J., Yang, Y., Sun, X., 2020. Region-level visual consistency verification for large-scale partial-duplicate image search. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16 (2), 1–25.

Akshat Gaurav received the M.Tech degree in Computer Engineering (cyber security) from NIT Kurukshetra, India. His research interests include information security, cyber security, cloud computing, web security, intrusion detection, and computer network.

B. B Gupta earned a Ph.D. degree in information and cybersecurity from IIT Roorkee, India. He has published more than 400 research articles (including four books and 20 book chapters) in international journals and high-repute conferences, including the IEEE, Elsevier, ACM, Springer, and Inderscience. His research interests include cybersecurity, information security, smartphone, web security, cloud computing, computer networks, intrusion detection, and phishing.

Prabin Kumar Panigrahi is a professor of Information Systems department at the Indian Institute of Management Indore, India. He earned his Ph.D from Indian Institute of Technology, Kharagpur. His research interests include Emerging Technologies, Machine

Learning, Cyber Security, Cyber Laws, Information Security, Text Mining of Vernacular Languages, Technology Adoption, e-Governance, e-Learning, e-Participation, Social Inclusion in Information Systems, and Business Value of Information Systems.