International Journal of Advanced Engineering and Management Research

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Cyber-attack Detection and Isolation

Chetanpal Singh[2], *, Rahul Thakkar[2], Jatinder Warraich[1]
[1]Faculty of business, design and it, Holmesglen Institute Chadstone Campus Melbourne Australia.
[2]Faculty of ICT, Victorian Institute of Technology (VIT), Melbourne Australia.

doi.org/10.51505/ijaemr.2024.9110 http://dx.doi.org/10.51505/ijaemr.2024.9110

Received: Feb 8, 2024 Accepted: Feb 19, 2024 Online Published: Feb 23, 2024

Abstract
The research is depending on the increasing the rate of cyber-attacks within the issues in current platforms of an organizational. Companies are using various types of methods to minimize the chance of cyber-attacks. It is essential to help the criteria of the management that can help to establish various plans that can help to determining and controlling in the research. Companies are trying to establish strong firewalls that can reduce the hackers from the cyber-attack. It is necessary for the organization to establish such the attention that help the company to detect the basic ideas to handle the cyber-attacks. It is essential to develop to analyse that help the company to prevent issues.

Keywords: Cyber-attack, Data isolation, Malware detection, Firewall, DDoS Attack

I. Introduction

1.1 Background

Industrial Control System is generally used in controlling and operating different industrial procedures in various production units, chemical plants, hydro-generation plants and other foreign companies. Current industries are facing cyber-attacks due to the vulnerability of the organizational storage system. Organizations like a company, schools, universities or public and private firms have confidential data related to the stakeholders and internal management like employee information, customer information, strategic goals, financial information, and other significant data (4). All these data can be stolen due to cyber-attacks and organizational management, creating danger for the organisation's existence. The cyber-attack does not take place only in large and shared devices; it can also take place in personal systems. It turns out to be dangerous, and there is a constant fear of data theft. Thus, cyber-attack detection and isolation can play the most important role in reducing the situational crisis and safeguarding the storage network.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

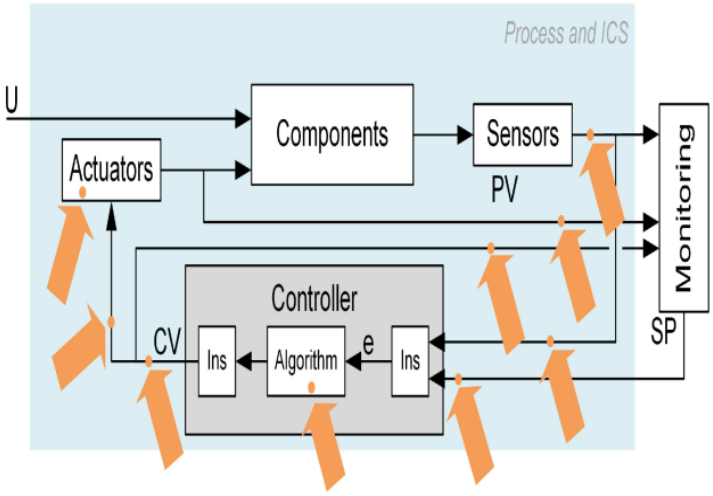


Figure 1: Cyber-attack Sensors
Source: [2]

Moreover, cybernetic faults also affect the system's performance by entering into the control system and being destructive in internal operations. It is also necessary to create awareness among users to prevent cyber-attacks and promote isolation. Poor internal management can also lead to a problem and upcoming external disturbance (5). Additionally, the users are likely to avoid creating program management that may help create different landscapes as it may determine negative influence in the social and organizational setup.

1.2 Research Focus

The research focus of the topic is related to the increasing rate of cyber-attacks along with problems in current organizational platforms. Companies are using different methods to reduce the chance of cyber-attacks. It is important to help other criteria management that can help the company to create different plans that can help in determining and controlling the focus of the research. Companies and personal devices are trying to develop strong firewalls that can prevent hackers from the attack (3). It is also necessary for the organization to create such attention that may help the company detect the basic ideas to control cyber-attacks. It is important to develop a proper analysis that may help the company to prevent issues. However, cyber-attack it is also important to create an appropriate plan that can determine the necessary aspects in choosing the best way to reduce cyber-attacks, so it is also important to create different issues and manage the best solution in addressing the cyber-attack. Thus, the company needs to manage and determine the best ways that may help in order to create the best way that can also help in making the best way to determine the cyber-attacks (7). On the other hand, it can also help create the best way to control cyber-attacks.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

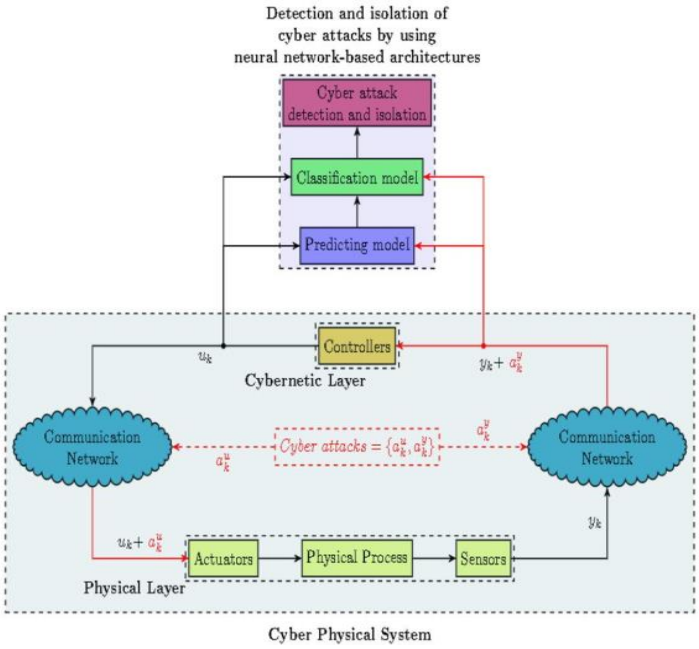


Figure 2: Detection and Isolation
Source: [2]

It is also important for the organization to plan something that may prevent any form of cyber-attack. On the other hand, it is also important for the organization to manage different situational crises so that it may help in controlling maximum seizures. The paper focuses on exploring various cyber-attacks in organizational and personal systems. The current situation of the pandemic has developed the trend of remote working, and different reports have suggested that remote working will persist for a longer time (1). Thus, it creates a new cyber security trend as the home environment is less prorated than an office. However, a maximum number of organizations are sending official systems to maintain security, though there is a slight chance that the company may face cyber security threats. While working from home, the employees connect to their domestic router or any data sources, leading to problems and a higher tendency of cyber-attacks. All the workplaces are not equipped enough to offer laptops or desktops to all the candidates; in that case, the employees are browsing the official data from home, enhancing the chance of cyber-attacks.

The Use of the Internet of Things is increasing in numbers, like the eerie body is using smart watches, smart TV, smart refrigerator and many more. Thus, internet connection and data sharing options are there in all these devices. However, these devices can only provide a little security like smart phones, laptops and desktops. As per the record, there will be 64 billion IoT devices being used around the globe (5). Users want all things to be easy, and the tendency to work from home has increased the usage of such devices. If there are more than one deuces connected to the internet, it will in case the attacking zone and lead to a cyber-attack.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

While digitalization can be good in increasing the efficiency of an organization, it can raise the chance of attacks as well. It has been seen that there are 120 separate families of ransom ware (8). It eventually makes it easier for hackers to steal data from the system. Digitalization of the organization and remote working trends have increased the chances of ransom ware as all these are becoming the targets. Such attacks can put the organization at severe risk as hackers are stealing the data from the system and encrypting it to make it inaccessible to the users. After that, the hackers control the data and blackmail the person or the organization for money and other reasons (10). Apart from that, it is also important to determine different ways to help people that can help determine if there is any operational aspect that can help the company create some external management through sustainable management operations. Ransom ware attackers have also developed different sections that may help in mitigating the issues related to cyber-attacks.

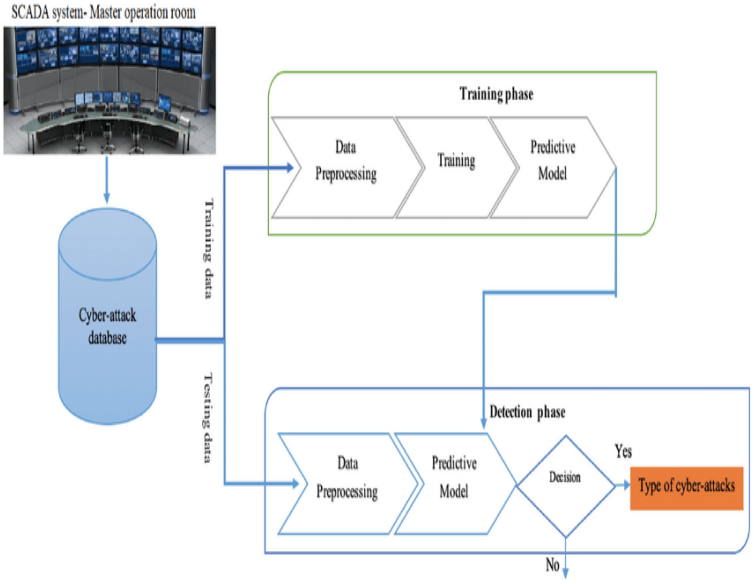


Figure 3: Flowchart of Cyber-attack detection
(Source: 13)

On the other hand, the attackers are becoming sophisticated with the application of machine learning with more coordinated action in the dark web. An attack of ransom ware may help in increasing good determinants that may also create a certain possibility that can also help in creating massive deduction that may help in exploring the ways to prevent any chance of cyberattacks (12). The companies are trying to resolve the issue, which can cost almost 3.86 billion dollars. Extended interest in cloud computing can also increase the potential that may help in determining factors, and it can also help the company to create potential challenges in an organizational setup.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

1.3 Aims and Objectives

The research aims to explore different areas related to cyber security attacks and isolation. The following research questions will be explored throughout the study.

- What is the concept of cyber-attacks and cases in real-life organizations?
- What can be the solution for isolation after a cyber-attack?
- How can it impact a cyber-attack in the current storage system?
- What actions can be taken to reduce the impact of cyber-attacks and isolation?

The objectives of the research will be

- To explore the concept of cyber-attacks and cases in the real-life organization
- To find out a solution for isolation after a cyber attack
- To examine the impact of cyber-attacks on the current storage system
- To determine the actions that can be taken to reduce the effects of cyber-attacks and isolation

1.4 Research Motivation

Cyber-attacks have turned out to create such an impact in the current organizational setup that every organisation is looking for a solution to reduce cyber-attacks through isolation. Personal devices are also trying to develop options that may help in reducing the impact, and it can also help in managing the security of each of the factors that may help the company to create major options creating cyber security (11). Apart from that, it is also important for all organizations to stay alert and to keep all the employees aware to help the company benefit from the current organizational structure. Security attacks are becoming rapid, and every time, hackers come coming with new ideas that may create issues in following the organizational structure, data management and storage management. In that case, it may need to be improved in leading the organization with a secure platform (9). Apart from that, it may help the company create different plans, which may help the company overcome such issues. These factors are motivating to take this research so that it can help in determining the maximum benefit to the individuals and overall organizations.

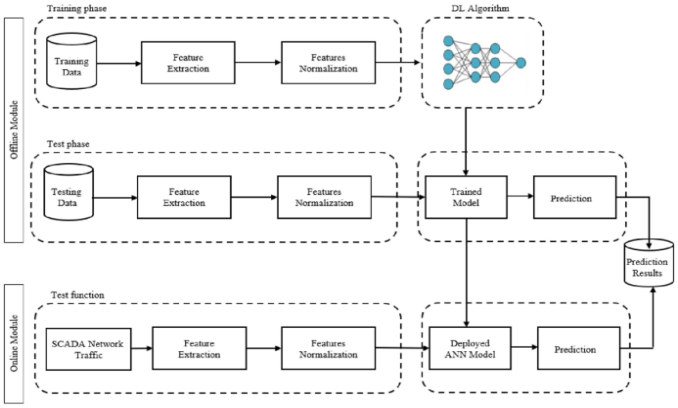


Figure 4: Intrusion Detection

(Source: 7)

Vol. 9, No. 01; 2024

ISSN: 2456-3676

According to the current digital ecosystem, every organization of any size has all information related to operations, brands, reputation and revenue share in the device. All are explored on the internet. Thus, it is important to manage and create rules that help the company be safe from cyber security. If any cybercrime occurs, it will become hard for the company to make rules that may orient the cybercrime. Loss of data can affect the reputation of the company at stake. Thus, it is important to make such rules that can help the company to determine different regulations in cyber security (2). The use of Artificial intelligence and machine learning is quite dominant according to the current trends. In that case, the company must ensure they are using all these devices actively by maintaining the needs. AI tools are widely used by hackers for advanced attacks. Intense fakes are already deployed. Current situation of war, pandemic and all those in between, it is becoming hard to protect data from any cyber-attacks. Cyber-attacks have turned so massive that during the last year, 34.5% of companies have suspected that hackers have targeted financial data. Among that, 22% have faced the attack, and 12.5% have faced more than one attack (14). However, the financial team is doing nothing to prevent such attacks. Maximum attacks are taking place in the accounting department, and loss of financial data can cause massive exploitation. Artificial intelligence is growing in the cybersecurity market, and it is growing at a CAGR of 23.6%, and it can be estimated that it can reach a market value of 46.3 billion dollars in 2027 (3). Thus,it can be seen from the analysis that all information available in the organizational database is not safe. It can be motivating to explore and determine ways to control and create different platforms to save the data.

II. Literature Review

Analyse the concept of Cyber-attack and its' impact on real-life organizations

COVID-19 impacted negatively on human lives, as well as organizational business processes. To stop the infection, people were advised by the WHO to stay at home and follow government rules and regulations. Different business institutions, schools, and shops closed to handle the virus expansion. According to [16], during the global pandemic, the rate of cyber-attacks increased innumerably. Based on the report of ILO (International Labour Organization), it has been noted that approximately 2.7 billion workers faced pressure due to the lockdown, and people became more dependent on the online process for fulfilling their daily purposes. Cyber threats were continuously evolving to take various advantage of digital trends. Based on the report of "*UK National Fraud & Cyber Security Centre*", it has been estimated that, during the global pandemic, in 2020, cyber-fraud related cases enhanced by almost 400%, while victims lost approximately 800 thousand pounds just in one month.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

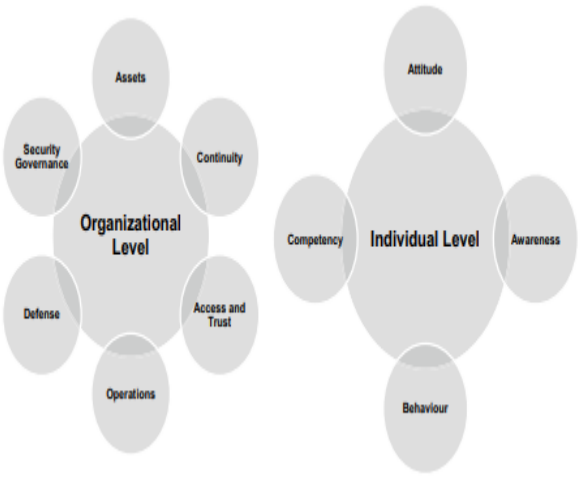


Figure 5: Cybersecurity culture model
Source: [16]

The above image defines key factors influencing as well as generating cybersecurity culture within the organization. This model defines two different levels, concluding organizational, as well as individual. Each one was divided by different dimensions that are considered various domains, with quantifiable indicators, as well as application areas.

Based on the previous statement, the other author [17] claimed that, during the quarantine period, major turn to the Internet, and work-from-home approach, cyber-attacks had increased dramatically. The social distancing approach increases the cyber criminality rate, as during the pandemic, it should be easy for attackers to harass their victims constantly. To make life easy, most people take advantage of cyberspace, which supports easier access for a major number of people, basically young people, and even to different illegal activities also. Even though people get various advantages from the network system, then on the other side, they also face major issues due to cyberspace. So it has been noted that the technology that makes human lives easier, while on the other side, it can exploit modern life also. In this article, it has been reported that it is important to take proper action and track the increasing rate of cyber threats, as it disrupts the privacy and security of internet users.

During the global pandemic, the other word has also become popular, and that is "***World Wide Web***". Along with this, in 2020, WHO also used the word "***Infodemic***", which meant the overmuch information that concluded accurate information as well as fraud information. During the pandemic period, almost every person became dependent on cyberspace; due to that reason, the rate of disinformation increased, and due to cyber-attacks, electronic mailboxes across the world received a major amount of fake emails based on the COVID-19 pandemic. By utilizing infodemic, cyber-attacks took extra advantage and sent mail to administrative staff to steal organizational information as well as financial information. According to [18], in an organizational system, different information is stored, concluding business process-based

Vol. 9, No. 01; 2024

ISSN: 2456-3676

information as well as employee-related information. The biggest threat to an organization is privacy and security loss. In an organizational security chain, employee security awareness is one of the most fundamental key facts. Through accessing the organizational system, it can be easy for hackers to damage the system, hack all important documents, and disrupt the organizational reputation. It has been reported that, during the pandemic period, the proportion of cyber-attacks has increased by 35%, and with this, some hackers also utilized different types of innovative approaches, concluding ML and AI-based techniques. Cyber attacks can cause loss of organizational money, threaten brand reputation, and with this utilize the information for illegal activities also, that negatively impact the organizational future business process also.

Identifying various solutions to prevent cybersecurity attacks

In the present age, basically from the COVID-19 period, it has been noted that the rate of cyber attacks is increasing constantly, and not just people but also business institutions faced various disruptions due to informational loss. In that situation, to overcome those threats, almost every business organizations adopt different types of innovative approaches so that it can be easy to detect fraudulent activities in the system and, with this, take proper actions also against the cyber attacks. According to [19], in the present technological era, ML (Machine Learning) has been defined as the most effective technique to avoid major cyber attacks. The constant development of Artificial Intelligence promotes enhanced security against computer attacks. Phishing attacks can increase major threats to business processes but can be mitigated through AI technology. In this article, the author provided the importance of AI technology to increase cyber security awareness and, with this, analyzes how it may influence various cyber-attacks. Among various types of cyber attacks, a phishing attack is one of the most common cyber security attacks, and AI has the capability to detect phishing activities and, with this support, take proper action.

The author indicated that AI technologies have the capability to produce significant results to avoid attacks. Most organizations increase their focus on “AI-based Cyber security” systems to protect administrative systems. The approach confirms that companies have mitigated cyber breaches in their organizations. Google produced an effective system. That had the ability to detect email phishing and avoid those to prevent users. The AI's ability to mitigate cyber attacks is featured through AI reasoning. ML (Machine Learning) technology development has also defined Cyber security. It can read real-time data and, with this, ensure that users are safe. Additionally, Thi systems have the capability to identify key threats as well as develop different processes through which cyber attacks could be mitigated. The major advantage of utilising AI and ML techniques in the system is that those can understand attack patterns and, with this, decide what types of approaches should be utilized to avoid those. In the current age, with technological innovation, hackers also utilize different types of innovative technologies, but AI can easily detect new threat patterns and, with this, handle vast amounts of data also.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Figure 6: The Priorities of Cyber security Approaches in Manufacturing
Source: [20]

To develop business process efficacy, most business institutions transfer their traditional process to digitalization, but with this, they face various types of new challenges also in their service processes. According to [20], Cyber security is an essential facilitator of the digitalization process, but if it can be managed properly, then it can be easy to get positive outcomes and develop organizational processes also. Based on the findings of this research process, it has been noted that the author defined that almost all security experts increase their aware for measuring security breaches. Effective cyber security management is important for handling the digitalization process. To ensure security concerns, possible actions should be taken, concluding with implementing a cloud security system and adopting IT infrastructure.

The utilization of innovative techniques provides incredible opportunities to society, organizations, as well as governments. But with this, ***digital fear*** is also becoming an essential concern in the current age. According to [21], the ***"Intrusion Detection System"*** has been developed to manage digital faults and make the organizational system more secure. In this system, ML technology has been utilized to develop a detection rate, as well as adaptability. In the present age, cyber-security and, with this protecting against malware attacks are becoming essential questions. The main reason behind utilising the computer system is to fulfil professional, as well as personal factors. In that situation, to handle the cyber-attack smoothly, proper actions should be taken, and IDS ("Intrusion Detection System") has the capability to identify infectious activities, as well as policy violations within the system. An IDS has the capability to identify abnormal behaviour in the network during the daily activities in a system. Along with this, an IDS also supports to locate, determine, and with this manage unauthorized network behaviour. Additionally, it has also been defined in this article that, among various types of IDS, basically two types of IDS have been utilized, and are "***Network intrusion detection***

Vol. 9, No. 01; 2024

ISSN: 2456-3676

system" or NIDS, and another is "***Host intrusion detection system***" or HIDS. The main advantage of utilizing IDS is that it analyses various attacks, detects attack patterns, and with this, supports corporate managers in setting and implementing appropriate control systems against attacks.

Importance of isolation solution after cyber-attack

After a cyber-attack, protocol isolation prevents damage from attacks, and with this, stops them from further attacks so that it can be easy to set a limitation of the attacker's movement. The fundamental purpose of utilizing web isolation technology is that it can ensure that no malicious content reach in the corporate network. According to [22], in the cyber-attack detection process, an isolation mechanism can prevent the network from cyber threats, as well as infections. The network system is not an easy process, and due to high complexities, often, it can be difficult to handle fraudulent activities. In that situation, it should be essential to utilise an isolation technique that has the capability to manage the virtualized environment. Normally, NSaaS ("Network Software as a service") may provide different isolation levels, concluding "*application segmentation isolation*", "*virtual machine isolation*", "*network segmentation isolation*", and with this ", *resource isolation.*" to ensure high-level security in the infrastructure of "***Mobile network operator***" (MNO), and with this to prevent the tent's privacy, and as well as their services. In this article, it has been mentioned that, as in the virtualization process, the network allows the different systems to merge easily; due to that reason, faults can be cultivated to other networks also through the virtualized environment, and with this attacker can easily cross those network slices to utilize the system for unauthorized activities. To handle those types of activities, different types of isolation activities were utilized, concluding traffic isolation, firewalls, encryption and so on, which could not provide efficient performance in handling the cyber-attacks. To resolve those issues properly, and achieve positive outcomes, in the current age, AI mechanisms are utilized to increase efficiency in performance. The main benefit of utilizing web isolation technology is that it can prevent malicious websites, malicious links, and fraudulent emails, and with this, handle fraud ads and fraud file download processes also.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

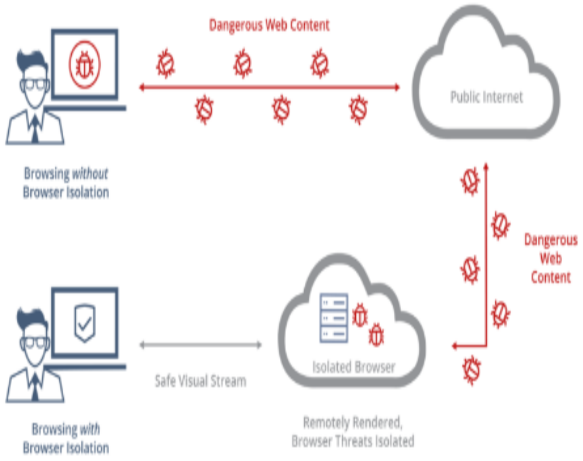


Figure 7: Browser isolation
Source: [22]

The above image provides an outline of browsing without browser isolation and browsing with browser isolation. By adopting the isolation approach, it can be easy to isolate browser threats and save browser content.

In the current technological era, the transportation system also becoming smart every day by utilizing different types of smart technologies. In one side, it makes the transportation system more efficient than the traditional processes, while on the other side, it increases various cyber threats also. According to [23], to confirm the higher quality security system of the intelligent transportation system, a proper isolation method should be utilized because the isolation method has the capability to detect false data through a "***Robust State Observer***". In a smart transportation system, passengers can easily communicate with another person through the network system. In that time, by hacking transportation networks, hackers could easily hack essential information. The isolation method has the capability to handle the information interchange process, and with this increases concerns about cyber-attacks within networked control systems, as well as cyber-physical systems. To isolate the multiple FDI smoothly, the "robust state observer bank" is developed.

It is evaluated based on the opinion of [24] that the networking system of smart vehicles should conclude an effective isolation strategy. The isolation strategy concludes with some steps. The initial step is to detect the system via the algorithm, whether the system is good or abnormal. After that, the outputs of the networking system are differentiated into two different subsets, and those are utilized to manage two robust state observers. This process should be repeated again. Based on the above steps, it can be easy to isolate multiple attacks and make sure that the intelligent transportation system can be secured properly.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

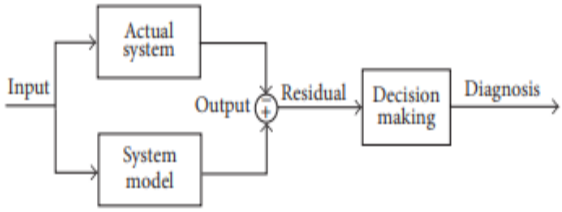


Figure 8: Conceptual structure of FDI (model-based)
Source: (Self-made)

"Model-based FDI) is normally a method which is mainly utilized to determine system fault by measuring the system with existing information. Through the above image, it has been defined that where the fault-identifying signal, as well as the decision-making, is to measure the residual for analyzing the probability of faults.

Measuring the impact of cyber-attack in current storage system

For business purposes, different business institutions utilize a major amount of data, and due to that reason, often, it can be difficult to handle vast amounts of data. In that situation, various organizations utilize the cloud system, which has the capability to measure big data. It is a widely utilized system, as it can prevent data loss. Among various types of advantages, the major disadvantage of the cloud system is privacy and security-based issues. Conducting DoS attacks ("Denial of Service") on cloud service can create lacking access to users' accounts. The main process of DoS attacks is to disrupt device resources by providing several requests, concluding with transferring malicious data to the server, which exhausts the application process, and with this, locks user accounts. In a cloud storage system, the other risk is the high probability of unauthorized access. According to [25], there are also various issues developed due to cyber-attacks, such as data leakage, account hacking, as well as a malicious insider. Malicious insider risks are also a major security concern within the cloud environment. Additionally, with different challenges, various solutions have also been mentioned here, concluding filtering the traffic regularly, enforcing multi-factor authentication, and implementing "*Data Loss Prevention*". For developing the data storage system, big organizations utilize cloud technology, but during implementing cloud, often due to the complexities of the cloud environment, the probability of risk is high, and that needs proper responsibility to mitigate those threats.

According to [26], cloud technology faces different risks, concluding human error, malware attacks, weak credentials, as well as criminal activities. The most fundamental feature which makes cloud services more easily accessible for the employee also makes it more difficult to handle unauthorized access.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Figure 9: Cyberattack in the cloud storage system
Source: [26]

The above image defines the cyber threats in the cloud storage system, concluding targeted attacks, lateral movement insider threats, data exfiltration, organized crime, as well as botnet monetization opportunistic threat. Cloud services are structured to support the information-sharing process easier, which provides support to hackers to target organizational systems easily. *Multi-factor authentication* has been utilized in a cloud storage system. Often weak passwords and weak management systems also increase cyber threats.

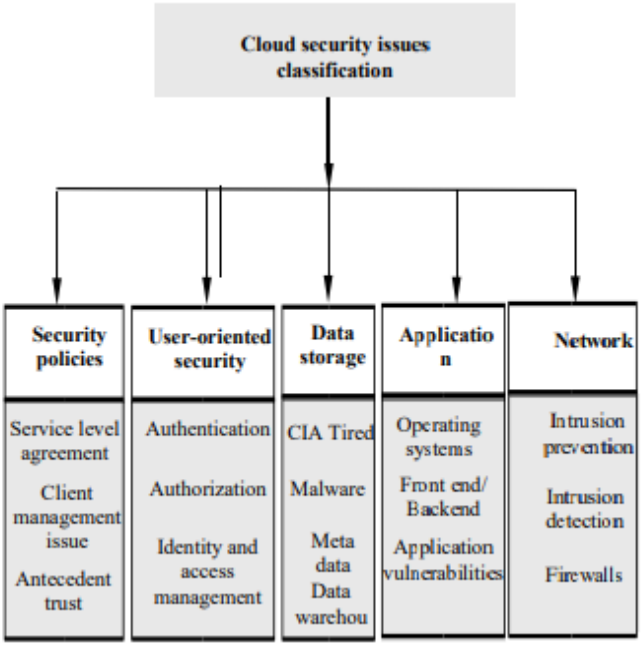


Figure: Various cloud security issues
Source: [27]

According to [27], it has been noted that cloud technologies are becoming a vital part of current corporate life, and producing innumerable opportunities to develop business processes more

Vol. 9, No. 01; 2024

ISSN: 2456-3676

efficiently. Cloud provides various advantages, but with this face different security challenges also. The above picture provides various security issues related to the cloud storage system, concluding security policies, user-oriented security, data storage, application, as well as network issues. To handle cyber-attack issues, a brief description has been provided based on ***Honeypot***, which is basically utilized in a cloud environment to capture information, sent from unauthorized resources. The main purpose of utilizing this solution is that it has the capability to prevent, identify, and with this manage cyber-attacks. By merging with firewalls, honeypots detect internal threats, as well as external threats.

Theoretical Application

Cybersecurity framework can help the company by providing a common language and a set of standards for security determinants across different countries and sectors to determine security postures and their vendors. A framework is also important to incorporate information through a framework that can help in attending the ***NIST*** cybersecurity framework that may help in improving to create greater collaboration between the public and private sector [28]. It helps in identifying, assessing and managing cyber risk. This framework is considered to be the gold standard for assessing the maturity of cybersecurity after the identification of security gaps and meeting the cybersecurity regulation [29]. As of February 2022, this framework is launched in the public interest to seek feedback and suggestions to improve the existing framework. In order to meet the requirements of the current trends, the company seeks information that can help in preventing cyber-attacks from different areas [30]. Another framework**, *ISO 27001 and ISO 27002***, which is created and issued by the International Organization of Standardization. If a company has such a framework, then that company can easily help in order to determine different aspects that can reduce the chance of cybersecurity. It can determine if the company can save the information related to the customers and other stakeholders. It may also help the company in restoring the reputation of the company.

Figure 9: Cybersecurity Framework
Source: [15]

Vol. 9, No. 01; 2024

ISSN: 2456-3676

The American Institute of Certified Public Accountants has developed the framework of Service Organisation Control Type 2. It is a security and auditing framework that can determine the tabard of audit and security of a company. It helps in verifying the vendors and partners securely managing the data related to the client. The comprehensiveness of the framework has turned it to be a complex framework to implement; this framework is specifically used in the accounting and financial sectors. This framework can help in resolving any issue that can bring conflict, and it can also help to mitigate any third-party risk. Healthcare organizations also face cyber-attacks. Thus, the Health Insurance Portability and Accountability Act is the framework to be used in the healthcare sector. The application of this act can protect the electronic medical data of the patients.

GDPR is one of the latest cybersecurity frameworks that can make the data procedure stronger. The framework has been designed for the data safety of the European Union and the citizens of the nation. This framework is adopted by the business organization of the country that uses, collects and stores personal information that can help in determining that all the data are safe and secured. It has 99 articles that explain the responsibility and compliances along with the data access rights, data protection policies and data breach [31]. The company needs to notify the national regulation about the data breach after the exploration of that. The federal Information Security management act is another framework that is used in protecting the information of the federal government. It can also help the company to determine if the company can help in getting the maximum outcome from that framework through the implication of different guidelines. It can act as a guideline to handle different issues associated with the different parameters.

Challenges

The research is going to explore different issues of cyber-attacks that can help the researcher to meet objectives. However, data breach is such an issue that can bring the reputation of that organization breach as stakeholders may not agree to work with this company. On the other hand, it is also important for the readers to be aware of the facts that have resolved the issue. In that case, the company may not like to share that information with external organizational context. This may help the company to create different rules that can reduce the chances of such an incident. Thus, finding out the cyber security instances in real-life examples can be hard to find. It can also help the company to create and generate ideas about different instances. On the other hand, a problem can also be encountered while searching for information.

III. Research Methodology

Overview of this methodology

Research methodology has been representing the various kinds of procedures of the methods here that can help to complete the reported work. Various kinds of techniques have been used to select, identify, analyze, and process to collect the data. Proper procedures and methods are very important to finish the collection of the data and the data analysis. The entire technology is used to composed of various sectors, including the method, research approaches, and research philosophy which are included with the various kind of designs which are used during the

Vol. 9, No. 01; 2024

ISSN: 2456-3676

collection, analyses, and research of various kinds of data during this research. This study compiles the reliability and validity of the research, which is comprehensive. It can involve different kinds of ethical considerations and addressing the research limitations.

Prediction Models

In this section, predicting the opinions is an essential activity to ensure the accurate compliance of cyber-attack detection and isolation. Different kinds o data mining techniques are used to develop an efficient predictive model [32]. There are three techniques used for detecting cyber-attacks and for isolation, such as Support Vector Machines or SVM, Random Forest or RF, and K-Nearest Neighbor or KNN for developing the predicting models. Every of these techniques offers innovative advantages, and this can adapt to many various kinds of contexts [33]. The limitations and strengths of every approach that can be used to make predicting analytics in the work and it can enhance the quality of the outcomes.

Support Vector Machine

SVM is a famous method that is used to classify the modelling that can be widely used for the data analytics for both the classifications and regression activity. SVM depends on the theory of statistical learning that is used to make it a relevant tool for measuring the data mathematically. It can be used as a linear hyperplane optically for splitting the data within the two different classes within their maximum margin in between the optical hyperplane and its nearest point [34]. The techniques of SVM have mapped the inputs in a vector with the high-dimensional characteristics distance into the transformations of the nonlinear that can make it efficient in solving complicated issues.

One of the major benefits of SVM is that it is optimal, innovative, and it is universal. For this reason, the solution of the SVM can be achieved by solving the constraint of the linearly quadratic problems. Another benefit is that this can be depends on the structural risk of the reduction and of their principle that can be used to reduce the top bound of their actual risks. This is a set of SVM from another classification where the SVM has been used for different areas and their applications and its theory that have been studied extensively [36]. The techniques of the SVM have been proven to be in different applications in cyber security, including the time series for the predictions. SVM techniques are performed in different areas comparable to the traditional classification.

Random Forests

Regression and classification activity can benefit from random forests or random decisions because of their assembled method of learning that can establish various decision trees. The accuracy of the random forests is lower compared to the gradient-booster despite trees where the fact of that outperform their former. Random forest, which has to be used for the DDOS that can attack detection. There are classified attacks by the architectures of the RF classifier. It is general architecture. This model has been developed using the assembling with their model of the logistic regression with the RF classifier that can be used to improve for both model's accuracy.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

K-Nearest Neighbor

The techniques of K-Nearest Neighbor which are used in the field of cybersecurity. It can work using analyzing the objects within the N characteristics in the space of N distance, where every object is considered as the one spot [38]. KNN has introduced the resemblance of the metric that can be used to for every classifier and object for the latest items using the comparison within the existing ones, and it can utilize the distance to measurement. This algorithm can be used to calculate every distance of the sampling set, and it can determine the nearest K neighbours for the unknown observation in cybersecurity. The cases of K are used to classify the latest observation using the specific classes. KNN is a remarkable classification of the algorithm to detect cyberattacks. The effectiveness of the techniques of the KNN algorithm in their comparison to their methods which are used for the discriminant analysis and logical analysis. The reporter determined that the KNN is more efficient and effective in terms of its average categorization accuracy.

Results

all relevant modules are imported for data wrangling, the modulus are matplotlib, panda, numpy, seaborn, sklearn, imblearn, and warning. For imporung the pyplot use the matplotlib.pyplot modules [41]. module pandas is import for data manipulation and analysis. To Provide the support to analyse the numeric and array import numpy modules. seaborn library is imported for visualizew the data.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Load Data

```
[ ]  # Dataset field names
     col_names = ["duration","protocol_type","service","flag",
                  "src_bytes","dst_bytes","land","wrong_fragment",
                  "urgent","hot","num_failed_logins","logged_in",
                  "num_compromised","root_shell","su_attempted",
                  "num_root","num_file_creations","num_shells",
                  "num_access_files","num_outbound_cmds","is_host_login",
                  "is_guest_login","count","srv_count","serror_rate",
                  "srv_serror_rate","rerror_rate","srv_rerror_rate",
                  "same_srv_rate","diff_srv_rate","srv_diff_host_rate",
                  "dst_host_count","dst_host_srv_count","dst_host_same_srv_rate",
                  "dst_host_diff_srv_rate","dst_host_same_src_port_rate",
                  "dst_host_srv_diff_host_rate","dst_host_serror_rate",
                  "dst_host_srv_serror_rate","dst_host_rerror_rate",
                  "dst_host_srv_rerror_rate","attack", "last_flag"]

     # Load NSL_KDD train dataset
     df_train = pd.read_table("KDDTrain+.txt",sep=",",names=col_names)
     df_train = df_train.iloc[:,:-1] # removes an unwanted extra field

     # Load NSL_KDD test dataset
     df_test = pd.read_table("KDDTest+.txt",sep=",",names=col_names)
     df_test = df_test.iloc[:,:-1]
```

Define the entire dataset in a variable as col_names to load the data. To load the dataset of NSL_KDD use the pd.raed_table() to read the data and df_train.iloc[: , -1] functions remove the unwanted extra space from the dataframe of df_train [42].

Train dataset dimensions

```
[ ]  df_train.head(3)
     print('Train set dimension: {} rows, {} columns'.format
      (df_train.shape[0], df_train.shape[1]))

     Train set dimension: 125973 rows, 42 columns
```

Test dataset Dimensions

```
[ ]  df_test.head(3)
     print('Test set dimension: {} rows, {} columns'.format
      (df_test.shape[0], df_test.shape[1]))

     Test set dimension: 22544 rows, 42 columns
```

Printing the train dataset dimensions using the shape attribute in the dataframe of df_train. Using the df_test it follow the logic of thre previous one.

Information About the both The Datasets like DataType and count of values

```
df_train.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 125973 entries, 0 to 125972
Data columns (total 42 columns):
 #   Column                       Non-Null Count   Dtype
---  ------                       --------------   -----
 0   duration                     125973 non-null  int64
 1   protocol_type                125973 non-null  object
 2   service                      125973 non-null  object
 3   flag                         125973 non-null  object
 4   src_bytes                    125973 non-null  int64
 5   dst_bytes                    125973 non-null  int64
 6   land                         125973 non-null  int64
 7   wrong_fragment               125973 non-null  int64
 8   urgent                       125973 non-null  int64
 9   hot                          125973 non-null  int64
 10  num_failed_logins            125973 non-null  int64
 11  logged_in                    125973 non-null  int64
 12  num_compromised              125973 non-null  int64
 13  root_shell                   125973 non-null  int64
 14  su_attempted                 125973 non-null  int64
 15  num_root                     125973 non-null  int64
 16  num_file_creations           125973 non-null  int64
 17  num_shells                   125973 non-null  int64
```

Vol. 9, No. 01; 2024

ISSN: 2456-3676

```
 17  num_shells                   125973 non-null  int64
 18  num_access_files             125973 non-null  int64
 19  num_outbound_cmds            125973 non-null  int64
 20  is_host_login                125973 non-null  int64
 21  is_guest_login               125973 non-null  int64
 22  count                        125973 non-null  int64
 23  srv_count                    125973 non-null  int64
 24  serror_rate                  125973 non-null  float64
 25  srv_serror_rate              125973 non-null  float64
 26  rerror_rate                  125973 non-null  float64
 27  srv_rerror_rate              125973 non-null  float64
 28  same_srv_rate                125973 non-null  float64
 29  diff_srv_rate                125973 non-null  float64
 30  srv_diff_host_rate           125973 non-null  float64
 31  dst_host_count               125973 non-null  int64
 32  dst_host_srv_count           125973 non-null  int64
 33  dst_host_same_srv_rate       125973 non-null  float64
 34  dst_host_diff_srv_rate       125973 non-null  float64
 35  dst_host_same_src_port_rate  125973 non-null  float64
 36  dst_host_srv_diff_host_rate  125973 non-null  float64
 37  dst_host_serror_rate         125973 non-null  float64
 38  dst_host_srv_serror_rate     125973 non-null  float64
 39  dst_host_rerror_rate         125973 non-null  float64
 40  dst_host_srv_rerror_rate     125973 non-null  float64
 41  attack                       125973 non-null  object
dtypes: float64(15), int64(23), object(4)
memory usage: 40.4+ MB
```

Based on the train dataset all null values are defined as 125573 for some column and mentioned data types are object, int, and flot. And based on the Test dataset all nul values are defined in 22544 for some column.

Checking for the Descriptive Statistics on Each and Every Column

```
df_train.describe()
```

	duration	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised	...	dst_host_co
count	125973.00000	1.259730e+05	1.259730e+05	125973.000000	125973.000000	125973.000000	125973.000000	125973.000000	125973.000000	125973.000000	...	125973.000
mean	287.14465	4.556674e+04	1.977911e+04	0.000198	0.022687	0.000111	0.204409	0.001222	0.395736	0.279250	...	182.148
std	2604.51531	5.870331e+06	4.021269e+06	0.014086	0.253530	0.014366	2.149968	0.045239	0.489010	23.942042	...	99.206
min	0.00000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000
25%	0.00000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	82.000
50%	0.00000	4.400000e+01	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	255.000
75%	0.00000	2.760000e+02	5.160000e+02	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	...	255.000
max	42908.00000	1.379964e+09	1.309937e+09	1.000000	3.000000	3.000000	77.000000	5.000000	1.000000	7479.000000	...	255.000

8 rows × 38 columns

The statistical deescriptive along with the several information like numbers of values of non null value, average value, standard deviation, min value, percentile values, and max value.

```
df_test.describe()
```

	duration	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised	...	dst_host_count
count	22544.000000	2.254400e+04	2.254400e+04	22544.000000	22544.000000	22544.000000	22544.000000	22544.000000	22544.000000	22544.000000	...	22544.000000
mean	218.859076	1.039545e+04	2.056019e+03	0.000311	0.008428	0.000710	0.105394	0.021647	0.442202	0.119899	...	193.869411
std	1407.176912	4.727864e+05	2.121930e+04	0.017619	0.142599	0.036473	0.928428	0.150328	0.496659	7.269597	...	94.035663
min	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	121.000000
50%	0.000000	5.400000e+01	4.600000e+01	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	255.000000
75%	0.000000	2.870000e+02	6.010000e+02	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	...	255.000000
max	57715.000000	6.282565e+07	1.345927e+06	1.000000	3.000000	3.000000	101.000000	4.000000	1.000000	796.000000	...	255.000000

8 rows × 38 columns

Descriptive statistics for some column for the test dataset including the several information like count, mean, min, std, percentile, and max.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Data Preprocessing

```
[ ]  mapping = {'ipsweep': 'Probe','satan': 'Probe','nmap': 'Probe',
                'portsweep': 'Probe','saint': 'Probe','mscan': 'Probe',
             'teardrop': 'DoS','pod': 'DoS','land': 'DoS','back': 'DoS',
                'neptune': 'DoS','smurf': 'DoS','mailbomb': 'DoS',
             'udpstorm': 'DoS','apache2': 'DoS','processtable': 'DoS',
             'perl': 'U2R','loadmodule': 'U2R','rootkit': 'U2R',
                'buffer_overflow': 'U2R','xterm': 'U2R','ps': 'U2R',
             'sqlattack': 'U2R','httptunnel': 'U2R',
             'ftp_write': 'R2L','phf': 'R2L','guess_passwd': 'R2L',
                'warezmaster': 'R2L','warezclient': 'R2L','imap': 'R2L',
             'spy': 'R2L','multihop': 'R2L','named': 'R2L',
                'snmpguess': 'R2L','worm': 'R2L','snmpgetattack': 'R2L',
             'xsnoop': 'R2L','xlock': 'R2L','sendmail': 'R2L',
             'normal': 'Normal'
             }

[ ]  # df_train['attack'].value_counts()

[ ]  # Apply attack class mappings to the dataset
     df_train['attack_class'] = df_train['attack'].apply(lambda v: mapping[v])
     df_test['attack_class'] = df_test['attack'].apply(lambda v: mapping[v])
```

use a mapping dictionary to map each type of attack with their corresponding class. From the dataset of training used to get all counts of every type of attack [35]. Mapped attack class from the attack column creates the new column to assign its corresponding class depending on this mapping dictionary.

```
# df_train.info()

# Drop attack field from both train and test data
df_train.drop(['attack'], axis=1, inplace=True)
df_test.drop(['attack'], axis=1, inplace=True)

# View top 3 train data
df_train.head(3)
```

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv_rate	dst_host_diff_srv_rate	ds
0	0	tcp	ftp_data	SF	491	0	0	0	0	0	...	25	0.17	0.03	
1	0	udp	other	SF	146	0	0	0	0	0	...	1	0.00	0.60	
2	0	tcp	private	S0	0	0	0	0	0	0	...	26	0.10	0.05	

3 rows × 42 columns

```
df_train.isnull().sum()
```

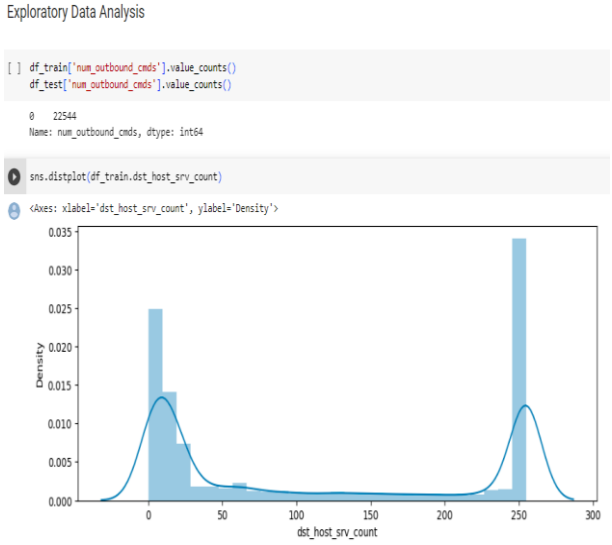
```
duration                     0
protocol_type                0
service                      0
flag                         0
src_bytes                    0
dst_bytes                    0
land                         0
wrong_fragment               0
urgent                       0
hot                          0
num_failed_logins            0
logged_in                    0
num_compromised              0
root_shell                   0
su_attempted                 0
num_root                     0
num_file_creations           0
num_shells                   0
num_access_files             0
num_outbound_cmds            0
is_host_login                0
is_guest_login               0
count                        0
```

Vol. 9, No. 01; 2024

ISSN: 2456-3676

```
srv_count                    0
serror_rate                  0
srv_serror_rate              0
rerror_rate                  0
srv_rerror_rate              0
same_srv_rate                0
diff_srv_rate                0
srv_diff_host_rate           0
dst_host_count               0
dst_host_srv_count           0
dst_host_same_srv_rate       0
dst_host_diff_srv_rate       0
dst_host_same_src_port_rate  0
dst_host_srv_diff_host_rate  0
dst_host_serror_rate         0
dst_host_srv_serror_rate     0
dst_host_rerror_rate         0
dst_host_srv_rerror_rate     0
attack_class                 0
dtype: int64
```

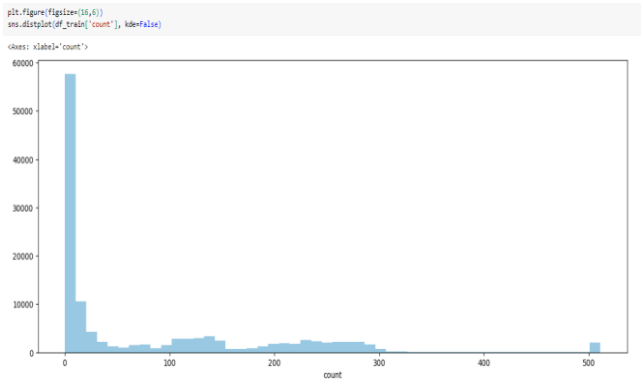
Drops the field of attack for the dataset of train use the drop() and set using in place parameter as true. Use head() to display all the column. Useisnull ().sum() to count all the null values [43].



Using the EDA functionalities to analyse the dataset and draw the graph and for that use to create the distribution plot.

Vol. 9, No. 01; 2024

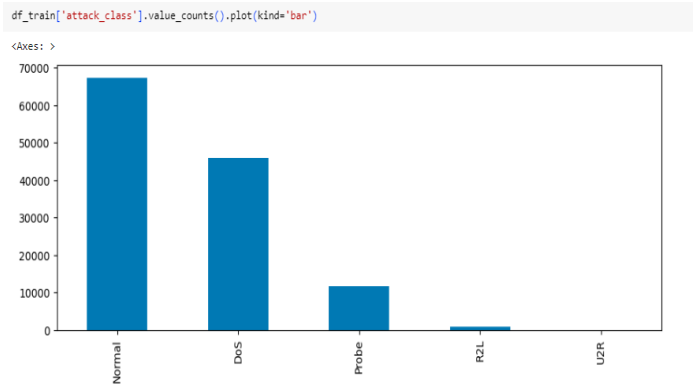
ISSN: 2456-3676



Draw a graph regarding the distplot function inckusin the counts of the dataset of df_train and set value as false [39].

```
print('Label distribution Training set:')
print(df_train['attack_class'].value_counts())
print()
print('Label distribution Test set:')
print(df_test['attack_class'].value_counts())
```

```
Label distribution Training set:
Normal    67343
DoS       45927
Probe     11656
R2L         995
U2R          52
Name: attack_class, dtype: int64

Label distribution Test set:
Normal    9711
DoS       7458
R2L       2754
Probe     2421
U2R        200
Name: attack_class, dtype: int64
```



Using the value.count() to count all the values and using the plot() to draw a graph [44].

Vol. 9, No. 01; 2024

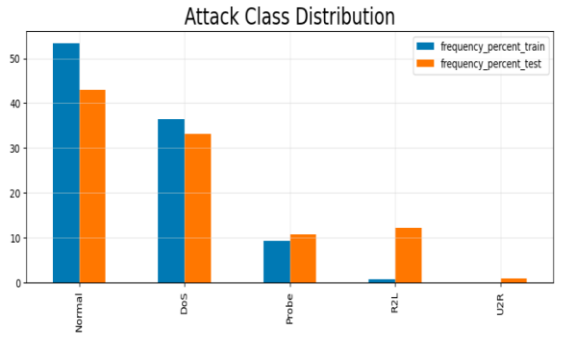
ISSN: 2456-3676

```
# 'num_outbound_cmds' field has all 0 values. Hence, it will be removed from both train and test dataset since it is a redundant field.
df_train.drop(['num_outbound_cmds'], axis=1, inplace=True)
df_test.drop(['num_outbound_cmds'], axis=1, inplace=True)
```

```
# Attack Class Distribution
train_attack_freq = df_train[['attack_class']].apply(lambda x: x.value_counts())
test_attack_freq = df_test[['attack_class']].apply(lambda x: x.value_counts())
train_attack_freq['frequency_percent_train'] = round((100 * train_attack_freq / train_attack_freq.sum()),2)
test_attack_freq['frequency_percent_test'] = round((100 * test_attack_freq / test_attack_freq.sum()),2)

dist_attack = pd.concat([train_attack_freq,test_attack_freq], axis=1)
dist_attack
```

	attack_class	frequency_percent_train	attack_class	frequency_percent_test
Normal	67343	53.46	9711	43.08
DoS	45927	36.46	7458	33.08
Probe	11656	9.25	2421	10.74
R2L	995	0.79	2754	12.22
U2R	52	0.04	200	0.89

```
# Attack class bar plot
plot = dist_attack[['frequency_percent_train', 'frequency_percent_test']].plot(kind="bar");
plot.set_title("Attack Class Distribution", fontsize=20);
plot.grid(color='lightgray', alpha=0.5);
```



The results are displayed on a table with the class of attack, training dataset frequency, frequency of the dataset [37].

```
df_train.head()
```

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv_rate	dst_host_diff_srv_rate	dst_host_same_src_port_rate	dst_host_srv_di
0	0	tcp	ftp_data	SF	491	0	0	0	0	0	...	25	0.17	0.03	0.17	
1	0	udp	other	SF	146	0	0	0	0	0	...	1	0.00	0.60	0.88	
2	0	tcp	private	S0	0	0	0	0	0	0	...	26	0.10	0.05	0.00	
3	0	tcp	http	SF	232	8153	0	0	0	0	...	255	1.00	0.00	0.03	
4	0	tcp	http	SF	199	420	0	0	0	0	...	255	1.00	0.00	0.00	

5 rows × 41 columns

Scaling Numerical Attributes

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()

# extract numerical attributes and scale it to have zero mean and unit variance
cols = df_train.select_dtypes(include=['float64','int64']).columns
train_sc = scaler.fit_transform(df_train.select_dtypes(include=['float64','int64']))
test_sc = scaler.fit_transform(df_test.select_dtypes(include=['float64','int64']))

# turn the result back to a dataframe
train_scdf = pd.DataFrame(train_sc, columns = cols)
test_scdf = pd.DataFrame(test_sc, columns = cols)
```

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Encoding of Categorical Attributes

```
[ ] from sklearn.preprocessing import LabelEncoder
    encoder = LabelEncoder()

    # extract categorical attributes from both training and test sets
    train_cat = df_train.select_dtypes(include=['object']).copy()
    test_cat = df_test.select_dtypes(include=['object']).copy()

    # encode the categorical attributes
    traincat = train_cat.apply(encoder.fit_transform)
    testcat = test_cat.apply(encoder.fit_transform)

    # separate target column from encoded data
    train_enc = traincat.drop(['attack_class'], axis=1)
    test_enc = testcat.drop(['attack_class'], axis=1)

    cat_Ytrain = traincat[['attack_class']].copy()
    cat_Ytest = testcat[['attack_class']].copy()
```

For Encoding the categorical attributes import the LabelEncoder module from sklearn. Extract the data from training sets and test sets [48]. Separate the target values from the encoded data.

Data Sampling

```
[ ] from imblearn.over_sampling import RandomOverSampler
    from collections import Counter

    # define columns and extract encoded train set for sampling
    train_scdf = df_train.select_dtypes(include=['float64','int64'])
    refclasscol = pd.concat([train_scdf, train_enc], axis=1).columns
    refclass = np.concatenate((train_sc, train_enc.values), axis=1)
    X = refclass

    # reshape target column to 1D array shape
    c, r = cat_Ytest.values.shape
    y_test = cat_Ytest.values.reshape(c,)

    c, r = cat_Ytrain.values.shape
    y = cat_Ytrain.values.reshape(c,)

    # apply the random over-sampling
    ros = RandomOverSampler(random_state=42)
    X_res, y_res = ros.fit_resample(X, y)   # got an error here sample --->fit_resample
    print('Original dataset shape {}'.format(Counter(y)))
    print('Resampled dataset shape {}'.format(Counter(y_res)))

    Original dataset shape Counter({1: 67343, 0: 45927, 2: 11656, 3: 995, 4: 52})
    Resampled dataset shape Counter({1: 67343, 0: 67343, 3: 67343, 2: 67343, 4: 67343})
```

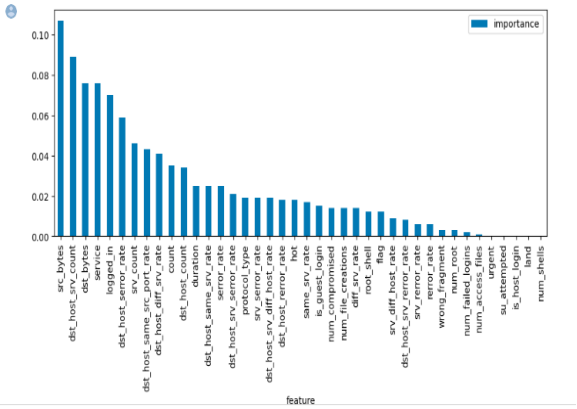
Import RandomOverSampler from the imblearn to define the columns and extract the set of encoded train for data sampling [47]. The target value is transferred to the one-dimensional array.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Feature Selection

```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier();
# fit random forest classifier on the training set
rfc.fit(X_res, y_res);
# extract important features
score = np.round(rfc.feature_importances_,3)
Imp_feat = pd.DataFrame({'feature':refclasscol,'importance':score})
Imp_feat = Imp_feat.sort_values('importance',ascending=False).set_index('feature')
# plot Imp_feat
plt.rcParams['figure.figsize'] = (11, 4)
Imp_feat.plot.bar();
```



Use the random forest algorithm with two axes where x axes define the input features and y axes define the target variable. extract the importance of the features by feature_importances attributes [45]. Create the dataframe of pandas to store the names, features and its score. Plot the grzaph using the Imp_feat.plot() [46].

```
from sklearn.feature_selection import RFE
import itertools
rfc = RandomForestClassifier()

# create the RFE model and select 10 attributes
rfe = RFE(rfc, n_features_to_select=10)
rfe = rfe.fit(X_res, y_res)

# summarize the selection of the attributes
feat_map = [(i, v) for i, v in itertools.zip_longest(rfe.get_support(), refclasscol)]
feat_selected = [v for i, v in feat_map if i==True]
```

```
feat_selected
```

```
['src_bytes',
 'dst_bytes',
 'logged_in',
 'count',
 'srv_count',
 'dst_host_srv_count',
 'dst_host_diff_srv_rate',
 'dst_host_same_src_port_rate',
 'dst_host_serror_rate',
 'service']
```

To create the model of RFE use to import the RFE from sklearn and summarize rhe attributes use map () to map the values.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Dataset Partition

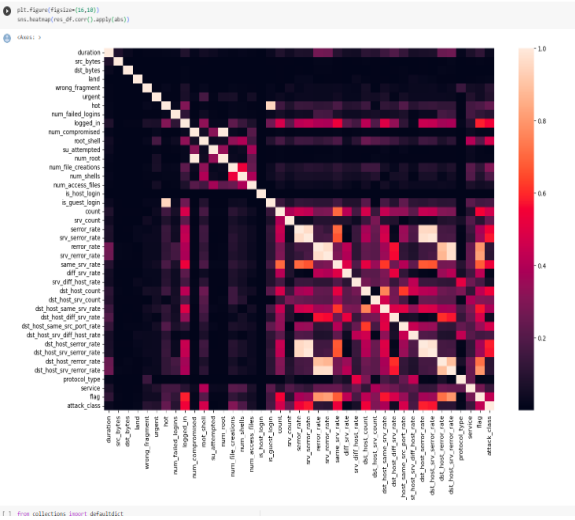
```
# define columns to new dataframe
newcol = list(refclasscol)
newcol.append('attack_class')

# add a dimension to target
new_y_res = y_res[:, np.newaxis]

# create a dataframe from sampled data
res_arr = np.concatenate((X_res, new_y_res), axis=1)
res_df = pd.DataFrame(res_arr, columns = newcol)

# create test dataframe
test_df_ref = pd.concat([test_scdf, testcat], axis=1)
test_df_ref['attack_class'] = test_df_ref['attack_class'].astype(np.float64)
test_df_ref['protocol_type'] = test_df_ref['protocol_type'].astype(np.float64)
test_df_ref['flag'] = test_df_ref['flag'].astype(np.float64)
test_df_ref['service'] = test_df_ref['service'].astype(np.float64)

res_df.shape
test_df_ref.shape
```

```
(22544, 41)
```

Correlation for Transformed Data

```
plt.figure(figsize=(16,10))
sns.heatmap(res_df.corr().apply(abs))
```

Correlation for Transformed Data



In the data partition define the new dataframe column with list() to list out all reference classes and append them using append(). Create the test data frame using concat() to concating the features of test_scdf and categorical features testcat. Visulalize the dataset from the above figure along with their features.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

```
[ ] from collections import defaultdict
    classdict = defaultdict(list)

    # create two-target classes (normal class and an attack class)
    attacklist = [('DoS', 0.0), ('Probe', 2.0), ('R2L', 3.0), ('U2R', 4.0)]
    normalclass = [('Normal', 1.0)]

    def create_classdict():
        '''This function subdivides train and test dataset into two-class attack labels'''
        for j, k in normalclass:
            for i, v in attacklist:
                restrain_set = res_df.loc[(res_df['attack_class'] == k) | (res_df['attack_class'] == v)]
                classdict[j +'_' + i].append(restrain_set)
                # test labels
                test_df_ref_set = test_df_ref.loc[(test_df_ref['attack_class'] == k) | (test_df_ref['attack_class'] == v)]
                classdict[j +'_' + i].append(test_df_ref_set)

    create_classdict()
```

```
for k, v in classdict.items():
    k
```

```
[ ] pretrain = classdict['Normal_DoS'][0]
    pretest = classdict['Normal_DoS'][1]
    grpclass = 'Normal_DoS'
```

Finalize data preprocessing for training

```
from sklearn.preprocessing import OneHotEncoder
enc = OneHotEncoder(handle_unknown='ignore')    #OneHotEncoder()

Xresdf = pretrain
newtest = pretest

Xresdfnew = Xresdf[feat_selected]
Xresdfnum = Xresdfnew.drop(['service'], axis=1)
Xresdfcat = Xresdfnew[['service']].copy()

Xtest_features = newtest[feat_selected]
Xtestdfnum = Xtest_features.drop(['service'], axis=1)
Xtestcat = Xtest_features[['service']].copy()


# Fit train data
enc.fit(Xresdfcat)

# Transform train data
X_train_1hotenc = enc.transform(Xresdfcat).toarray()
```

```
# Transform test data
X_test_1hotenc = enc.transform(Xtestcat).toarray()

X_train = np.concatenate((Xresdfnum.values, X_train_1hotenc), axis=1)
X_test = np.concatenate((Xtestdfnum.values, X_test_1hotenc), axis=1)

y_train = Xresdf[['attack_class']].copy()
c, r = y_train.values.shape
Y_train = y_train.values.reshape(c,)

y_test = newtest[['attack_class']].copy()
c, r = y_test.values.shape
Y_test = y_test.values.reshape(c,)
```

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Train Models

```
from sklearn.naive_bayes import BernoulliNB
from sklearn import tree
from sklearn.model_selection import cross_val_score
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import VotingClassifier

# Train KNeighborsClassifier Model
KNN_Classifier = KNeighborsClassifier(n_jobs=-1)
KNN_Classifier.fit(X_train, Y_train);

# Train LogisticRegression Model
LGR_Classifier = LogisticRegression(n_jobs=-1, random_state=0)
LGR_Classifier.fit(X_train, Y_train);

# Train Gaussian Naive Baye Model
BNB_Classifier = BernoulliNB()
BNB_Classifier.fit(X_train, Y_train)

# Train Decision Tree Model
DTC_Classifier = tree.DecisionTreeClassifier(criterion='entropy', random_state=0)
DTC_Classifier.fit(X_train, Y_train);

# Train RandomForestClassifier Model
RF_Classifier = RandomForestClassifier(criterion='entropy', n_jobs=-1, random_state=0)
RF_Classifier.fit(X_train, Y_train);
```

Evaluate Models

Test Models

```
[ ] from sklearn import metrics

    models = []
    models.append(('Naive Baye Classifier', BNB_Classifier))
    models.append(('Decision Tree Classifier', DTC_Classifier))
    models.append(('RandomForest Classifier', RF_Classifier))
    models.append(('KNeighborsClassifier', KNN_Classifier))
    models.append(('LogisticRegression', LGR_Classifier))
```

```
for i, v in models:
    accuracy = metrics.accuracy_score(Y_test, v.predict(X_test))
    confusion_matrix = metrics.confusion_matrix(Y_test, v.predict(X_test))
    classification = metrics.classification_report(Y_test, v.predict(X_test))
    print()
    print('============================== {} {} Model Test Results ==============================='.format(grpclass, i))
    print()
    print ("Model Accuracy:" "\n", accuracy)
    print()
    print("Confusion matrix:" "\n", confusion_matrix)
    print()
    print("Classification report:" "\n", classification)
    print()
```

Heatmap visualization used to visualize the matrix of the data frame of res_df to identify various features. To create a dictionary, the defaultdict class is used to store the training datasets and test datasets. Divide the data frame in the attack classe depends on the attack list along with the normal class. To perform the train dataset includes the categorical features use the OneHotEncoder for relevant features and build separate sets of features.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

```
============================== Normal_DoS Naive Baye Classifier Model Test Results ==============================

Model Accuracy:
 0.8336536781408352

Confusion matrix:
 [[5487 1971]
 [ 885 8826]]

Classification report:
              precision    recall  f1-score   support

         0.0       0.86      0.74      0.79      7458
         1.0       0.82      0.91      0.86      9711

    accuracy                           0.83     17169
   macro avg       0.84      0.82      0.83     17169
weighted avg       0.84      0.83      0.83     17169
```

The accuracy of the native bayes is 83%

```
============================== Normal_DoS Decision Tree Classifier Model Test Results ==============================

Model Accuracy:
 0.8165880365775525

Confusion matrix:
 [[5591 1867]
 [1282 8429]]

Classification report:
              precision    recall  f1-score   support

         0.0       0.81      0.75      0.78      7458
         1.0       0.82      0.87      0.84      9711

    accuracy                           0.82     17169
   macro avg       0.82      0.81      0.81     17169
weighted avg       0.82      0.82      0.82     17169
```

The accuracy of the decision tree classifier is 82%

```
============================== Normal_DoS RandomForest Classifier Model Test Results ==============================

Model Accuracy:
 0.829518317898538

Confusion matrix:
 [[5489 1969]
 [ 958 8753]]

Classification report:
              precision    recall  f1-score   support

         0.0       0.85      0.74      0.79      7458
         1.0       0.82      0.90      0.86      9711

    accuracy                           0.83     17169
   macro avg       0.83      0.82      0.82     17169
weighted avg       0.83      0.83      0.83     17169
```

The accuracy of the Random Forest Classifier is 83%

Vol. 9, No. 01; 2024

ISSN: 2456-3676

```
============================== Normal_DoS KNeighborsClassifier Model Test Results ==============================

Model Accuracy:
 0.8666200710583027

Confusion matrix:
 [[5787 1671]
 [ 619 9092]]

Classification report:
              precision    recall  f1-score   support

         0.0       0.90      0.78      0.83      7458
         1.0       0.84      0.94      0.89      9711

    accuracy                           0.87     17169
   macro avg       0.87      0.86      0.86     17169
weighted avg       0.87      0.87      0.86     17169
```

The accuracy of KNeighbors Classifier model is 87%

```
============================== Normal_DoS LogisticRegression Model Test Results ==============================

Model Accuracy:
 0.8418661541149747

Confusion matrix:
 [[5963 1495]
 [1220 8491]]

Classification report:
              precision    recall  f1-score   support

         0.0       0.83      0.80      0.81      7458
         1.0       0.85      0.87      0.86      9711

    accuracy                           0.84     17169
   macro avg       0.84      0.84      0.84     17169
weighted avg       0.84      0.84      0.84     17169
```

The accuracy of Logistic Regression model is 84%

From the comparison, the accuracy is high for the KNeighbors Classifier model which is 87% more efficient and effective.

V. Conclusion

The paper has depicted the participation of cyber-attacks and the ways that can reduce the chance of cyber-attacks. In order to do that, the paper has explored different cases of cyber threats and explored the ways to detect the attack. In order to do that, it is also important that a company uses a proper framework. The maximum cyber-attacks are taking place due to the inefficient data protection in the financial sector. Thus, it may help in creating the best outcome through organizational landscape. Security attacks are becoming rapid, and every time, hackers come coming with new ideas that may create issues in following the organizational structure, data management and storage management. In that case, it may create problems in leading the organization with a secured platform. Apart from that, it may help the company to create different plans, which may help the company to get over such issues. These factors are motivating to take this research so that it can help in determining the maximum benefit to the individuals and overall organizations.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

The paper is focusing on exploring different cyber-attacks that are happening in organizational and personal systems. The current situation of the pandemic has developed the trend of remote working, and different reports have suggested that remote working will persist for a longer time. Thus, it is creating a new cyber security trend as the home environment is not as prorated as an office. However, a maximum number of organizations are sending official systems to maintain the security, though there is a slight chance that the company may face threats of cyber security. The use of Artificial intelligence and machine learning is quite dominant according to the current trends. In that case, it is necessary for the company to make sure that they are using all these devices actively by maintaining the needs. AI tools are widely used by hackers for advanced attacks.

VI. Acknowledgment

I would like to thank God for all the blessings and opportunities bestowed on me. I would like to thank my family for their love, prayers, encouragement and unconditional support throughout my research process. Special thank you to my sisters, and other family members without whom I wouldn't have made it so far and enjoyed life half as much! I would also like to express my gratitude to my supervisor for guidance, support, clarity and constructive feedback throughout the dissertation project. Lastly, thank you to the survey participants without whom this project would not be possible.

References

Xiao, J., &Feroskhan, M., 2022. Cyber-attack detection and isolation for a quadrotor uav with modified sliding innovation sequences. *IEEE Transactions on Vehicular Technology*, *71*(7), 7202-7214.

Biroon, R. A., Biron, Z. A., &Pisu, P., 2021. False data injection attack in a platoon of CACC: real-time detection and isolation with a PDE approach. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 8692-8703.

Shi, D., Lin, P., Wang, Y., Chu, C. C., Xu, Y., & Wang, P., 2021. Deception attack detection of isolated DC microgrids under consensus-based distributed voltage control architecture. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, *11*(1), 155-167.

Taheri, M., Khorasani, K., Shames, I., &Meskin, N., 2020. Cyber-attack and machine induced fault detection and isolation methodologies for cyber-physical systems. *arXiv preprint arXiv:2009.06196*.

Ye, L., Zhu, F., & Zhang, J., 2020. Sensor attack detection and isolation based on sliding mode observer for cyber-physical systems. *International Journal of Adaptive Control and Signal Processing*, *34*(4), 469-483.

Zhou, Q., Shahidehpour, M., Alabdulwahab, A., &Abusorrah, A., 2020. A cyber-attack resilient distributed control strategy in islanded microgrids. *IEEE Transactions on Smart Grid*, *11*(5), 3690-3701.

Yan, J., Guo, F., & Wen, C., 2020. Attack detection and isolation for distributed load shedding algorithm in microgrid systems. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, *1*(1), 102-110.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Khan, A. S., Khan, A. Q., Iqbal, N., Sarwar, M., Mahmood, A., & Shoaib, M. A., 2020. Distributed fault detection and isolation in second order networked systems in a cyber–physical environment. *ISA transactions*, *103*, 131-142.

Elnour, M., Meskin, N., Khan, K., & Jain, R., 2020. A dual-isolation-forests-based attack detection framework for industrial control systems. *IEEE Access*, *8*, 36639-36651.

Aluko, A. O., Carpanen, R. P., Dorrell, D. G., &Ojo, E. E., (2022). Real-Time Cyber-attack Detection Scheme for Standalone Microgrids. *IEEE Internet of Things Journal*, *9*(21), 21481-21492.

Siddiqui, M. A., Stokes, J. W., Seifert, C., Argyle, E., McCann, R., Neil, J., & Carroll, J., 2019, May. Detecting cyber attacks using anomaly detection with explanations and expert feedback. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2872-2876). IEEE.

Al-Dabbagh, A. W., Barboni, A., &Parisini, T., 2020. Distributed Detection and Isolation of Covert Cyber Attacks for a Class of Interconnected Systems. *IFAC-PapersOnLine*, *53*(2), 772-777.

Elnour, M., Meskin, N., & Khan, K. M., 2020, August. Hybrid attack detection framework for industrial control systems using 1D-convolutional neural network and isolation forest. In *2020 IEEE Conference on Control Technology and Applications (CCTA)* (pp. 877-884). IEEE.

Tiwari, D. D., Naskar, S., Sai, A. S., &Palleti, V. R., 2021. Attack detection using unsupervised learning algorithms in cyber-physical systems. In *Computer Aided Chemical Engineering* (Vol. 50, pp. 1259-1264). Elsevier.

Wang, X., Luo, X., & Guan, X., 2017, July. Unknown cyber-attack detection and isolation for power systems via Luenberger observer. In *2017 4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS)* (pp. 673-678). IEEE.

Georgiadou, A., Mouzakitis, S., &Askounis, D. (2022). Working from home during COVID-19 crisis: a cyber security culture assessment survey. *Security Journal*, *35*(2), 486-505.

Vilić, V. (2022). COVID-19 and Cyber Threats: Aggression, Frauds, and Infodemic in Cyberspace during the Pandemic.

Georgiadou, A., Mouzakitis, S., Bounas, K., &Askounis, D. (2022). A cyber-security culture framework for assessing organization readiness. *Journal of Computer Information Systems*, *62*(3), 452-462.

Ansari, M. F., Sharma, P. K., & Dash, B. (2022). Prevention of phishing attacks using AI-based Cybersecurity Awareness Training. *Prevention*.

Ghelani, D. (2022). Cyber security, cyber threats, implications and future perspectives: A Review. *Authorea Preprints*.

Uyyala, P. (2022). DETECTION OF CYBER-ATTACK IN NETWORK USING MACHINE LEARNING TECHNIQUES. *Journal of interdisciplinary cycle research*, *14*(3), 1903-1913.

Wong, S., Han, B., &Schotten, H. D. (2022). 5G Network Slice Isolation. *Network*, *2*(1), 153-167.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Huang, X. and Wang, X., 2022. Detection and Isolation of False Data Injection Attack in Intelligent Transportation System via Robust State Observer. *Processes*, *10*(7), p.1299.

AbuAlghanam, O., Alazzam, H., Alhenawi, E. A., Qatawneh, M., &Adwan, O. (2023). Fusion-based anomaly detection system using modified isolation forest for internet of things. *Journal of Ambient Intelligence and Humanized Computing*, *14*(1), 131-145.

Al Nafea, R., &Almaiah, M. A. (2021, July). Cyber security threats in cloud: Literature review. In *2021 International Conference on Information Technology (ICIT)* (pp. 779-786). IEEE.

Aldasoro, I., Gambacorta, L., Giudici, P., & Leach, T. (2022). The drivers of cyber risk. *Journal of Financial Stability*, *60*, 100989.

Omer, M. A., Yazdeen, A. A., Malallah, H. S., & Abdulrahman, L. M. (2022). A Survey on Cloud Security: Concepts, Types, Limitations, and Challenges. *Journal of Applied Science and Technology Trends*, *3*(02), 47-57.

Markopoulou, D., Papakonstantinou, V. and De Hert, P., 2019. The new EU cybersecurity framework: The NIS Directive, ENISA's role and the General Data Protection Regulation. *Computer Law & Security Review*, *35*(6), p.105336.

Srinivas, J., Das, A.K. and Kumar, N., 2019. Government regulations in cyber security: Framework, standards and recommendations. *Future generation computer systems*, *92*, pp.178-188.Srinivas, J., Das, A.K. and Kumar, N., 2019. Government regulations in cyber security: Framework, standards and recommendations. *Future generation computer systems*, *92*, pp.178-188.

Efthymiopoulos, M.P., 2019. A cyber-security framework for development, defense and innovation at NATO. *Journal of Innovation and Entrepreneurship*, *8*(1), pp.1-26.

Sani, A.S., Yuan, D., Jin, J., Gao, L., Yu, S. and Dong, Z.Y., 2019. Cyber security framework for Internet of Things-based Energy Internet. *Future Generation Computer Systems*, *93*, pp.849-859.

Islam, U., Muhammad, A., Mansoor, R., Hossain, M. S., Ahmad, I., Eldin, E. T., ... & Shafiq, M. (2022). Detection of distributed denial of service (DDoS) attacks in IOT based monitoring system of banking sector using machine learning models. *Sustainability*, *14*(14), 8374.

Musumeci, F., Ionata, V., Paolucci, F., Cugini, F., &Tornatore, M. (2020, June). Machine-learning-assisted DDoS attack detection with P4 language. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)* (pp. 1-6). IEEE.

Miao, Y., Chen, C., Pan, L., Han, Q. L., Zhang, J., & Xiang, Y. (2021). Machine learning–based cyber attacks targeting on controlled information: A survey. *ACM Computing Surveys (CSUR)*, *54*(7), 1-36.

Das, S., Venugopal, D., & Shiva, S. (2020). A holistic approach for detecting ddos attacks by using ensemble unsupervised machine learning. In *Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), Volume 2* (pp. 721-738). Springer International Publishing.

Vol. 9, No. 01; 2024

ISSN: 2456-3676

Awan, M. J., Farooq, U., Babar, H. M. A., Yasin, A., Nobanee, H., Hussain, M., ... & Zain, A. M. (2021). Real-time DDoS attack detection system using big data approach. *Sustainability*, *13*(19), 10743.

Hindy, H., Atkinson, R., Tachtatzis, C., Colin, J. N., Bayne, E., &Bellekens, X. (2020). Utilising deep learning techniques for effective zero-day attack detection. *Electronics*, *9*(10), 1684.

Yungaicela-Naula, N. M., Vargas-Rosales, C., & Perez-Diaz, J. A. (2021). SDN-based architecture for transport and application layer DDoS attack detection by using machine and deep learning. *IEEE Access*, *9*, 108495-108512.

Salloum, S., Gaber, T., Vadera, S., &Shaalan, K. (2021). Phishing email detection using natural language processing techniques: a literature survey. *Procedia Computer Science*, *189*, 19-28.

Anwer, M., Khan, S. M., & Farooq, M. U. (2021). Attack detection in IoT using machine learning. *Engineering, Technology & Applied Science Research*, *11*(3), 7273-7278.

Saharkhizan, M., Azmoodeh, A., Dehghantanha, A., Choo, K. K. R., &Parizi, R. M. (2020). An ensemble of deep recurrent neural networks for detecting IoT cyber attacks using network traffic. *IEEE Internet of Things Journal*, *7*(9), 8852-8859.

Dasgupta, D., Akhtar, Z., & Sen, S. (2022). Machine learning in cybersecurity: a comprehensive survey. *The Journal of DefenseModeling and Simulation*, *19*(1), 57-106.

Mondal, B., Koner, C., Chakraborty, M., & Gupta, S. (2022). Detection and investigation of DDoS attacks in network traffic using machine learning algorithms. *Int. J. Innov. Technol. Explor. Eng.*, *11*(6), 1-6.

Truong, T. C., Zelinka, I., Plucar, J., Čandík, M., &Šulc, V. (2020). Artificial intelligence and cybersecurity: Past, presence, and future. In *Artificial intelligence and evolutionary computations in engineering systems* (pp. 351-363). Springer Singapore.

Acosta, M. R. C., Ahmed, S., Garcia, C. E., & Koo, I. (2020). Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks. *IEEE access*, *8*, 19921-19933.

Yin, X. C., Liu, Z. G., Nkenyereye, L., &Ndibanje, B. (2019). Toward an applied cyber security solution in IoT-based smart grids: An intrusion detection system approach. *Sensors*, *19*(22), 4952.

Ravikumar, G., &Govindarasu, M. (2020). Anomaly detection and mitigation for wide-area damping control using machine learning. *IEEE Transactions on Smart Grid*.

Strecker, S., Van Haaften, W., & Dave, R. (2021). An analysis of IoT cyber security driven by machine learning. In *Proceedings of International Conference on Communication and Computational Technologies: ICCCT 2021* (pp. 725-753). Springer Singapore.