IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Machine Learning for Cloud Security: A Systematic Review

Ali Bou Nassif[1]*, Manar Abu Talib[2], Qassim Nassir[3], Halah Albadani[2], Fatima Dak Albab[2]
[1]Department of Computer Engineering, University of Sharjah, Sharjah, United Arab Emirates
[2]Department of Computer Science, University of Sharjah, Sharjah, United Arab Emirates
[3]Department of Electrical Engineering, University of Sharjah, Sharjah, United Arab Emirates
Email: {anassif, mtalib, nasir, U16104005, falbab}@sharjah.ac.ae
*Corresponding Author: Ali Bou Nassif
*Corresponding Author Email: anassif@sharjah.ac.ae

ABSTRACT **The popularity and usage of Cloud computing is increasing rapidly. Several companies are investing in this field either for their own use or to provide it as a service for others. One of the results of Cloud development is the emergence of various security problems for both industry and consumer. One of the ways to secure Cloud is by using Machine Learning (ML). ML techniques have been used in various ways to prevent or detect attacks and security gaps on the Cloud. In this paper, we provide a Systematic Literature Review (SLR) of ML and Cloud security methodologies and techniques. We analyzed 63 relevant studies and the results of the SLR are categorized into three main research areas: (i) the different types of Cloud security threats, (ii) ML techniques used, and (iii) the performance outcomes. We have defined 11 Cloud security areas. Moreover, distributed denial-of-service (DDoS) and data privacy are the most common Cloud security areas, with a 16% level of use and 14%respectively. On the other hand, we found 30 ML techniques used, some used hybrid and others as standalone. The most popular ML used is SVM in both hybrid and standalone models. Furthermore, 60% of the papers compared their models with other models to prove the efficiency of their proposed model. Moreover, 13 different evaluation metrics were enumerated. The most applied metric is true positive rate and least used is training time. Lastly, from 20 datasets found, KDD and KDD CUP'99 are the most used among relevant studies.**

INDEX TERMS Cloud Security; Machine Learning; DDos; Privacy; Security

I. INTRODUCTION

Cloud computing is a technological advance that offers the facilities, platform and software of information technology as Internet services [1]. It is considered to be the conversion of a long-lasting dream called "Computing for Use" and it is being gradually embraced by organizations as private, public or hybrid Clouds [2]–[4] . Its main objective is to let users use and pay for what they want, promising on-demand services for their software or infrastructure needs [5]–[7]

Although Cloud computing is seen as a significant and positive IT infrastructure shift, much security work is still needed to minimize its deficiencies. Since a significant amount of personal and corporate information is placed in the Cloud data centers, those Cloud security issues and vulnerabilities need to be identified and prevented. Because Cloud infrastructure runs through standard Internet protocols and uses virtualization techniques, it may be vulnerable to attacks. Those attacks may come from traditional sources such as Address Resolution Protocol, IP spoofing, Denial of Service (DoS) etc [8], [9]. They may also come from other sources. Zero-day attacks, for example, referred to as unknown attacks, are seen as a significant challenge in the cyber security domain [10]. Traditional techniques used for detection and prevention are not efficient enough to handle those attacks while also dealing with a large data flow.

Machine Learning (ML) techniques are very helpful for identifying attacks, whether traditional or zero-day attacks. Machine learning includes a series of algorithms that can learn patterns from data and predict accordingly[11]. ML combines computer science and statistics to enhance the prediction [11]. ML comprises three main types of learning, supervised, unsupervised and semi-supervised [12] [13]. Supervised machine learning depends on classified data that are trained to build the classification model. Unsupervised learning algorithms enable training a model without guidance [11]. There are different algorithms for each, such as Nearest Neighbor, Naïve Bayes, Decision Trees, Linear Regression, Support Vector Machines (SVM)... etc. K-means clustering is an example of unsupervised algorithms. Deep Learning (DL) enables multi-layered computing models to learn data depictions with various abstraction levels [14]. It has achieved significant improvements in multiple applications such as image analysis, speech recognition and text recognition [15].

The main objective of this study is to conduct a systematic review of the ML techniques used to solve, detect or prevent Cloud security issues and vulnerabilities.

Despite the large number of research studies conducted on Cloud security using machine learning, to the best of our knowledge, there are very few Systematic Reviews on this topic. For our study, research papers were carefully collected and selected with regards to: (I) the ML techniques used for Cloud security, (II) the security areas that ML techniques are used for, and (III) the estimation and accuracy of the ML techniques used.

The remainder of this study is divided into five sections: Section II provides the literature review. The methodology for

conducting this review is described in Section III. The findings and results are listed in Section IV. Section V addresses the limitations of this review, while Section VI includes discussion and suggestions for future work.

Table 1: Literature Review: Survey papers that discussed similar topics as our paper and the difference between our work and their work.

Survey	Year	Description	Difference
"A survey on security challenges in cloud computing: issues, threats, and solutions" [16]	2020	This paper examines the numerous Cloud infrastructure components and emerging security and privacy challenges raised in cloud platforms. Besides, they provide a new classification of the current security technologies in this field	It covers Cloud computing and its security issues, threats, and provide solutions.
"A Survey on the Security of Cloud Computing"[17]	2019	This paper provides a review of the significant attacks against Cloud computing. Adding more, it offered solutions and possible countermeasures for comparative analysis	It covers Cloud security attacks, threats, and protection methods. Does not cover Machine learning methods
"On cloud security requirements, threats, vulnerabilities, and countermeasures: A survey"[18]	2019	This paper provides an extensive survey that defines a coherent security taxonomy, risks, vulnerabilities, and counter measurement criteria. Furthermore, it emphasizes security issues in other relevant fields, such as trust-based security architectures, large-scale, IoT, SDN, and Network Function Virtualization (NFV)	It covers a wide aspect of Cloud security issues varying from the vulnerabilities, risks, threats in different fields. But it does not cover machine learning techniques.
"A Survey of Cloud Computing Security Challenges, Issues and their Countermeasures"[19]	2019	This paper addresses problems related to security in Cloud at the conceptual level, data level, and reviews the cloud identities and Cloud access control. Further, the paper discusses several methods applied to prevent or mitigate Cloud security threats.	It covers Cloud computing and its security issues, threats, and provide solutions.
"Survey on android malware detection and protection using data mining algorithms" [20]	2019	This research paper reviews malware on mobile devices and compares data mining algorithms to find the most accurate one among them.	It covers ML techniques on mobile malware detection.
"Cloud security issues and challenges" [21]	2018	This paper reviews Cloud computing deployment models, their issues, and the issues of service models as well.	It covers Cloud security issues and challenges.
"Survey on machine learning algorithms as Cloud service for CIDPS" [22]	2018	This research study identifies current attack detection and prevention strategies based on ML, in addition to analyzing algorithms to get the most accurate algorithm.	It covers Cloud security in big data.
"A survey of deep learning-based network anomaly detection" [13]	2017	This paper provides an overview of deep learning techniques with a focus on network anomaly detection. It provides a background to the topic along with a literature review and an analysis to compare accuracies.	It covers Cloud anomaly detection using DL.
"A survey on attack detection on Cloud using supervised learning techniques" [23]	2016	This survey discusses Cloud architecture, types, risks and threats.	Covers only Cloud computing and its security issues only.
"Security and privacy for big data: A systematic literature review" [24]	2016	The purpose of this research is to categorize and analyze, both in a quantitative and a qualitative way, big data papers related to security or privacy.	It covers the big data issues in Cloud.
"A review on intrusion detection techniques for Cloud computing and security challenges" [25]	2015	This research study presents an overview of Cloud intrusion attacks, types of systems, and analysis of existing techniques. One of those techniques is ML.	It covers Cloud intrusion detection only.
"A survey on security issues and solutions at different layers of Cloud computing" [26]	2013	This paper discusses vulnerabilities, threats and attacks to Cloud computing in detail.	It covers Cloud security only. They do not discuss ML techniques.
"A survey of intrusion detection techniques in Cloud" [27]	2013	In this research paper, forms of intrusion that can threaten the integrity, confidentiality and availability of Cloud services, are discussed, along with the techniques for solving these threats.	Although some of the solutions are ML related, they are not thoroughly discussed.
"An intrusion detection and prevention system in Cloud computing: A systematic review" [28]	2013	This is a thorough review of intrusion detection along with the methods and comparative analysis of their features.	It covers intrusion detection in Cloud only.
"A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in Cloud computing" [29]	2012	This research study focused on Cloud computing gaps and identifying types of attacks and related solutions. ML techniques are used, and their accuracies are compared to determine the best choice.	It covers both Cloud security and ML techniques. However, those experiments are done by the authors and not by related researchers.
"Trust issues that create threats for cyber attacks in Cloud computing" [1]	2011	This paper is divided to two parts. The first is a survey on Cloud computing, including the gaps and threats. The second proposes solutions using ML techniques.	Although it covers Cloud computing, the proposed solution is limited to one "proactive attack detection."

II. Literature Review

A. Cloud Computing Security

In this section, we discuss the security and privacy issues that currently exist in Cloud computing. Cloud computing itself is a very broad field, because it transmits and hosts its facilities on the Internet. It provides services to meet the needs of the clients and charges accordingly [21]. This makes the Cloud crucial as people start to depend on it and organizations can now hire a Cloud service easily.

According to Khorshed et al. [1], gaps in Cloud computing are defined as the trust issues between customers and Cloud providers, where customers fear policies that are hidden from them. On the other hand, Cloud providers are afraid that customers might take advantage and conduct attacks using their Cloud services. The main determinants of the selection of a Cloud provider are the expectations of their organizations and what facilities they will obtain from a specific provider. Vulnerabilities, according to Modi et al. [26], are defined as Cloud safety loopholes, which an opponent can use to obtain access to the network and other infrastructure resources. A Cloud threat is a possible negative occurrence that can be malicious or incidental [26]. An attack involves a Cloud resource damage activity, and a vulnerability exploitation may influence the accessibility of Cloud computing and financial benefits [26].

Major vulnerabilities in Cloud computing that can pose serious threats are vulnerabilities in virtualization/multi tenancy, vulnerabilities in Internet protocol, unauthorized access to management interface, injection vulnerabilities, and vulnerabilities in browsers and APIs [26], [29]. Those vulnerabilities pose consequent effects, such as allowing network attacks, giving access control to intruders, allowing unauthorized service access and disclosure of private data. All of these vulnerabilities expose Cloud to threats, directly or indirectly, such as with business. Some of these threats are (i) the changes to a business model which can hinder the usage of Cloud computing services, (ii) abusive use of Cloud computing, (iii) insecure interfaces and API, (iv) malicious insiders, (v) data loss and leakage, (vi) service hijacking, and (vii) unknown risk profile.

In order to protect the Cloud from those threats and prevent any damage, the attacks that can be launched need to be identified and understood. The attacks most often discussed in Cloud computing [23], [26], [29] are the following:

Denial of Service (DoS) attack: is an attempt to affect service availability for users. Distributed Denial of Services (DDoS) is used to launch DoS using multiple computers.

Zombie attack: when an attacker floods the victim with requests from innocent hosts in the network. Such an assault interrupts Cloud's anticipated behavior, influencing the accessibility of Cloud services.

Phishing attack: is an attempt to manipulate and gain personal information from innocent people by redirecting them to a false link. At Cloud, an attacker may be hosting a Cloud service to hide the accounts and services of other Cloud users via a phishing attack site.

Man-in-the Middle attack: where an attacker is able to access the communication path between two users. An intruder can access information interactions between data centers in the Cloud

There are other attacks such as Cloud malware injection attack, breach of confidentiality, authentication attacks, attacks on virtualization, etc.

B. Attack Detection for Cloud Computing using Machine Learning Techniques

From the papers we have collected, we noticed that ML is used in many ways for Cloud attack detection. One of the main ways is traditional detection, where it detects and alerts users when an attack happens. Another detection approach is preventing the attack before it happens by checking the Cloud security itself for any gaps or vulnerabilities.

Table 1 represents related work or surveys/reviews conducted on cloud security issues or ML techniques used for cloud security. It shows the paper title and year it was published, describes what the paper discusses, and notes the difference between our review and the paper.

The papers above include cloud security issues and some of them include ML techniques. However, they either limit the research to one or two cloud security issues or do not include ML techniques at all. Our research reviews cloud security that uses ML techniques in their research up to the present date.

Our systematic review differs from those described above as we present a comprehensive cloud security research study with machine learning techniques. Moreover, to the best of our knowledge there is no systematic literature review that covers the same areas our review provides. Adding more, our study differs in several aspects from the related work listed in Table 1, such as:

1. Review machine learning techniques, the types of the models and whether the model is hybrid or standalone.
2. Precision comparison of the advantages and disadvantages of each technique.
3. A comprehensive cloud security analysis of the issues in this area.
4. Present the highest precision values in terms of security area.
5. It covers the very recent period from 2004 to 2019.

III. Methodology

In this review, we conducted a systematic review based on Kitchenham's and Charters' methodology [30]. Their methodologies divide the process into several phases, and each phase includes several stages. These phases are planning, conducting and finally reporting.

The following sections illustrate the review protocol that this paper followed.

A. Research Questions

This SLR aims to summarize and clarify the Machine Learning (ML) techniques and implementations that were used in Cloud security from 2004 to 2019 inclusive. Towards this end, the following three research questions (RQs) are raised:

RQ1: Which Cloud security areas are addressed by this review?

RQ1 addresses the Cloud security areas that are explored in the collected papers, including the categories, number of studies and whether they are conference papers or journal papers.

RQ2: Which machine learning algorithms are used in Cloud computing security?

RQ2 addresses machine learning model type, analysis type and features used in the collected papers. This RQs analyzes the common features among the studies.

RQ3: What is the overall estimation and accuracy of machine learning models?

RQ3 focuses in four aspects of estimation accuracy found on the papers that mention them: the accuracy metric, accuracy value, data set of construction, and model validation methods. It compares these aspects with the other papers.

B. Search Strategy

This section of the paper presents on the following:

The main search terms from the research question are identified.

New terms have been defined to substitute for main terms

To make search results more relevant, Boolean logic is added in the form of search operators. (AND, OR, quotations, parentheses)

We used search terms that are related to Cloud Security and Machine learning, such as "Cloud security" AND "(Machine learning" OR "ML").

Survey resources

In the search for the necessary research papers, the following digital libraries were used:

- Google Scholar
- ACM Digital Library
- Springer Online
- Scopus
- IEEE Xplore

In this review, 63 publications were used, based on our inclusion / exclusion criteria; 24 of them are Journal papers and 36 are conference papers.

C. Study Selection

We originally obtained 230 search studies based on the specific search conditions. The authors carried out further filtration to guarantee that only appropriate documents were included in this review as shown in Table 2.

Table 2:Inclusion and Exclusion criteria

Inclusion criteria	Exclusion criteria
• Using machine learning in the Cloud security area • Using hybrid models that employ at least 2 machine learning techniques for Cloud security • Include journal papers and conference papers only • Scopus indexed conference papers • Q1,Q2 ranked Scimago journal papers	• Papers that use machine learning in an area other than Cloud security • Papers that discuss Cloud security without machine learning • Non-refereed publications

D. Quality Assessment Rules (QARs)

The last stage in identifying the final list of articles included in this review was the application of quality assessment rules. We developed six QARs to determine the quality and relevancy of the articles to our research, where each QAR is worth 1 mark out of 10. Each QAR score is assigned as follows: "fully answered" = 1, "above average" =0.75, "average" = 0.5, "below average" = 0.25, "not answered" = 0. The total value of all 6 QARs is the overall score of each paper. A rating of less than 3 implies that this review has not included the document. The following QARs were used:

1. Are the objectives of the study clearly stated in the article?
2. Is the paper well-structured?
3. Does the article provide appropriate background information?
4. Is the specific area of Cloud security clearly defined?
5. Are machine learning techniques explained in enough detail?
6. Are the outcomes and conclusions of the experiment clearly reported?

Accordingly, we got 63 relevant papers with a quality score higher than or equal to 3. The quality scores of these studies are presented in the Appendix B.

E. Data Extraction Strategy

This step aims to provide a semi-structured response to the research questions for each article. The following data is obtained from every article: paper number, paper title, publication year, publication type, domain, RQ1, RQ2 and RQ3. It should be noted that not all articles addressed all research questions.

F. Synthesis of extracted data

In order to synthesize the data derived from the selected papers, we use different methods to extract evidence to address the RQs. This explains in detail the synthesis procedure we adopted:

- RQ1 and RQ2: narrative synthesis method is used to tabulate extracted information according to RQ1 and RQ2.
- RQ3: for quantitative data extraction, we use binary outcomes to calculate the results of RQ3.

IV. Results and discussion

For each RQ, the outcomes of this SLR will be provided and addressed in the following subsections. Appendix A shows all papers collected with their IDs and titles.

A. RQ1: Cloud security areas

From the 63 papers we have collected, we deduce that 11 Cloud security issues are addressed and researched. These are: anomaly detection, attack detection, confidentiality of data, data privacy, DoS, DDoS, intrusion detection (ID), malware, privacy preservation, security and vulnerability detection.

Table 3 shows the number of research papers in each cloud security area and the frequency of the area, as well as the

percentage.

Table 3:Cloud Security Area found from collected research papers

Cloud Security Applications	Research Papers	Percentage
Anomaly Detection	6	9%
Attack Detection	4	6%
Confidentiality of Data	3	5%
Data Privacy	9	14%
DoS	3	5%
DDoS	10	16%
ID	9	14%
Malware	6	10%
Privacy Preservation	6	10%
Security	5	8%
Vulnerability Detection	2	3%

Detection of anomalies involves finding patterns in data that do not correspond to anticipated behavior. Anomaly detection is important because data anomalies are essential and often critical information that can be acted on in a broad range of applications. We found a total of 6 papers that cover anomaly detection. However, we noticed that 3 of these papers cover anomaly detection in user behavior. As shown in Table 4, the first three papers discuss frameworks or systems involving anomalies detection in the Cloud. Even though A4-A5 involve research on anomalies, they focused on behavior anomalies.

Table 4:Papers that discusses anomaly detection security aspect in Cloud

ID	Paper summary on Anomaly Detection
A1	Develops a system to improve the accuracy of detection systems at the hypervisor Cloud layer
A2	Demonstrates the feasibility of ML techniques for anomaly detection and categorizes attacks on Cloud
A3	Proposes an approach for detection and classification of attacks in network traffic
A4	Proposes a framework to operate anomaly detection for profiling user's behavior
A5	Presents a mobile Cloud infrastructure that monitors abnormal behavior and detects malware
A6	Proposes an authentication mechanism based in verifying facial features along with two factor authentications

One of the top Cloud security area discussed in our sample of papers is data privacy. Security and privacy are the biggest challenges for both clients and Cloud service providers, since many confidential and sensitive data are stored in the Cloud that can be of value to an attacker [31].

Murakami et al. [32] propose a system that encrypts a secret message that is embedded into an image file by using a dynamically generated morphing image based steganography technique. This method ensures the security of the data because human beings cannot perceive the image hiding the message internally.

Another paper that is about encrypting Cloud data is by Wang et al. [33]. However, this paper uses another approach where users encrypt their data using their own key and store the data in the Cloud so that they can securely access and retrieve them without compromising data privacy. They propose two efficient schemes, Vitamin+ and Vitamin*.

Eskandari et al. [34] focus on the data privacy of geolocation that is stored and processed in the Cloud. In some specific compliance situations, knowing and managing the physical place of information for storing and handling reasons could be crucial for organizations using Cloud. Therefore, VLOC (a Verifier for physical LOCation of a virtual machine) a geolocation approach that is able to verify the physical location of a VM and notify the user if the location is unauthorized—comes into play. Moreover, it does not rely on a network of fixed landmarks. The experimental results indicate that VLOC is accurate enough to be used in practice.

According to Babu et al. [35], data breach is the biggest problem in the Cloud, and an insider attack is the worst threat. Hence, they propose an approach using a hot based user profiling technique that analyzes user keystrokes to provide authentication to the user. In the case of an abnormality in the behavior, an alarm will be raised and the current session in VM will be locked.

The rest of the papers are summarized in Table 5. One of the top Cloud security area discussed in our sample of papers is data privacy. Security and privacy are the biggest challenges for both clients and Cloud service providers, since many confidential and sensitive data are stored in the Cloud that can be of value to an attacker [31].

Murakami et al. [32] propose a system that encrypts a secret message that is embedded into an image file by using a dynamically generated morphing image based steganography technique. This method ensures the security of the data because human beings cannot perceive the image hiding the message internally.

Another paper that is about encrypting Cloud data is by Wang et al. [33]. However, this paper uses another approach where users encrypt their data using their own key and store the data in the Cloud so that they can securely access and retrieve them without compromising data privacy. They propose two efficient schemes, Vitamin+ and Vitamin*.

Eskandari et al. [34] focus on the data privacy of geolocation that is stored and processed in the Cloud. In some specific compliance situations, knowing and managing the physical place of information for storing and handling reasons could be crucial for organizations using Cloud. Therefore, VLOC (a Verifier for physical LOCation of a virtual machine)— a geolocation approach that is able to verify the physical location of a VM and notify the user if the location is unauthorized—comes into play. Moreover, it does not rely on a network of fixed landmarks. The experimental results indicate that VLOC is accurate enough to be used in practice.

According to Babu et al. [35], data breach is the biggest problem in the Cloud, and an insider attack is the worst threat. Hence, they propose an approach using a hot based user profiling technique that analyzes user key strokes to provide authentication to the user. In the case of an abnormality in the behavior, an alarm will be raised and the current session in VM will be locked.

The rest of the papers are summarized in Table 5.

Table 5: Papers that discusses Data Privacy in Cloud

ID	Paper summary on Data Privacy

A29	Developed a model that provides high level of protection to sensitive information for distributed privacy preservation online learning
A30	Proposes a model to encrypt private data and employs Cloud servers to perform the high-order back-propagation algorithm on the encrypted data efficiently for deep computation model training
A47	A new scheme is proposed on encrypted Cloud data that is able to resist potential attacks
A50	Proposes a scheme that guarantees the security of input dataset and output results, while being secured from Cloud so that it cannot learn about the training data owners

The fundamental function of IDSs, which calls for elevated accuracy, low false alarm rates and effectiveness to predict alarms based on positive or true alarms when intrusion occurs and false positive or false alarms in the event of a failure [8]. Those can be used to defend the systems from different kind of attacks.[11] Intrusion, which implies access to the scheme by force or without consent from anyone, is the biggest threat to the Internet [11]. Nine of the selected papers take up intrusion detection for Cloud security as seen in Table 6. They propose or develop techniques and systems to secure Cloud.

Table 6: Papers that discusses ID in Cloud

ID	Paper summary on ID
A49	Proposes techniques for an efficient speedy detection mechanism using Machine learning and parallelization
A45	A highly accurate neural network based IDS is built on a Cloud platform
A43	Presents a method that addresses the intrusion severity analysis problem for Cloud using a machine learning approach
A42	A two tier architecture of security system is proposed for Cloud data. The second tier is SVM based IDS that is designed for a Cloud server to detect intruders
A37	Proposes an Incremental k-NN SVM intrusion detection method that has the ability to learn and update with new data in an acceptable amount of time
A24	Demonstrates a way to increase intrusion detection accuracy for a robotic vehicle using RNN-based deep learning, enhanced by LSTM
A13	An efficient security architecture is proposed that deals with intrusions at the Network and Virtualization layer in the Cloud Environment
A40	Applies ML using SVM as base classifier for detecting intrusions in network, and compares these to find the best Multi-Classifier algorithm that can outperform SVM
A60	Proposed a multi-cloud cooperative IDS that exploits previous feedback data to make decisions about suspicious intrusions

Privacy plays a main role in the management of many security services. Low tier encryption and decryption are given before the data is secured. Different public and private keys are needed to provide encryption and decryption [36]. Through privacy, we can safely and securely maintain and communicate information in any channel. six papers cover privacy preservation.

According to Hesamifard et al. [37], ML algorithms based on deep neural networks are the mainstream in current AI research. However, training the models requires access to raw data that are often privacy sensitive and might create privacy risks. Hence, they provide a solution that enables parties to use the service without revealing their sensitive data to other parties. This is done by applying neural network algorithms to encrypted data.

A similar problem is presented by Yuan et al. [38] However it focuses on back-propagation neural network learning privacy preserving. In order to enhance the precision of the learning outcome, multiple parties can cooperate on the union of their corresponding information sets by undertaking joint back-propagation neural network learning. Yet none of these parties wants to disclose their private data to others, so Yuan solves this problem by having parties encrypt their data privately and upload them to the Cloud. This enables the Cloud to execute the operations pertaining to the learning algorithms without knowing the original private data.

Ma et al. [39] propose a novel privacy preservation deep learning model, named PDLM, to solve privacy issues around collected data used for deep learning training. The PDLM applies deep learning over the data owners' encrypted data under multiple keys and uploads the encrypted data to service providers. Service providers and the Cloud platform train the model based on the multi-key encrypted data with a privacy preservation calculation toolkit.

Even though the Wang model [40] is discussed in data privacy Cloud security, it also covers privacy preserving for online learning. EXPLORER framework offers an additional tool for privacy preservation where it provides a high-level guarantee for protecting sensitive encrypted information that is exchanged between the server and the client.

Table 7 summarizes the rest of the papers in privacy preservation area.

Table 7: Papers that discusses Privacy Preservation in Cloud

ID	Paper summary on Privacy preservation
A35	Designs a simple framework for a secure outsourced collaborative data mining scheme with multi-owners in Cloud and proposes multiple enhanced frameworks and schemes
A32	MSCryptoNet is a novel framework which is a collaborative privacy preservation deep neural network architecture based on a fully homomorphic cryptosystem

Denial of service attacks (DoS) prevent access to the network and other resources by lawful customers. Distributed denial of service (DDoS) is a type of DoS attack, in which the attacker uses a bunch of remotely controlled computers for the attack instead of a single machine [41]. A total of 13 papers cover this type of security, 3 of them focus on Dos and the rest on DDos.

He et al. [41] propose a detection system for DoS attack to prevent attacks on the source side in the Cloud. The paper analyzes in their framework the statistical features of different kinds of attacks including flooding attacks, spoofing attacks and brute-force attacks. Those 3 attacks are the most prevalent DDoS attacks.

To protect virtual machine (VM) against Dos attacks in a Cloud environment, Kumar et al. [5] propose a platform named Eucalyptus. This platform is an intrusion detection system that is designed to detect DoS attacks on VMs in the Cloud by any external or internal machine in the Internet.

According to Chonka et al. [42], two of the major threats to Cloud computing are HTTP Denial of Service and XML Denial of Service. For this, SOTA model (Service-Oriented Traceback Architectural) is updated to a Cloud model to protect Cloud computing from X_DoS?H-DoS attacks. It traces back to find

Table 8: ML Models used in each research paper collected

ML Models	Ref.	Freq.	ML Models	Ref.	Freq.
SVM	,A39 ,A38 ,A37 ,A35 ,A28 ,A18 ,A16 ,A11 ,A8 ,A3 A55 ,A54 ,A42 ,A41 ,A40	15	Linear SVM	A11	1
Random Forest	A59 ,A58 ,A40, A56 ,A13 ,A9 ,A5 ,A2	8	C2DF	A12	1
KNN	A62, A61, A58 ,A47 ,A46 ,A37 ,A35 ,A6	8	Polynomial Regression	A21	1
Decision Tree	A26 ,A44 ·A49 ,A51 ,A53	5	RNN	A24	1
Back Propagation NN	A23, A30, A33	3	PDLM	A31	1
ANN	A45, A46, A53	3	DNN	A32	1
Linear Regression	A2, A13	2	Fuzzy C-means	A38	1
Linear Means Classifier	A7, A22	2	C4.5	A43	1
Neural Network	A55 ,A49 ,A19	3	LS-SVM	A48	1
One-class SVM	A25, A34	2	CKNN	A52	1
Naïve Bayes	A49, A50	2	FCM-ANN	A1	1
K-means	A56 ,A17	3	Logistic Regression	A29 ,A60	2
Linear Kernel	A14, A16	2	XGBoost classifier	A36	1
Fisher's Linear Discriminant (FLD) Classifier	A7	1	CNN	A57	1
Bayes net	A10	1	LSTM	A57	1

the source of these attacks and detects them though a back propagation neutral network, called Cloud Protector.

SDN controller, software defined network, manages the whole system yet is vulnerable to any DDoS attack that will cause paralysis of the entire network. XGBoost [43], extreme gradient boosting, is a detection method in SDN based Cloud that is proven to have a higher accuracy and lower false positive rate than other algorithms.

Modi et al. [44] propose a structure that integrates a network intrusion detection system (NIDS) into the Cloud infrastructure to detect and prevent DoS arracks and other malicious activities at the network layer. This is done by monitoring network traffic while sustaining performance and service quality. The remaining papers are shown in Table 9.

Security is researched in 5 of the selected papers. The research of Gai et al. [45] shows concern about privacy information leakage among financial service institutions and customers. They propose the SEB-SIC model, Supervised Earning-Based Secure Information Classification, which classifies information in a way that avoids leaking data that can be harmful to the institution or its customers. Using decision tree techniques, the model predicts the potential risks of the data that is shared between institutions, along with reducing the chances of privacy leakage.

Table 9: Papers that discusses DDoS in Cloud

ID	Paper summary on DDoS
A27	Proposes a self-learning method that adapts a detection model to network changes, which minimizes false detection and marks legitimate users as malicious or vice versa.
A48	Presents a new classifier system to detect and prevent DDoS TCP flood attacks in public Clouds by classifying incoming packets and making a decision based on the classification results
A51	Uses data mining techniques to detect internal and external DDoS attacks in Cloud computing. In addition, it proposes an architecture of the Cloud security component for detecting DDoS attacks without affecting a client's data
A52	An approach is presented for detection of DDoS attacks based on CKNN that is able to detect the existing attacks by examining flow features only
A53	Uses the ANN algorithm to detect and mitigate known and unknown DDoS attacks in real time environments based on patterns that separate DDoS attack traffic from real traffic
A12	Presents a statistical technique to detect and filter DDoS attacks and has the ability to mitigate most TCP attacks
A54	Uses different ML algorithms to test and analyze their accuracy and compare against different types of DDoS attacks
A55	Addresses Economic Denial of Sustainability, that is a form of DDoS attack, and proposes a framework to detect it by learning the types of attacks and classifying normal and abnormal behaviors

To secure password authentication, Omri et al. [46] uses handwritten recognition to secure access to data in the cloud on mobile phones. According to their research, biometrics provides better security than traditional authentication methods. Their framework is composed of pre-processing, feature extraction, classification and authentication process.

The issue of the security of medical image analysis over cloud computing is addressed by Marwan et al. [47]. They propose a method that alleviates security and privacy concerns when preforming image analysis using cloud computing. Their cloud framework is designed to not reveal hidden medical data to cloud providers. In addition, it secures data processing in the cloud environment. The remaining papers are listed in Table 10.

Table 10: Papers that discusses Security attacks in Cloud

A54	Focuses on protecting communication and proposes a new firewall scheme that detects and classifies traffic packets and malicious users
A56	Propose and design an architecture that detect and classify network threat based on ML approach to respond to network threats. They build it in a hybrid way to improve performance of detection and classification

Although Khorshed et al. [1] presents a survey of gaps that

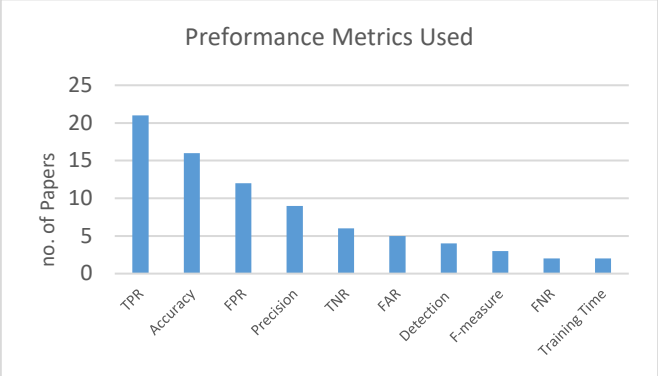


Figure 1: Performance metrics usedin collected papers

slow down cloud adoption at the beginning of the research, they still propose a solution for attack detection. Their "Proactive Attack Detection" model is able to detect an attack when it starts or during the attack. It also alerts the customer if the Cloud provider tries to hide attack information. Testing several popular ML techniques, they found SVM to be the most efficient for their model.

Moon et al. [48] on the other hand, propose and analyze attacks in CMS, CyberManufacturing system, which is considered as a blueprint for future manufacturing systems. CMS has multiple layers, therefore attacks on CMS is enlarged by the additional layers and Internet connections. As a solution of detecting malicious attacks, they use ML methods on CMS environment for security and in 3D printers and CNC machines, providing experimental results in the end. The remaining papers are listed in Table 11.

Table 11: Papers that discusses Attack Detection on Cloud

A57	Propose an approach for detecting attacks at VM-layer in cloud environment, which is applicable to most cloud service models such as Saas, Paas and Iaas. They apply DL on 2 layers, CNN for layer-1 and LSTM for layer-2.
A59	Focus on protecting Paas against malicious behavior by proposing a noble security mechanism that classifies malicious behavior of worker threads of web applications.

Graepel et al. [49] demonstrate a way to implement confidentiality of ML training and test data. According to their research, encrypting the data before uploading it to the cloud is one way to preserve confidentiality. However, this may limit the utility of the data. Therefore, the homomorphic Encryption scheme makes it possible to provide confidentiality, by applying polynomial approximations to known ML algorithms.

A new system is developed by Vijayakumar et al. [50] that assesses the vulnerabilities on the applications before and after deploying them into the Cloud. The system assesses the online vulnerabilities at regular intervals and checks to see if there is any change in the structure of the code or the code itself in the application.

B. RQ2: Machine learning algorithms

Machine learning offers a highly responsive and automated security solution, and it is used since it solves security problems and handle data in a more effective way. Instead of focusing

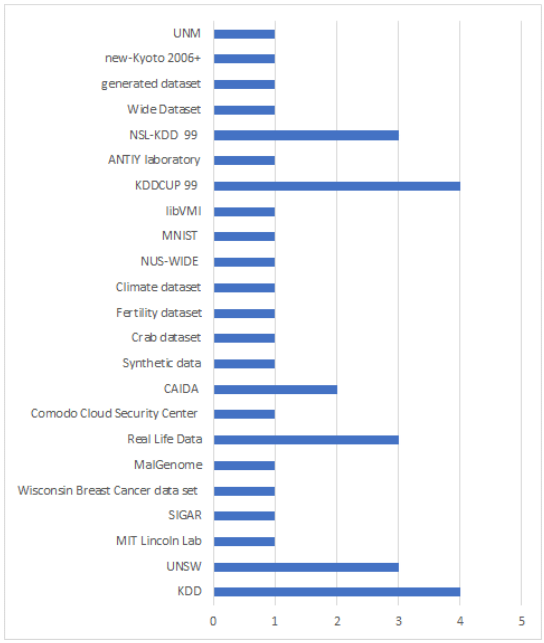


Figure 2: Data sets used in collected papers

only on detecting confidential data trends, ML solutions should use a comprehensive approach to protecting organizational information across all cloud applications of an organization. ML focuses on advancing computer programs which can find the right rate for their own learning [51].

We identified ML algorithms that had been applied by researchers in Cloud security areas. The list of these algorithms is shown in Table 8.

Out of the 30 ML techniques used, SVM is the most common.

LS-SVM, One-class SVM and Linear SVM are part of SVM technique. LS-SVM and One-class SVM have been used as standalone models. Table 12 shows the hybrid models used for SVM and other techniques.

Random Forest was used 8 times in total. Research paper A43 uses C4.5 algorithm, which is part of Random Forest. Four papers (A2, A13, A40, A59) used it as a standalone model, while 3 out of 6 papers used KNN as a standalone model. CKNN is part of KNN that is proposed by A52 research and is a standalone model. It is a lightweight method used to detect DDoS attacks. Decision Tree is used as a standalone method in A26, A44, A51 and A53 research papers.

Ten models from Table 8 are DL models. DL uses the backpropagation algorithm to discover complex structures in large information sets [14]. Neural networks (NN) have been developed as an important classification tool [52]. Those models are ANN, DNN, Back Propagation NN, PDLM and KNN.

Table 12: Papers that used Hybrid Models

Hybrid Model	Reference
Linear Regression + Random Forest	A2
SVM + Linear SVM	A11
Random Forest + Linear Regression	A13
KNN + SVM	A37
SVM + Fuzzy C-Means	A38
SVM + Random Forest	A40
ANN + KNN	A46

Neural Network + Naïve Bayes + Decision Tree	A49
SVM+ Neural Network	A54
CNN + LSTM	A57

According to Table 8, SVM is also the most used technique in the hybrid models. All hybrid models were found to use two techniques except A49, which combines 3 techniques to provide efficient detection speed and mechanism.

To choose the right ML technique or to prove that it is effective and accurate, some papers compare their models to other ML. Almost 60% of the papers collected use the comparison method. Some compare with multiple types of ML, others with only one ML that is the most relevant to their model.

Appendix table D present each research paper ID along to the ML technique applied and the advantages and disadvantages of that technique.

C. RQ3: Estimation and accuracy

For this section, we assembled all of the Estimations of accuracy that we found from the collected papers. We focused on four aspects: the accuracy metric, accuracy value, data set of construction, and model validation methods.

Thirty-six papers showed performance metrics in their research. After collecting the metrics they used, we ended up with more than 30 metrics. For this reason, we will include only the metrics that are used more than once. From the 13 metrics that we got; True Positive Rate is the most used among them. It is also known as sensitivity, recall or detection rate. TPR is the value of normal data correctly predicted or classified. 16 papers use Accuracy for performance evaluation as it shows the efficiency of their ML model. False Positive Rate is used by 12 papers, where a value of normal data incorrectly predicted or classified. Precision is how often a model correctly predicts a positive result. 6 papers used True Negative Rate, also known as specificity, to get the value of normal data that are predicted as normal. Following this in frequency of use is False Alarm Rate, which is used in 5 studies. This is followed by Detection value and Specificity, which are both used in 4 studies. The mean of recall and precision is F-measure, also known as F-score or F-value. False Negative Rate is used only twice to calculate the data that were falsely predicted. Figure 1 shows the performance metrics

Data sets are essential for arriving at an evaluation of the model and play an essential part in getting the best result. Thus, we conducted a review of the data sources that were used in the relevant articles, ending up with a total of 36 datasets. Figure 2 represent the datasets that were used, along with their frequency. As shown, KDD and KDD CUP '99 are the most used among the datasets, where 4 research papers in each dataset were used for evaluation. Following them is the NSL KDD '99 dataset and Real datasets, which is used in 3 papers. CADIA and UNSW are used in 2 papers each. The rest of the datasets are only used once.

D. Critical Analysis

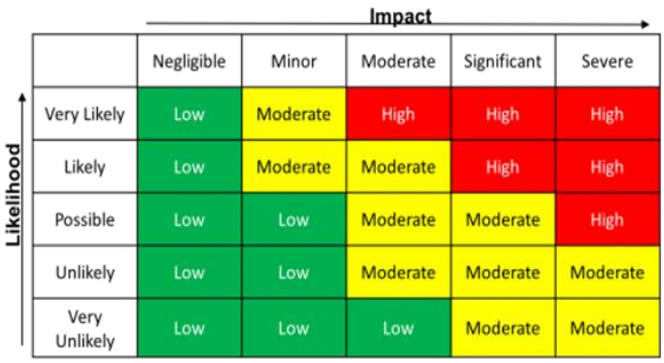


Figure 3: Risk Levels in terms impact and the risk likelihood

In this section, we present the analysis we established after answering the research questions, along with the open 3 challenges and future works of some papers. Table 14 shows the overall summary of the research papers that showed the highest accuracy according to their security aspect. For example, paper A58 covered attack detection in cloud security and the model they presented had an accuracy of 96.67% which is considered to be the highest among other collected research articles that covered the same security aspect. Appendix table C summarizes the performance metrics used in the papers collected that mentioned their research results.

In accordance with policy IT-19, Institutional Data Access, Business Owners (as defined in IT-16, Roles and Responsibilities for Information Security Policy) classify systems into one of three risk categories:

- Low Risk
  - Cloud processes and/or stores public data
  - Cloud is easily recoverable and reproducible
  - Cloud provides an informational / non-critical service
- Moderate Risk
  - Cloud processes and/or stores non-public or internal-use data
  - Cloud is internally trusted by other networked systems
  - Cloud provides a normal or important service
- High Risk
  - Cloud processes and/or stores confidential or restricted data
  - Cloud is highly trusted by UI networked systems
  - Cloud provides a critical or campus-wide service

Risk Analysis must take into consideration the sensitivity of data processed, as well as the likelihood and impact of potential threat events. These probabilities into risk levels and an overall system risk level.

Although there are many threat events related to a system, they can be generally organized into three main categories:

- Loss of Confidentiality:
  - The system and its data are compromised
  - The system and its data are released publicly

Table 14: Highest Accuracy Values in Collected Research Papers According to the Security Aspect

Security Aspect	Papers with highest accuracy metric	Consequences and risk level
Anomaly Detection	paper A2 has the highest accuracy 99%, as well as paper A12 with an accuracy of 97%.	Low
Attack Detection	paper A58 has an accuracy of 96.67% which is the highest among other collected papers.	Moderate
Confidentiality of Data	paper A1 reported an accuracy of 94.52% which is the highest in this security aspect.	High
Data privacy	paper A46 used recognition rate as a performance metric and it has the highest value of 96.60%.	High
DDoS	papers A32, A55 had the significant accuracy 99.64% and 99.70% respectively.	High
Intrusion Detection	paper A36 has an accuracy of 98.53%	High
Malware Detection	paper A40 has an accuracy of 97.50% which is considered the highest among other papers in this security aspect.	Low
Privacy Preserving	paper A34 used detection rate as a performance metric and it has the highest value of 90%.	High
Security	paper A45 and A57 has the highest accuracy number which is 99% for both of them.	High, Moderate

  - The system and its data erroneously publish data on public
- Loss of Integrity:
  - The system and its data cannot be trusted
  - The system and its data are not complete or incorrect
- Loss of Availability:
  - The system and its data no longer exist
  - The system and its data no longer respond to valid queries
  - The system and its data cannot be retrieved by an authorized user (e.g. DDOS)

Risk levels are calculated as the product of the likelihood and impact of a potential threat event / threat event category. The risk level for each threat event category is then calculated. The overall risk level for the system is equal to the highest risk level for any risk event. Figure 3 shows the risk levels.

As discussed in many papers, there are many challenges to be addressed. In two papers, the need of more data that are properly trained is a challenge to them. Some papers find it a challenge to enhance and the ML algorithm, or that the ML contains a performance overhead such as paper A4. Paper A53 states that accuracy detection might degrade against some heavy attacks, while paper A32 states that the consideration of new features would naturally invoke a computational trade off. Furthermore, according to Table 13 we classified the open challenges into several fields machine learning, data, security, DDoS attack, and performance. There are 5 papers that discuss the open challenge of training the ML classifier or enhance the model. As for the data aspect, there are 4 papers that found challenge in the lack of data available due the privacy concern

Table 13: Summary of Open Challenges in Collected Papers

Field	Open Challenges	Paper ID
Machine learning	Training the ML classifier, classification generation, improving and enhancing the algorithm and model	A5, A25, A40, A43, A50
Data	The privacy concern of data sharing results in a need for more data or features to distinguish attacks	A9, A25, A28, A29
	Handling and analyzing big data as traditional data system is not suitable anymore	A14
	Training data properly and extract optimal features	A36
Security	Acquiring thousands of malware samples and limitation of dynamic analysis of the malware	A5, A37
	Cyber security functions are resource constrained	A23
DDoS	Handle HTTPs DDoS Attack traffic	A31
	DDoS issues with spoofed addresses	A47
	Introduce an interactive engine that searches for new DDoS attacks information	A52
Performance	Runtime problem cannot be applied in functional homomorphic encryption schemes due to activation functions	A17
	SMC's computational and coordination complexity	A22
	The problems facing public cloud studies are hypervisor noise	A15

Table 15: Future Work Similarity in Collected Papers

Future Work	Paper ID	Freq.
Evaluate other ML techniques, algorithms, and models. As well as consider combining the techniques to achieve higher results	A2, A3, A11, A13, A17, A19, A25, A27, A35, A38, A59	11
Enhance models from several aspects such as increasing detection rate, extend security, make it more scalable, and improve efficiency of the model	A5, A11, A14, A21, A24, A27, A29, A37, A47, A50	10
Add more applications to datasets, obtain more datasets, and conduct other and more experiments	A5, A16, A17, A25, A30, A51, A58	7
Investigate more security attacks , and feature selection techniques	A12, A13, A47, A52, A54, A55, A58	7
Implement on a real-world scenario, teat with real long term dataset scenario	A2, A16, A17	3
Research why selected features yielded high results	A4, A11	2
Consider other monitoring metrics and framework	A18, A20	2
Apply models to multi-cloud environment and study the measurement of trusted cloud environment	A9, A33	2
Develop real time mobile malware detection, investigate, and monitor mobile cloud infrastructure	A4, A18	2

of data sharing.

For future work, we noticed that most papers will further evaluate their research work by adding more tests or combining different ML. Other papers state for their future work is to improve their proposed work by adding more features or enable it to detect more types of attacks. Table 15 shows the similarity of future work of the collected papers. We found that 11 papers want to work on trying other ML classifiers or using hybrid methods. As well as 10 papers want to focus on enhancing their proposed model from several aspects, and other papers want to conduct more experiments in different datasets.

Furthermore, we collected the type of cloud security problems as reported in the papers. We found that almost 51% of the problems are considered as classification problems, 7% are clustering, and 5% are regression. Some papers used more than one type in same research. Please note that machine learning techniques are very powerful to tackle such problems.

As for the future work of our study, we plan to look for more security threats in cloud that can be solved through ML techniques. Moreover, we will explore more of ML techniques and which achieve the highest results in almost all security aspects in cloud. Furthermore, we will look into a wider aspect of the applications of Cloud security such as Cyber Physical Systems (IoTs) and SDNs.

V. Limitations of This Review

This work is restricted to journal and conference papers related to ML in Cloud security. By applying our search approach strategy, we excluded a large number of non-relevant research papers. The papers we selected fully match our research objective, hence the small number of papers collected. In addition, we applied quality assessment criteria to select articles that provide synthesized results.

VI. Conclusion

We carried out a systematic literature review to analyze ML techniques used in Cloud security. The review investigated relevant studies that answered 3 RQs; Cloud security area, type of ML techniques used, and the accuracy estimation of the ML model. Overall, we obtained 60 research papers after applying our selection criteria. Our conclusions are summarized as follows:

- RQ1 findings are the 11 Cloud security areas identified; anomaly detection, attack detection, privacy preservation, security, vulnerability detection, confidentiality of data, data privacy, DDoS, DoS, and intrusion detection (ID). DDoS and data privacy are analyzed the most, with a 16% frequency of usage and 14% respectively.
- RQ2 counted 30 ML techniques used, some used as hybrid and others as standalone. The most popular ML used is SVM in both hybrid and standalone models. 60% of the papers compared their models with other ML models to get the best evaluation to either prove their accuracy or to further improve their model.
- RQ3 enumerated 13 different evaluation metrics; TPR, Accuracy, FPR Precision, TNR, FAR, Detection, F-measure, FNR and Training time. The most used metric was TPR, and the least used was Training time, respectively. Furthermore, datasets have been used to evaluate models' performance. From the 20 datasets found, KDD and KDD CUP '99 are the most used.

Our study also found very few surveys based on ML techniques in Cloud security form, with no usage of their feature selection/extraction strategy. Therefore, we recommend more thorough research and more empirical experiments to address the need for ML in Cloud security. In addition, research papers should present their results using multiple evaluation metrics when considering imbalanced datasets.

Moreover, we noticed that little work has been done using deep learning techniques in cloud security. We encourage researchers to take advantage of the deep learning in this regard.

Another important observation is that most of the datasets used are relatively old such as KDD. Researchers are encouraged to use recent datasets such as CICIDS2017, CSE-CIC-IDS2018 and Kyoto 2006+ for intrusion detection.

Appendix

Appendix A: Referenced papers used in this research study

ID	Paper	Type	Year	Refs.
A1	"Anomaly Detection System in Cloud Environment Using Fuzzy Clustering Based ANN"	Journal	2015	[53]
A2	"Machine Learning for Anomaly Detection and Categorization in Multi-Cloud Environments"	Conference	2017	[54]
A3	"A hybrid machine learning approach to network anomaly detection"	Journal	2007	[55]
A4	"Frequent Pattern Based User Behavior Anomaly Detection for Cloud System"	Conference	2014	[56]
A5	"Monitoring and detecting abnormal behavior in mobile Cloud infrastructure"	Conference	2012	[57]
A6	"Insider Threat Detection with Face Recognition and KNN User Classification"	Conference	2018	[58]
A7	"ML Confidential: Machine Learning on Encrypted Data"	Conference	2013	[49]
A8	"Trust Issues that Create Threats for Cyber Attacks in Cloud Computing"	Conference	2011	[1]
A9	"Evaluation of machine learning classifiers for mobile malware detection"	Journal	2014	[59]
A10	"Applying machine learning classifiers to dynamic Android malware detection at scale"	Conference	2013	[60]
A11	"Combining file content and file relations for Cloud based malware detection"	Conference	2011	[61]
A12	"Statistical-based filtering system against DDOS attacks in Cloud computing"	Conference	2014	[62]
A13	"NvCloudIDS: A security architecture to detect intrusions at network and virtualization layer in Cloud environment"	Conference	2016	[63]
A14	"Continuous security assessment of Cloud based applications using distributed hashing algorithm in SDLC"	Journal	2017	[50]
A15	"Machine Learning Based DDoS Attack Detection from Source Side in Cloud"	Conference	2017	[41]

A16	"Secure Data Mining in Cloud Using Homomorphic Encryption"	Conference	2015	[31]
A17	"Cache-Based Application Detection in the Cloud Using Machine Learning"	Conference	2017	[64]
A18	"Privacy-preserving Machine Learning in Cloud"	Journal	2017	[37]
A19	"Improvement of security in Cloud systems based on steganography"	Conference	2014	[32]
A20	"VLOC: An Approach to Verify the Physical Location of a Virtual Machine In Cloud"	Conference	2015	[34]
A21	"Computing encrypted Cloud data efficiently under multiple keys"	Conference	2013	[33]
A22	"Privacy Preserving Back-Propagation Neural Network Learning Made Practical with Cloud Computing"	Journal	2013	[38]
A23	"Cloud-Based Cyber-Physical Intrusion Detection for Vehicles Using Deep Learning"	Journal	2017	[65]
A24	"Detecting Denial of Service Attacks in the Cloud"	Conference	2016	[5]
A25	"Security-Aware Information Classifications Using Supervised Learning for Cloud-Based Cyber Risk Management in Financial Big Data"	Conference	2016	[45]
A26	"Self-learning method for DDoS detection model in Cloud computing"	Conference	2017	[66]
A27	"Analyzing User Behavior Using Keystroke Dynamics to Protect Cloud from Malicious Insiders"	Conference	2015	[35]
A28	"EXpectation Propagation LOgistic REgRession (EXPLORER): Distributed privacy-preserving online model learning"	Journal	2013	[40]
A29	"Privacy Preserving Deep Computation Model on Cloud for Big Data Feature Learning"	Journal	2015	[67]
A30	"PDLM: Privacy-Preserving Deep Learning Model on Cloud with Multiple Keys"	Journal	2018	[39]
A31	"MSCryptoNet: Multi-Scheme Privacy-Preserving Deep Learning in Cloud Computing"	Journal	2019	[68]
A32	"Cloud security defence to protect Cloud computing against HTTP-DoS and XML-DoS attacks"	Journal	2011	[42]
A33	"Malware Detection in Cloud Computing Infrastructures"	Journal	2015	[69]
A34	"Secure Collaborative Outsourced Data Mining with Multi-owner in Cloud Computing"	Conference	2012	[70]
A35	"XGBoost Classifier for DDoS Attack Detection and Analysis in SDN-Based Cloud"	Conference	218	[43]
A36	"Incremental k-NN SVM method in intrusion detection"	Conference	2018	[8]
A37	"Security Enhancement in Healthcare Cloud using Machine Learning"	Conference	2018	[47]
A38	"Malicious Executables Classification Based on Behavioral Factor Analysis"	Conference	2010	[71]
A39	"A comparative analysis of SVM and its stacking with other classification algorithm for intrusion detection"	Conference	2016	[11]
A40	"Malware behavioural detection and vaccine development by using a support vector model classifier"	Journal	2015	[72]
A41	"Designing encryption and IDS for Cloud security"	Conference	2017	[73]
A42	"A novel intrusion severity analysis approach for Clouds"	Journal	2013	[12]
A43	"A novel framework for intrusion detection in Cloud"	Conference	2012	[44]
A44	"A neural network based distributed intrusion detection system on Cloud platform"	Conference	2013	[74]
A45	"Cloud-based mobile system for biometrics authentication"	Conference	2013	[46]
A46	"Secure and controllable KNN query over encrypted Cloud data with key confidentiality"	Journal	2016	[75]
A47	"An Efficient DDoS TCP Flood Attack Detection and Prevention System in a Cloud Environment"	Journal	2017	[76]
A48	"Efficient approaches for intrusion detection in Cloud environment"	Conference	2017	[77]
A49	"Secure Naïve Bayesian Classification over Encrypted Data in Cloud"	Conference	2016	[78]
A50	"DDoS Attacks Detection in Cloud Computing Using Data Mining Techniques"	Conference	2016	[79]
A51	"Detecting DDoS attacks against data center with correlation analysis"	Journal	2015	[80]
A52	"Detection of known and unknown DDoS attacks using Artificial Neural Networks"	Journal	2016	[81]
A53	"A Combined Decision for Secure Cloud Computing Based on Machine Learning and Past Information"	Conference	2019	[82]
A54	"Analysis and Detection of DDoS Attacks on Cloud Computing Environment using Machine Learning Techniques"	Conference	2019	[83]
A55	"Machine Learning-Based EDoS Attack Detection Technique Using Execution Trace Analysis"	Journal	2019	[84]
A56	"Design of network threat detection and classification based on machine learning on cloud computing"	Journal	2018	[85]
A57	"VMAnalyzer: Malware Semantic Analysis using Integrated CNN and Bi-Directional LSTM for Detecting VM-level Attacks in Cloud"	Conference	2019	[86]
A58	"Detecting cyber-physical attacks in CyberManufacturing systems with machine learning methods"	Journal	2019	[48]
A59	"DeMETER in clouds: detection of malicious external thread execution in runtime with machine learning in PaaS clouds"	Journal	2019	[87]
A60	"A deep learning approach for proactive multi-cloud cooperative intrusion detection system"	Journal	2019	[88]
A61	"K-NN classifier for data confidentiality in cloud computing"	Conference	2014	[89]
A62	"A machine learning algorithm TSF k-Nn based on automated data classification for securing mobile cloud computing model"	Conference	2019	[90]
A63	"Cloud based emails boundaries and vulnerabilities"	Conference	2013	[91]

Appendix B: QAR and the score of the collected papers

Paper	QA1	QA2	QA3	QA4	QA5	QA6	Score
A1	1	1	1	0.75	1	1	**5.25**
A2	1	1	1	0.75	1	1	**5.75**
A3	1	0.5	1	1	1	1	**5.5**
A4	1	1	0.5	1	0.5	0.75	**4.75**
A5	1	1	1	1	0.5	1	**5.5**
A6	1	1	1	1	1	0.25	**5.25**
A7	1	0.25	1	1	1	1	**5.25**
A8	1	1	1	0.5	1	1	**5.5**
A9	1	1	1	0.75	1	1	**5.75**
A10	1	1	0.5	1	1	1	**5.5**
A11	1	1	1	1	0.25	1	**5.25**
A12	1	1	0.75	1	0.5	1	**5.25**
A13	1	1	1	1	1	1	**6**
A14	1	1	0.5	1	0.25	0.75	**4.5**
A15	1	1	1	1	0.75	0.75	**5.5**
A16	1	1	0.75	1	0.25	0.25	**4.25**
A17	1	1	1	1	0.25	1	**5.25**
A18	1	1	1	0.5	0.75	1	**5.25**
A19	1	1	1	1	0.75	0.5	**5.25**
A20	1	1	0.5	1	0.25	0.25	**4.0**
A21	1	1	0.5	1	0.75	0.75	**5.0**
A22	1	1	0.25	0.75	1	1	**5.0**
A23	1	1	1	1	1	1	**6.0**
A24	1	1	0.75	1	1	1	**5.75**
A25	1	1	1	1	0.5	0.75	**5.25**
A26	1	1	1	1	0.5	0.5	**5.0**
A27	1	1	0.75	1	0.25	0.75	**4.75**
A28	1	1	1	1	1	1	**6.0**
A29	1	1	1	1	0.5	1	**5.5**
A30	1	1	1	1	0.75	1	**5.75**
A31	1	1	1	1	0.75	0.75	**5.5**
A32	1	1	1	1	0.25	1	**5.25**

A33	1	1	1	1	0.75	1	**5.75**
A34	1	1	1	1	0.5	0.5	**5.0**
A35	1	1	1	0.75	1	1	**5.75**
A36	1	1	0.5	0.75	1	1	**5.25**
A37	1	0.25	0.75	1	1	1	**5.0**
A38	1	1	1	0.75	0.25	0.25	**4.25**
A39	1	1	0.75	0.75	1	1	**5.5**
A40	1	0.75	1	1	1	1	**5.75**
A41	1	1	0.25	1	0.75	1	**5.0**
A42	1	1	1	1	1	1	**6.0**
A43	1	1	1	1	1	1	**6.0**
A44	1	1	0.75	1	1	0.75	**5.5**
A45	1	1	0.5	1	1	0.75	**5.25**
A46	1	1	0.5	0.75	1	0.75	**5.0**
A47	1	1	1	1	1	1	**6.0**
A48	1	1	1	1	0.5	1	**5.5**
A49	1	0.5	1	1	1	0.21	**5.25**
A50	1	0.5	1	1	1	1	**5.5**
A51	1	1	0.5	0.75	0.75	1	**5.0**
A52	1	1	1	1	0.75	0.25	**5.0**
A53	1	1	1	1	0.75	0.5	**5.25**
A54	1	1	1	1	0.5	0.75	**5.25**
A55	1	1	1	1	1	0.75	**5.75**
A56	1	1	1	1	0.5	0.5	**5.0**
A57	1	1	1	1	0.75	1	**5.75**
A58	1	1	1	1	1	1	**6.0**
A59	1	1	1	1	1	0.75	**5.75**
A60	1	1	1	0.75	0.75	1	**5.5**
A61	1	1	0.75	1	1	0.25	**5.0**
A62	1	1	0.5	1	0.75	0.75	**5.0**
A63	1	1	0.75	1	0.25	0.25	**4.25**

Appendix C: Performance Metrics used in collected papers

ID	Performance Metrics	Results
A1	Precision	94.52%
	Recall	60.73%
	F-Value	75.92%
A2	Detection Accuracy	99%
	Categorization Accuracy	93.60%
	False Alarm Rate	1.90%
	UND	2.00%
	Overall Error	0.90%
	Misclassification Error	0.06%
A3	Filtering Rate	94.40%
	Detection Rate	87.74%
	False Positive Rate	10.20%
	False Negative Rate	27.27%
A4	False Positive Rate	4.60%
	Detection	86.00%
A5	True positive Rate	98.80%
	False Positive Rate	0.80%
	Precision	98.90%
	Recall	0.988
A8	Classification Accuracy	96.27%
	Area Under ROC	1
A9	Accuracy	93.35%
A10	Correctly Classified	81.25%
	True positive Rate	97.30%
	False Positive Rate	31.03%
A11	True positive Rate	34675 file
	False Positive Rate	563 file
	True Negative Rate	356991 file
	False Negative Rate	1798 file
	Accuracy	99.40%
A12	Accuracy	97%
	False Alarm Rate	1%
	Time	0.31s
A13	Accuracy	94.54%
	False Positive Rate	2.81%
A14	Risk Identification Time	0.0005s
A15	False Negative Rate	6.60%
	Recall	93.40%
	F1-Score	0.9643
A24	Sensitivity	100% against ICMP Flood
	Specificity	100% against ICMP Flood
	Accuracy	100% against ICMP Flood
A26	Precision	90%
	Accuracy	95%
	F1-Score	94%
A33	Accuracy	Above 90%
A36	TPR	99.68%
	FPR	0.43%
	Training Time	9.32s
A38	Accuracy	83.30%
A39	Accuracy	97.50%
	Sensitivity	93.49%
	Specificity	98.38%
	Precision	97.60%
	Recall	97.60%
A40	Accuracy	94% out of 100 samples
	True Negative + True Positive	94 out of 100 samples
A41	Error Rate	5.05%
	Recall	98.9%%
	Performance	95%
A42	Success rate	Over 90%
A43	Precision	99.32%
	Recall/True Positive	96.25%
	True Negative	98.08%
	False Positive	1.91%
	False Negative	3.74%
	Accuracy	96.71%

A44	Accuracy	99%
A45	Recognition Rate	96.60%
	Error Rate	3.40%
A46	NA	NA
A47	Accuracy	97%
	Sensitivity	97%
	Specificity	97%
A49	Detection Rate	99%
A50	False Positive	0.05%
	False Negative	0.25%
	Recall	99.80%
	Precision	99.80%
	F-Measure	99.80%
A51	Accuracy	95% out of 7 samples
	Time	120s out of 7 samples
A53	Accuracy	98%
	Sensitivity	96%
	Specificity	100%
	Precision	100%
A54	Accuracy	89.96%

Appendix D: ML techniques used in collected papers with their pros and cons

Paper ID	ML technique	ML type	Advantages	Disadvantages
A1	FCM-ANN	classifier	The Nv Detector advance feature facilitates IP spoofing and advanced threat detection.	na
A2	Linear Regression + Random Forest	na	na	na
A3	SVM	classifier	We can solve a complex problem with kernel function. Also, SVM doesn't solve local optima. It manages the high dimensional data fairly well.	For large sets of data, the SVM algorithm is not suitable. In fact, when the data collection is noisy, SVM does not perform very well
A4	na	na	It improves the performance. In addition, its fast with an efficient computation and it known for its quick data training	Output is usually difficult to analyze. As well as, it is not practical with many-featured datasets.
A5	Random Forest	classifier	Can be adjusted in multi-arm correlations. The probability of each drug being the best treatment can be directly calculated	Cannot yield exact results for one loop network Needs more statistical knowledge than other approaches
A6	KNN	cluster (Fuzzy) + classifier (ANN)	na	Depends on the hardware.
A7	Linear Means + Fisher's Linear Discriminant (FLD) Classifier	classifier	na	na
A8	SVM	na	Support Vector Machine (SVM) displayed state of the art results in classification problems with the benefit of managing wide functional space without overfitting	na
A9	Random Forest	classifier and regression	It simplifies the assessment procedure	It is not suitable for noisy data; overfitting models occur with noisy data.
A10	Bayes net	classifier	Predictor performance can compete with the best controlled learning algorithms and provide a reliable estimate of feature importance	na
A11	Linear SVM	na	na	na
A12	C2DF	classifier	na	na
A13	Random Forest + Linear Regression	classifier	their approach can be extended to broader DOS attacks, as well as it doesn't degrade performance	They are only passive defenses after the attack, they cannot use the outbound statistical features of attacks, and it is hard to trace back to the attacker with these approaches
A14	Linear Kernel	cluster	na	Clustering of data in varying sizes and distribution is based on initial values.
A15	na	na	Open on the same framework for all processes	not suitable for large data sets
A16	Linear Kernel + SVM	na	na	na
A17	k-means	classifier	*Scales to large sets of data *Convergence protections. *Will heat up the centroid locations. *Adapts seamlessly to different cases	does not use bootstrapping because of the expense of computation and needs contact between client and server in place
A18	SVM	classifier	Predictive achievement can contend with the strongest supervised learning algorithms, offering a robust function evaluation	na
A19	Neural Network	na	na	na
A20	na	na	na	na
A21	Polynomial Regression	classifier	na	na
A22	Linear Means	classifier	Transfer information throughout the network, function with inadequate information	Depends on the hardware.
A23	Back Propagation NN	classifier	na	na
A24	RNN	classifier	can model data sequence (i.e. time series) so that every sample can be supposed to be dependent on previous ones and used to extend the effective pixel neighborhood with coevolutionary layers	If tanh or relu is an activation function, it cannot process very long sequences.

A25	One-class SVM	classifier	The key benefit is to use only patterns that relate to the target class distribution to train the classifier. When wide samples are accessible for correct labeling, the OC-SVM is efficient.	For large sets of data, the SVM algorithm is not suitable. If the data collection contains more noise, SVM does not do very well
A26	Decision Tree	classifier	na	na
A27	na	classifier	na	Na
A28	SVM	Regression	Detect known threats and determine whether or not the event is harmful	For large sets of data, the SVM algorithm is not suitable. If the data collection contains more noise, SVM does not do very well
A29	Logistic Regression (EXPLORER)	classifier	Don't require so much computing tools, they are easily interpretable	can't solve non-linear problems with logistic regression since its decision surface is linear
A30	Back Propagation NN	na	na	na
A31	PDLM	na	detect and filter out X-DoS messages	na
A32	DNN	classifier	produce a good decision function	recall is not good
A33	Back Propagation NN	classifier	ensure the outsourced data privacy	na
A34	One-class SVM	classifier	scalability and is well suited to the cloud with continuous expanding network scale	machine cannot continue to increase the number of threads due to its limited configuration. in SDN it CAN handle them
A35	SVM + KNN	cluster and classifier	model incrementally trained without steep climbing in training time and growth in prediction time	unstably in score of misclassification cost.
A36	XGBoost classifier	classifier	prevent the potential disclosure of confidential data	na
A37	SVM + KNN	classifier	na	na
A38	SVM + Fuzzy C-means	classifier	best performance in all evaluating parameters	na
A39	SVM	classifier	suitable means of clustering data from small data sets	selecting appropriate kernel function and parameters is difficult
A40	SVM + Random Forest	classifier	na	performance becomes poor if input data size is big
A41	SVM	classifier	simplistic and requires less training	if tree size is not considered it can cause an overhead
A42	SVM	classifier	detect known attacks and predict that the given event is malicious or not	na
A43	C4.5	classifier	better flexibility, scalability and performance.	na
A44	Decision Tree	classifier	each classifier treats all patterns independently from the other classifiers	na
A45	na	classifier	to detect known attacks and predict that the given event is malicious or not	na
A46	KNN	classifier	its capability for multi model classification	na
A47	KNN	classifier	Overall, the proposed CS_DDoS system is more effective and stable in resisting a single-source attack when adopting the LS-SVM classifier regardless of the window size and threshold.	not the least time-consuming
A48	LS-SVM	classifier	na	na
A49	Decision Tree + Neural Network + Naïve Bayes	classifier	predict the class label of unclassified samples	na
A50	Naïve Bayes	classifier	it constructs well-defined dependencies between different features for detecting the most popular DDoS attacks	when there is almost no traffic in the cloud network the false positive rate grows up.
A51	Decision Tree	classifier	- able to detect the existing DDoS attacks by examining flow features only - easy to be implemented since 549 it uses correlation information of training data	na
A52	CKNN	classifier	zero measure for the False Positive Rate.	the ANN only detects attacks that are similar to those it was trained with, and 5% of the attacks were completely different. ANN algorithm requires retraining every 5–6 years.
A53	Decision Tree	classifier	used in previous work related to security concept	na
A54	SVM	classifier	find global optimal solution	na
A55	Neural Network, SVM	cluster and classifier	improves the detection and classification of malicious users	there has to be enough number of packets in data set for machine to learn
A56	Random forest K means	classifier	provide a feature importance value for each used feature	na
A57	CNN LSTM	classifier	extract and select the relevant system call sequences and learn the behavior of observed processes	ne classifier which is good in detecting a particular type of intrusive, may not be good in detecting other type of intrusive process
A58	Random forest, k-nearest neighbors (k-NN) machine	classifier	na	different types of infill have a minor influence on system accuracy
A59	Random forest	classifier	best classification accuracy with enriched features	needs better data sets
A60	logistic regression	classifier	extract features that are robust to IDS feedback	na

Acknowledgment

We would like to thank University of Sharjah and OpenUAE Research and Development Group for funding this research study. We are also grateful to our research assistants who helped in collecting, summarizing, and analyzing the research articles for this SLR study.

Funding

This work is funded by OpenUAE Research and

Development Group, University of Sharjah

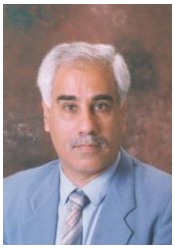
Availability of data and materials

The data extracted in this research is tabulated in the Appendix

Authors' information

ALI BOU NASSIF is currently the Assistant Dean of Graduate Studies at the University of Sharjah, UAE. Ali is also an Associate Professor in the department of Computer Engineering, as well as an Adjunct Research Professor at Western University, Canada. He obtained a Master's degree in Computer Science and a Ph.D. degree in Electrical and Computer Engineering from Western University, Canada in 2009 and 2012, respectively. Ali's research interests include the applications of statistical and artificial intelligence models in different areas such as software engineering, electrical engineering, e-learning, security, networking, signal processing and social media. Ali has published more than 65 refereed conference and journal papers. Ali is a registered professional engineer (P.Eng) in Ontario, as well as a member of IEEE Computer Society.

MANAR ABU TALIB is teaching at the University of Sharjah in the UAE. Dr. Abu Talib's research interest includes software engineering with substantial experience and knowledge in conducting research in software measurement, software quality, software testing, ISO 27001 for Information Security and Open Source Software. Manar is also working on ISO standards for measuring the functional size of software and has been involved in developing the Arabic version of ISO 19761 (COSMIC-FFP measurement method). She published more than 50 refereed conferences, journals, manuals and technical reports. She is the ArabWIC VP of Chapters in Arab Women in Computing Association (ArabWIC), Google Women Tech Maker Lead, Co-coordinator of OpenUAE Research & Development Group and the International Collaborator to Software Engineering Research Laboratory in Montreal, Canada.



QASSIM NASIR is currently an associate professor at the University Of Sharjah since 2009 and the chairman of scientific publishing unit. Dr. Nasir current research interests are in telecommunication and network security such as in CPS, IoT. He also conducts research in drone and GPS jamming as well. He is a co-coordinator in OpenUAE research group which focuses on blockchain performance and security, and the use of artificial intelligence in security applications. Prior to joining the University of Sharjah, Dr. Nasir was working with Nortel Networks, Canada, as a senior system designer in the network management group for OC-192 SONET. Dr. Nasir was visiting professor at Helsinki University of Technology, Finland, during the summers of 2002 to 2009, and GIPSA lab, Grenoble France to work on a Joint research project on "MAC protocol and MIMO" and "Sensor Networks and MIMO" research projects. Dr. Nasir has published over 90 refereed conferences, journals, book chapter, and technical reports.

HALAH ALBADANI is a research assistant at the University of Sharjah in the UAE. Halah is majoring in Information Technology Multimedia. She is currently working as a research assistant in OpenUAE Research and Development Group. Her interest in research Internet of Vehicles (IoV) and Artificial Intelligence (AI). Halah is also a member at Sharjah Google Developer Group (GDG) and Arab Women in Computing Association (ArabWIC) since 2016.

FATIMA DAKALBAB is a student pursuing her MSc. in Computer Science and a graduate research assistant at the University of Sharjah in the UAE. Fatima earned her bachelor's degree in information technology Multimedia with a 3.92/4 GPA. She is currently working as a graduate research assistant in OpenUAE Research and Development Group. Her research interest includes inter-blockchain communication, Internet of things (IoT), and Machine learning in anomaly detection and conducting systematic literature reviews. Moreover, Fatima is currently a member of the Sharjah Google Developer Group (GDG) and Arab Women in Computing Association (ArabWIC) since 2016. In addition to being a Events & Workshops Co-Coordinator in the student chapter in UAE for Association for Computing Machinery (ACM).

Competing interests:

The authors declare that they have no competing interests

References

[1] M. T. Khorshed, A. B. M. Shawkat Ali, and S. A. Wasimi, "Trust issues that create threats for cyber attacks in cloud computing," in *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS*, 2011, doi: 10.1109/ICPADS.2011.156.

[2] A. P. Achilleos *et al.*, "The cloud application modelling and execution language," *J. Cloud Comput.*, 2019, doi: 10.1186/s13677-019-0138-7.

[3] P. K. P, S. K. P, and A. P.J.A., "Attribute based encryption in cloud computing: A survey, gap analysis, and future directions," *Journal of Network and Computer Applications*. pp. 37–52, 2018, doi: 10.1016/j.jnca.2018.02.009.

[4] T. Halabi and M. Bellaiche, "Towards quantification and evaluation of security of Cloud Service Providers," *J. Inf. Secur. Appl.*, vol. 33, pp. 55–65, 2017, doi: 10.1016/j.jisa.2017.01.007.

[5] R. Kumar, S. P. Lal, and A. Sharma, "Detecting Denial of Service Attacks in the Cloud," in *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016, 2016 IEEE 2nd International Conference on Big Data*, 2016, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.70.

[6] T. Halabi and M. Bellaiche, "A broker-based framework for standardization and management of Cloud Security-SLAs," *Comput. Secur.*, vol. 75, pp. 59–71, 2018, doi: 10.1016/j.cose.2018.01.019.

[7] C. Mylara Reddy and N. Niranjan, "Fault tolerant software systems using software configurations for cloud computing," *J. Cloud Comput.*, vol. 7, 2018, doi: 10.1186/s13677-018-0104-9.

[8] B. Xu, S. Chen, H. Zhang, and T. Wu, "Incremental k-NN SVM method in intrusion detection," in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 2018, doi: 10.1109/ICSESS.2017.8343013.

[9] R. Moreno-Vozmediano, R. S. Montero, E. Huedo, and I. M. Llorente, "Efficient resource provisioning for elastic Cloud services based on machine learning techniques," *J. Cloud Comput.*, 2019, doi: 10.1186/s13677-019-0128-9.

[10] A. Aleroud and G. Karabatis, "A contextual anomaly detection approach to discover zero-day attacks," in *Proceedings of the 2012 ASE International Conference on Cyber Security, CyberSecurity 2012*, 2012, doi: 10.1109/CyberSecurity.2012.12.

[11] N. Chand, P. Mishra, C. R. Krishna, E. S. Pilli, and M. C. Govil, "A comparative analysis of SVM and its stacking with other classification algorithm for intrusion detection," *Proc. - 2016 Int. Conf. Adv. Comput. Commun. Autom. ICACCA 2016*, 2016, doi: 10.1109/ICACCA.2016.7578859.

[12] J. Arshad, P. Townend, and J. Xu, "A novel intrusion severity analysis approach for Clouds," *Futur. Gener. Comput. Syst.*, vol. 29, no. 1, pp. 416–428, 2013, doi: 10.1016/j.future.2011.08.009.

[13] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of deep learning-based network anomaly detection," *Cluster Comput.*, pp. 1–13, 2017, doi: 10.1007/s10586-017-1117-8.

[14] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[15] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Inf. Fusion*, vol. 42, pp. 146–157, 2018, doi: 10.1016/j.inffus.2017.10.006.

[16] H. Tabrizchi and M. Kuchaki Rafsanjani, "A survey on security challenges in cloud computing: issues, threats, and solutions," *J. Supercomput.*, vol. 76, no. 12, pp. 9493–9532, 2020, doi: 10.1007/s11227-020-03213-1.

[17] L. Alhenaki, A. Alwatban, B. Alamri, and N. Alarifi, "A Survey on the Security of Cloud Computing," in *2nd International Conference on Computer Applications and Information Security, ICCAIS 2019*, 2019, doi: 10.1109/CAIS.2019.8769497.

[18] R. Kumar and R. Goyal, "On cloud security requirements, threats, vulnerabilities and countermeasures: A survey," *Computer Science Review*, vol. 33. pp. 1–48, 2019, doi: 10.1016/j.cosrev.2019.05.002.

[19] L. B. Bhajantri and T. Mujawar, "A Survey of Cloud Computing Security Challenges, Issues and their Countermeasures," in *Proceedings of the 3rd International Conference on I-SMAC IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2019*, 2019, pp. 376–380, doi: 10.1109/I-SMAC47947.2019.9032545.

[20] K. V. Uma and E. S. Blessie, "Survey on android malware detection and protection using data mining algorithms," in *Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2018*, 2019, doi: 10.1109/I-SMAC.2018.8653720.

[21] D. Dave, N. Meruliya, T. D. Gajjar, G. T. Ghoda, D. H. Parekh, and R. Sridaran, "Cloud security issues and challenges," *Adv. Intell. Syst. Comput.*, vol. 654, pp. 499–514, 2018, doi: 10.1007/978-981-10-6620-7_48.

[22] I. Avdagic and K. Hajdarevic, "Survey on machine learning algorithms as cloud service for CIDPS," in *2017 25th Telecommunications Forum, TELFOR 2017 - Proceedings*, 2018, doi: 10.1109/TELFOR.2017.8249467.

[23] G. Nenvani and H. Gupta, "A survey on attack detection on cloud using supervised learning techniques," in *2016 Symposium on Colossal Data Analysis and Networking, CDAN 2016*, 2016, doi: 10.1109/CDAN.2016.7570872.

[24] B. Nelson and T. Olovsson, "Security and privacy for big data: A systematic literature review," in *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 2016, doi: 10.1109/BigData.2016.7841037.

[25] S. G. Kene and D. P. Theng, "A review on intrusion detection techniques for cloud computing and security challenges," in *2nd International Conference on Electronics and Communication Systems, ICECS 2015*, 2015, doi: 10.1109/ECS.2015.7124898.

[26] C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan, "A survey on security issues and solutions at different layers of Cloud computing," *J. Supercomput.*, vol. 63, pp. 561–592, 2013, doi: 10.1007/s11227-012-0831-5.

[27] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in Cloud," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 42–57, 2013, doi: 10.1016/j.jnca.2012.05.003.

[28] A. Patel, M. Taghavi, K. Bakhtiyari, and J. Celestino Júnior, "An intrusion detection and prevention system in cloud computing: A systematic review," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 25–41, 2013, doi: 10.1016/j.jnca.2012.08.007.

[29] M. T. Khorshed, A. B. M. S. Ali, and S. A. Wasimi, "A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing," *Futur. Gener. Comput. Syst.*, vol. 28, no. 6, pp. 833–851, 2012, doi: 10.1016/j.future.2012.01.006.

[30] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3," *Engineering*, vol. 45, no. 4ve, p. 1051, 2007, doi: 10.1145/1134285.1134500.

[31] D. Mittal, D. Kaur, and A. Aggarwal, "Secure data mining in cloud using homomorphic encryption," in *2014 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM 2014*, 2015, doi: 10.1109/CCEM.2014.7015496.

[32] K. Murakami, R. Hanyu, Q. Zhao, and Y. Kaneda, "Improvement of security in cloud systems based on steganography," in *2013 International Joint Conference on Awareness Science and Technology and Ubi-Media Computing: Can We Realize Awareness via Ubi-Media?, iCAST 2013 and UMEDIA 2013*, 2013, doi: 10.1109/ICAwST.2013.6765492.

[33] B. Wang, M. Li, S. S. M. Chow, and H. Li, "Computing encrypted cloud data efficiently under multiple keys," in *2013 IEEE Conference on Communications and Network Security, CNS 2013*, 2013, doi: 10.1109/CNS.2013.6682768.

[34] M. Eskandari, A. S. De Oliveira, and B. Crispo, "VLOC: An approach to verify the physical location of a virtual machine in cloud," in *Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom*, 2015, doi: 10.1109/CloudCom.2014.47.

[35] M. B. Bondada and S. M. S. Bhanu, "Analyzing user behavior using keystroke dynamics to protect cloud from malicious insiders," in *2014 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM 2014*, 2015, doi: 10.1109/CCEM.2014.7015481.

[36] A. Nagaraja, N. Mangathayaru, and N. Rajashekar, "Privacy preserving and data security - A survey," in *Proceedings - 2016 International Conference on Engineering and MIS, ICEMIS 2016*, 2016, doi: 10.1109/ICEMIS.2016.7745350.

[37] E. Hesamifard, H. Takabi, M. Ghasemi, and C. Jones, "Privacy-preserving machine learning in cloud," in *CCSW 2017 - Proceedings of the 2017 Cloud Computing Security Workshop, co-located with CCS 2017*, 2017, pp. 39–43, doi: 10.1145/3140649.3140655.

[38] J. Yuan and S. Yu, "Privacy preserving back-propagation neural network learning made practical with cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 212–221, 2014, doi: 10.1109/TPDS.2013.18.

[39] X. Ma, J. Ma, H. Li, Q. Jiang, and S. Gao, "PDLM: Privacy-Preserving Deep Learning Model on Cloud with Multiple Keys," *IEEE Trans. Serv. Comput.*, pp. 1–1, 2018, doi: 10.1109/TSC.2018.2868750.

[40] S. Wang, X. Jiang, Y. Wu, L. Cui, S. Cheng, and L. Ohno-Machado, "EXpectation Propagation LOgistic REgRession (EXPLORER): Distributed privacy-preserving online model learning," *J. Biomed. Inform.*, vol. 46, no. 3, pp. 480–496, Jun. 2013, doi: 10.1016/j.jbi.2013.03.008.

[41] Z. He, T. Zhang, and R. B. Lee, "Machine Learning Based DDoS Attack Detection from Source Side in Cloud," in *Proceedings - 4th IEEE International Conference on Cyber Security and Cloud Computing, CSCloud 2017 and 3rd IEEE International Conference of Scalable and Smart Cloud, SSC 2017*, 2017, doi: 10.1109/CSCloud.2017.58.

[42] A. Chonka, Y. Xiang, W. Zhou, and A. Bonti, "Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks," *J. Netw. Comput. Appl.*, vol. 34, no. 4, pp. 1097–1107, 2011, doi: 10.1016/j.jnca.2010.06.004.

[43] Z. Chen, F. Jiang, Y. Cheng, X. Gu, W. Liu, and J. Peng, "XGBoost Classifier for DDoS Attack Detection and Analysis in SDN-Based Cloud," in *Proceedings - 2018 IEEE International Conference on Big Data and Smart Computing, BigComp 2018*, 2018, doi: 10.1109/BigComp.2018.00044.

[44] C. Modi, D. Patel, B. Borisanya, A. Patel, and M. Rajarajan, "A novel framework for intrusion detection in cloud," in *Proceedings of the 5th International Conference on Security of Information and Networks, SIN'12*, 2012, pp. 67–74, doi: 10.1145/2388576.2388585.

[45] K. Gai, M. Qiu, and S. A. Elnagdy, "Security-Aware Information Classifications Using Supervised Learning for Cloud-Based Cyber Risk Management in Financial Big Data," in *Proceedings - 2nd IEEE International Conference on Big Data Security on Cloud, IEEE BigDataSecurity 2016, 2nd IEEE International Conference on High Performance and Smart Computing, IEEE HPSC 2016 and IEEE*

International Conference on Intelligent Data and S, 2016, doi: 10.1109/BigDataSecurity-HPSC-IDS.2016.66.

[46] F. Omri, S. Foufou, R. Hamila, and M. Jarraya, "Cloud-based mobile system for biometrics authentication," in *2013 13th International Conference on ITS Telecommunications, ITST 2013*, 2013, pp. 325–330, doi: 10.1109/ITST.2013.6685567.

[47] M. Marwan, A. Kartit, and H. Ouahmane, "Security enhancement in healthcare cloud using machine learning," *Procedia Comput. Sci.*, vol. 127, pp. 388–397, 2018, doi: 10.1016/j.procs.2018.01.136.

[48] M. Wu, Z. Song, and Y. B. Moon, "Detecting cyber-physical attacks in CyberManufacturing systems with machine learning methods," *J. Intell. Manuf.*, vol. 30, pp. 1111–1123, 2019, doi: 10.1007/s10845-017-1315-5.

[49] T. Graepel, K. Lauter, and M. Naehrig, "ML confidential: Machine learning on encrypted data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7839 LNCS, no. July, pp. 1–21, doi: 10.1007/978-3-642-37682-5_1.

[50] K. Vijayakumar and C. Arun, "Continuous security assessment of cloud based applications using distributed hashing algorithm in SDLC," *Cluster Comput.*, vol. 22, pp. 10789–10800, 2019, doi: 10.1007/s10586-017-1176-x.

[51] U. A. Butt *et al.*, "A Review of Machine Learning Algorithms for Cloud Computing Security," *Electronics*, vol. 9, no. 9, p. 1379, Aug. 2020, doi: 10.3390/electronics9091379.

[52] G. P. Zhang, "Neural networks for classification: a survey," *ACS Appl. Mater. Interfaces*, vol. 10, no. 12, pp. 10513–10519, 2018, doi: 10.1021/acsami.7b16653.

[53] N. Pandeeswari and G. Kumar, "Anomaly Detection System in Cloud Environment Using Fuzzy Clustering Based ANN," *Mob. Networks Appl.*, vol. 21, no. 3, pp. 494–505, 2016, doi: 10.1007/s11036-015-0644-x.

[54] T. Salman, D. Bhamare, A. Erbad, R. Jain, and M. Samaka, "Machine Learning for Anomaly Detection and Categorization in Multi-Cloud Environments," in *Proceedings - 4th IEEE International Conference on Cyber Security and Cloud Computing, CSCloud 2017 and 3rd IEEE International Conference of Scalable and Smart Cloud, SSC 2017*, 2017, pp. 97–103, doi: 10.1109/CSCloud.2017.15.

[55] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," *Inf. Sci. (Ny).*, vol. 177, no. 18, pp. 3799–3821, 2007, doi: 10.1016/j.ins.2007.03.025.

[56] C. Y. Chiu, C. T. Yeh, and Y. J. Lee, "Frequent pattern based user behavior anomaly detection for cloud system," in *Proceedings - 2013 Conference on Technologies and Applications of Artificial Intelligence, TAAI 2013*, 2013, doi: 10.1109/TAAI.2013.25.

[57] T. Kim *et al.*, "Monitoring and detecting abnormal behavior in mobile cloud infrastructure," in *Proceedings of the 2012 IEEE Network Operations and Management Symposium, NOMS 2012*, 2012, doi: 10.1109/NOMS.2012.6212067.

[58] M. S. Sarma, Y. Srinivas, M. Abhiram, L. Ullala, M. S. Prasanthi, and J. R. Rao, "Insider threat detection with face recognition and KNN user classification," in *Proceedings - 2017 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM 2017*, 2018, doi: 10.1109/CCEM.2017.16.

[59] F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, "Evaluation of machine learning classifiers for mobile malware detection," *Soft Comput.*, vol. 20, pp. 343–357, 2016, doi: 10.1007/s00500-014-1511-6.

[60] B. Amos, H. Turner, and J. White, "Applying machine learning classifiers to dynamic android malware detection at scale," in *2013 9th International Wireless Communications and Mobile Computing Conference, IWCMC 2013*, 2013, pp. 1666–1671, doi: 10.1109/IWCMC.2013.6583806.

[61] Y. Ye *et al.*, "Combining file content and file relations for cloud based malware detection," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, pp. 222–230, doi: 10.1145/2020408.2020448.

[62] P. Shamsolmoali and M. Zareapoor, "Statistical-based filtering system against DDOS attacks in cloud computing," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, 2014, doi: 10.1109/ICACCI.2014.6968282.

[63] P. Mishra, E. S. Pilli, V. Varadharajant, and U. Tupakula, "NvCloudIDS: A security architecture to detect intrusions at network and virtualization layer in cloud environment," in *2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016*, 2016, doi: 10.1109/ICACCI.2016.7732025.

[64] B. Gulmezoglu, T. Eisenbarth, and B. Sunar, "Cache-based application detection in the cloud using machine learning," in *ASIA CCS 2017 - Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security*, 2017, pp. 288–300, doi: 10.1145/3052973.3053036.

[65] G. Loukas, T. Vuong, R. Heartfield, G. Sakellari, Y. Yoon, and D. Gan, "Cloud-Based Cyber-Physical Intrusion Detection for Vehicles Using Deep Learning," *IEEE Access*, vol. 6, pp. 3491–3508, 2017.

[66] A. Rukavitsyn, K. Borisenko, and A. Shorov, "Self-learning method for DDoS detection model in cloud computing," in *Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering Conference, ElConRus 2017*, 2017, doi: 10.1109/EIConRus.2017.7910612.

[67] Q. Zhang, L. T. Yang, and Z. Chen, "Privacy Preserving Deep Computation Model on Cloud for Big Data Feature Learning," *IEEE Trans. Comput.*, vol. 65, no. 5, pp. 1351–1362, 2016, doi:

10.1109/TC.2015.2470255.

[68] O. A. Kwabena, Z. Qin, Z. Qin, and T. Zhuang, "MSCryptoNet: Multi-Scheme Privacy-Preserving Deep Learning in Cloud Computing," *IEEE Access*, vol. 7, pp. 29344–29354, 2019, doi: 10.1109/ACCESS.2019.2901219.

[69] M. R. Watson, N. U. H. Shirazi, A. K. Marnerides, A. Mauthe, and D. Hutchison, "Malware Detection in Cloud Computing Infrastructures," *IEEE Trans. Dependable Secur. Comput.*, vol. 13, no. 2, pp. 192–205, 2016, doi: 10.1109/TDSC.2015.2457918.

[70] Q. Lu, Y. Xiong, X. Gong, and W. Huang, "Secure collaborative outsourced data mining with multi-owner in cloud computing," in *Proc. of the 11th IEEE Int. Conference on Trust, Security and Privacy in Computing and Communications, TrustCom-2012 - 11th IEEE Int. Conference on Ubiquitous Computing and Communications, IUCC-2012*, 2012, doi: 10.1109/TrustCom.2012.251.

[71] H. Zhao, M. Xu, N. Zheng, J. Yao, and Q. Hou, "Malicious executables classification based on behavioral factor analysis," in *IC4E 2010 - 2010 International Conference on e-Education, e-Business, e-Management and e-Learning*, 2010, doi: 10.1109/IC4E.2010.78.

[72] P. Wang and Y. S. Wang, "Malware behavioural detection and vaccine development by using a support vector model classifier," *J. Comput. Syst. Sci.*, vol. 81, no. 6, pp. 1012–1026, 2015, doi: 10.1016/j.jcss.2014.12.014.

[73] N. Sengupta, "Designing encryption and IDS for cloud security," in *ACM International Conference Proceeding Series*, 2017, pp. 1–5, doi: 10.1145/3018896.3018954.

[74] Z. Li, W. Sun, and L. Wang, "A neural network based distributed intrusion detection system on cloud platform," in *Proceedings - 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems, IEEE CCIS 2012*, 2013, doi: 10.1109/CCIS.2012.6664371.

[75] Y. Zhu, Z. Huang, and T. Takagi, "Secure and controllable k-NN query over encrypted cloud data with key confidentiality," *J. Parallel Distrib. Comput.*, vol. 89, pp. 1–12, 2016, doi: 10.1016/j.jpdc.2015.11.004.

[76] A. Sahi, D. Lai, Y. Li, and M. Diykh, "An Efficient DDoS TCP Flood Attack Detection and Prevention System in a Cloud Environment," *IEEE Access*, vol. 5, pp. 6036–6048, 2017, doi: 10.1109/ACCESS.2017.2688460.

[77] P. Mishra, E. S. Pilli, V. Varadharajan, and U. Tupakula, "Efficient approaches for intrusion detection in cloud environment," in *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, 2017, doi: 10.1109/CCAA.2016.7813926.

[78] X. Li, Y. Zhu, and J. Wang, "Secure naïve bayesian classification over encrypted data in cloud," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, pp. 130–150, doi: 10.1007/978-3-319-47422-9_8.

[79] K. Borisenko, A. Smirnov, E. Novikova, and A. Shorov, "DDoS attacks detection in cloud computing using data mining techniques," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9728, pp. 197–211, doi: 10.1007/978-3-319-41561-1_15.

[80] P. Xiao, W. Qu, H. Qi, and Z. Li, "Detecting DDoS attacks against data center with correlation analysis," *Comput. Commun.*, vol. 67, pp. 66–74, 2015, doi: 10.1016/j.comcom.2015.06.012.

[81] A. Saied, R. E. Overill, and T. Radzik, "Detection of known and unknown DDoS attacks using Artificial Neural Networks," *Neurocomputing*, vol. 172, pp. 385–393, 2016, doi: 10.1016/j.neucom.2015.04.101.

[82] Z. Chkirbene, A. Erbad, and R. Hamila, "A Combined Decision for Secure Cloud Computing Based on Machine Learning and Past Information," in *IEEE Wireless Communications and Networking Conference, WCNC*, 2019, doi: 10.1109/WCNC.2019.8885566.

[83] V. Sharma, V. Verma, and A. Sharma, "Detection of DDoS Attacks Using Machine Learning in Cloud Computing," in *International Conference on Advanced Informatics for Computing Research*, 2019, vol. 1076, pp. 260–273, doi: 10.1007/978-981-15-0111-1_24.

[84] H. Abbasi, N. Ezzati-Jivan, M. Bellaiche, C. Talhi, and M. R. Dagenais, "Machine Learning-Based EDoS Attack Detection Technique Using Execution Trace Analysis," *J. Hardw. Syst. Secur.*, vol. 3, pp. 164–176, 2019, doi: 10.1007/s41635-018-0061-2.

[85] H. Kim, J. Kim, Y. Kim, I. Kim, and K. J. Kim, "Design of network threat detection and classification based on machine learning on cloud computing," *Cluster Comput.*, vol. 22, pp. 2341–2350, 2019, doi: 10.1007/s10586-018-1841-8.

[86] P. Mishra, K. Khurana, S. Gupta, and M. K. Sharma, "VMAnalyzer: Malware Semantic Analysis using Integrated CNN and Bi-Directional LSTM for Detecting VM-level Attacks in Cloud," in *2019 12th International Conference on Contemporary Computing, IC3 2019*, 2019, doi: 10.1109/IC3.2019.8844877.

[87] M. T. Sandıkkaya, Y. Yaslan, and C. D. Özdemir, "DeMETER in clouds: detection of malicious external thread execution in runtime with machine learning in PaaS clouds," *Cluster Comput.*, vol. 23, pp. 2565–2578, 2020, doi: 10.1007/s10586-019-03027-8.

[88] A. Abusitta, M. Bellaiche, M. Dagenais, and T. Halabi, "A deep learning approach for proactive multi-cloud cooperative intrusion detection system," *Futur. Gener. Comput. Syst.*, vol. 98, pp. 308–318, 2019, doi: 10.1016/j.future.2019.03.043.

[89] M. A. Zardari, L. T. Jung, and N. Zakaria, "K-NN classifier for data confidentiality in cloud computing,"

in *2014 International Conference on Computer and Information Sciences, ICCOINS 2014 - A Conference of World Engineering, Science and Technology Congress, ESTCON 2014 - Proceedings*, 2014, doi: 10.1109/ICCOINS.2014.6868432.

[90] A. Inani, C. Verma, and S. Jain, "A machine learning algorithm TSF k-Nn based on automated data classification for securing mobile cloud computing model," in *2019 IEEE 4th International Conference on Computer and Communication Systems, ICCCS 2019*, 2019, pp. 9–13, doi: 10.1109/CCOMS.2019.8821756.

[91] T. Ayodele and D. Adeegbe, "Cloud based emails boundaries and vulnerabilities," in *Proceedings of 2013 Science and Information Conference, SAI 2013*, 2013, pp. 912–914.