A Systematic Review on Anomaly Detection for Cloud Computing Environments

Tanja Hagemann
tanja.hagemann@tu-berlin.de
Technische Universität Berlin,
Telekom Innovation Laboratories
Berlin, Germany

Katerina Katsarou
a.katsarou@tu-berlin.de
Technische Universität Berlin,
Service-centric Networking
Berlin, Germany

ABSTRACT

The detection of anomalies in data is a far-reaching field of research which also applies to the field of cloud computing in several different ways: from the detection of various types of intrusions to the detection of hardware failures, many publications address how far anomaly detection methods are able to meet the specific requirements of a cloud-based network. Since there is still no comprehensive overview of this constantly growing field of research, this literature review provides a systematic evaluation of 215 publications that can be considered as representative for the last ten years of this scientific development. Our analysis identifies three main methodological areas (machine learning, deep learning, statistical approaches) and summarizes how exactly the corresponding models are applied for the detection of anomalies. Furthermore, we clarify which concrete application areas are typically addressed by anomaly detection in the context of cloud computing environments and which related public datasets are often used for evaluations. Finally, we discuss the implications of the literature review and provide directions for future research.

CCS CONCEPTS

• **Computing methodologies** → **Anomaly detection**; • **Networks** → **Cloud Computing**; **Network monitoring**; **Network performance analysis**; • **Security and privacy** → **Intrusion detection systems**.

KEYWORDS

Anomaly Detection, Intrusion Detection, Cloud Computing, Machine learning, AIOps.

ACM Reference Format:
Tanja Hagemann and Katerina Katsarou. 2020. A Systematic Review on Anomaly Detection for Cloud Computing Environments. In *2020 3rd Artificial Intelligence and Cloud Computing Conference (AICCC 2020), December 18–20, 2020, Kyoto, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442536.3442550

The detection of anomalies in data has a long tradition and versatile applications. The most comprehensive overview of the topic is offered by Chandola et al. [33]. By definition [75], an anomaly (sometimes also referred to as an outlier) is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism. It can also be defined as an outlying observation that appears to deviate markedly from other members of the sample in which it occurs [23]. In each case, what defines an anomaly as such depends on the sample and the measurement method. In general, three different types of anomalies are distinguished: point anomalies, collective anomalies and contextual anomalies. Point anomalies are individual data instances that appear abnormal when compared to the rest of the dataset. If a collection of related data instances is anomalous with respect to the rest of the dataset, it is termed as a collective anomaly. The individual data instances in a collective anomaly may not be anomalies by themselves, but their occurrence together as a collection is anomalous. The third type of anomalies does take contextual attributes (such as time) into consideration. If a data instance is anomalous in a specific context (but not otherwise), then it is termed as a contextual anomaly (see figure 1). What may sound simple in theory is often a challenge in practice, as it is often not clear which phenomena manifest themselves in the data and how, and which anomalies are really relevant for a particular application. But the selected data and the choice of model is decisive in the end, because too many false positives (detected anomalies that are not caused by a relevant event) make a reliable application impossible. Thus, the goal of anomaly detection is to apply methods which are capable of identifying relevant anomalies in data without detecting too many false positives. In general these approaches can be distinguished if they are based on labeled data (binary: anomaly/normal, multiclass: attack types, failures, ...) or if they are independent of such a high degree of prior knowledge and do thus only take similarities and dependencies in the data into consideration for identifying the normal data structures and detecting deviations from it.

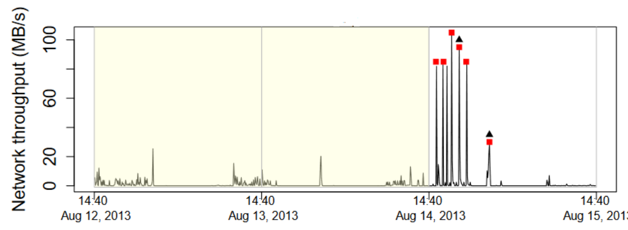

Figure 1: Typical example of contextual anomalies in a cloud workload behaviour observed from VM traces. The coloured shapes represent anomalies identified by a detection system called RADS [22].

AICCC 2020, December 18–20, 2020, Kyoto, Japan

ACM ISBN 978-1-4503-8883-2/20/12.
https://doi.org/10.1145/3442536.3442550

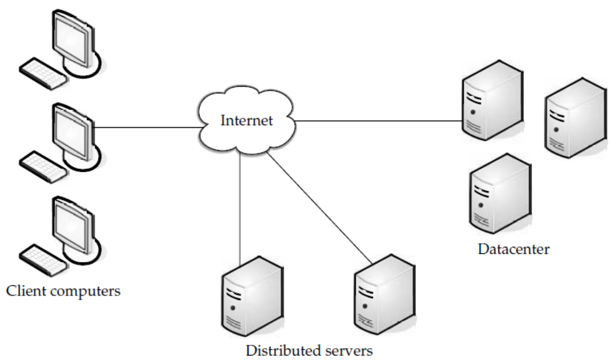

Figure 2: Three components make up a cloud computing solution [191].

A good overview of how far anomaly detection is helpful for traditional computer networks is provided by Bhattacharyya and Kalita [28]. In general, a computer network is formed when many individual entities are put together to contribute toward providing complex communication services. Things can go wrong with any of the interacting entities. Then, anomalies in a network may occur due to two major reasons: performance related and security related. A performance-related anomaly may occur due to malfunctions, such as network device malfunction, e.g., router misconfiguration. Security-related anomalies occur due to malicious activities that attempt to disrupt normal functioning of the network. They can be classified into six major categories: infection, exploding, probe, cheat, traverse and concurrency.

The US National Institute of Standards and Technology (NIST) defines cloud computing as follows [111]: Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three delivery models, and four deployment models. Thus, the general idea behind cloud computing is to provide a new model of infrastructure provisioning on which business can create elastic on-demand IT infrastructures according to their ever changing requirements. These on-demand infrastructures may enable end users to use the business services without installation and access them at any computer with internet access. To achieve this end, the cloud computing defines a stack composed of three well-known layers [100]: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (Saas). Due to virtualization, software can be installed allowing multiple instances of virtual servers (VMs) to be used. In this way, several VMs can run on one physical server. These physical servers do not all have to be housed in the same location. Often, servers are in geographically disparate locations to give the service provider more flexibility in options and security. All components are visualized in figure 2. However, cloud computing environments are multi-domain environments in which each domain can use different security, privacy, and trust requirements and potentially employ various mechanisms, interfaces, and semantics [182].

To conclude, the detection of anomalies in cloud-based infrastructures can benefit from traditional IT network approaches but additionally has its own challenges:

- Heterogeneity of services: Since many different services are run at the same time, there can also appear unpredictable side-effects which manifest themselves in new, previously unseen data patterns which in turn makes it difficult to apply common monitoring techniques. In addition, When failures occur or when certain services generate abnormal loads the general service delivery can be interrupted.
- Multi-tenancy: Several users share the same hardware resources which poses particular challenges with regard to the optimization and planning of resource utilization.
- Virtualization: Since the environment is abstracted, it is difficult to completely diagnose problems and monitor performance. In particular, problems which take place on the physical level might not be detectable on a virtual one and vice versa.
- Dynamics: the on-demand character of the cloud infrastructure ensures that services are quickly scaled up or down, making monitoring and troubleshooting even more difficult.

To tackle these challenges, tremendous efforts have been made, resulting in a rich literature of related papers and methods. The adopted methods and detecting strategies also vary greatly, ranging from supervised to unsupervised machine learning and from statistical to hybrid approaches. However, to the best of our knowledge, little effort has been made to systematically summarize the differences and connections between these approaches. In this paper, we try to fill this knowledge gap by comprehensively reviewing anomaly detection approaches for cloud computing environments.

1 RELATED WORK

During our systematic research we have also identified a certain amount of related surveys that in one way or another deal with cloud infrastructures and anomaly detection. These are summarized in the following.

Most surveys focus exclusively on the state-of-the-art for the identification of intrusions in cloud environments. Furthermore, these surveys are often limited to only one methodological field. For example, Alarqan and Zaaba [12] present an overview about how to detect and defend distributed denial of service (DDoS) attacks in cloud computing environments while focusing on statistical anomaly-based methods only and Prathyusha and Naseera [143] conduct an overview about monitoring solutions for detecting DDoS attacks in cloud environments with a focus on biologically inspired algorithms. We have summarized all surveys concerned with intrusion detection systems (IDS) for cloud environments in Table 1.

A survey that takes a more generic approach on autonomic software systems is from Dehraj and Sharma [47]. It provides an insight vision of the autonomic decision-making concept and its importance for the various purposes such as intrusion detection, cloud-based data security, wireless sensor network, Internet of Things, big data and other areas where management cannot be handled by a human in real time. Another general approach is from Fernandes et al. [55] which provides an article about artificial immune systems.

Table 1: Related Work Dealing with Intrusion Detection.

Scope	References
IDS for cloud computing in general	[122], [148], [117], [119], [20], [140], [9], [98]
DDoS / DoS detection for cloud computing	[12], [143], [176], [10], [154], [134], [19]

The article introduces the related principles and surveys several works applying such systems to computer security problems. Both approaches either do not focus on the specific methodological aspects of anomaly detection or do not take into account the special requirements of cloud computing environments. Other examples covering other than cloud-based networks and the detection of anomalies are [7, 29, 147, 184].

The only other survey dealing explicitly with anomaly detection for cloud computing infrastructures in general is by Ramachandra et al. [149]. This survey provides insights into how log data is useful in order to detect anomalies in cloud computing. Since this survey deals exclusively with anomalies in log data and its examples are also limited to intrusion detection again, we do not find a generic view of the topic here either.

To conclude, none of the surveys we have identified deals with the entirety of methods used so far for detecting different types of anomalies in cloud computing infrastructures and/or does take more than one related area of application into consideration. The goal of this survey is to fill this gap.

2 RESEARCH METHODOLOGY

Our contribution is a survey that systematically reviews the state-of-the-art for anomaly detection in the domain of cloud computing environments. For this purpose, we query three different scientific databases from which we assume that they represent a sufficiently good overall picture of the relevant research, namely SpringerLink[1], Web of Science[2] (WoS) and the open-access archive ArXiv[3]. Hereby, we address the following three research questions:

(1) Which methods are used to detect anomalies in cloud environments?
(2) What are specific purposes for using anomaly detection for cloud environments?
(3) How has this research area evolved over time?

To complement our manual query pipeline, we furthermore use text mining approaches to obtain a topical clustering of the resulting papers.

2.1 Search and Selection

The goal of our literature review is to collect and evaluate all those publications that use methods from the area of anomaly detection for identifying relevant patterns in a cloud-based infrastructure. As a first step, we use therefore two slightly different search strings bringing together the two domains of interest (Table 2): the first one connects "Anomaly Detection" with "Cloud Computing". This allows us to identify all those publications that in any case show aspects of both domains. As we already know that many of the relevant approaches use anomaly detection for monitoring the infrastructure, we would like to make sure that we capture related publications. Thus, as a second search string, we combine the two terms "Anomaly Detection" and "Cloud Monitoring". We query each database with both search strings and refine all responses by the following filter criteria:

- Since we are only interested in single, clearly defined approaches and their evaluation, we do not consider entire book chapters or survey papers. However, we select all survey papers to assess them in the context of the related work.
- As we are not interested in papers that implement anomaly detection methods by means of cloud computing technologies in a field such as geophysics or biology, we only consider those works that belong to a relevant field such as computer science or engineering.
- Furthermore, anomaly detection should not only be mentioned, but rather be the focus of the work. Therefore we only select those papers that contain the search strings in the abstract or title.
- In particular WoS contains work from several publishers, also from Springer. Since there are some overlaps in the content of the databases, we eliminate duplicates based on titles or digital object identifiers (DOI).
- Finally, we manually go through the resulting papers and discard those that still do not fit our review goal.

We receive a total of 215 relevant publications, which we analyse in the following and evaluate with respect to the specific problems addressed and the corresponding methods used to solve it.

2.2 Analysis and Synthesis

A first statistical evaluation of the identified papers shows that our search found about the same proportion (~ 45%) of results in SpringerLink and WoS (without SpringerLink). ArXiv contains a much smaller number of relevant publications (9.9%), but is also the smallest database. Figure 5 shows a further detailed distribution of the results from WoS with regard to the areas of application. Furthermore, a chronological analysis shows that the oldest relevant contribution dates back to 2011 and the number of total publications shows a positive trend up to and including 2019. This trend seems to continue until our analysis (May 2020) and we can therefore expect a new peak value for 2020 as well. The distribution of the papers over time is shown in figure 4. This growth can probably be explained by the increased use of cloud-based infrastructures as well as the application of new methods based on the advancement of machine learning, as both use cases and the variety of possible solutions increase continuously.

[1] https://link.springer.com/
[2] https://clarivate.com/webofsciencegroup/solutions/web-of-science/
[3] https://arxiv.org/

Table 2: Search Strings Used for the Literature Review.

("Cloud Computing" OR "Cloud Monitoring") AND "Anomaly Detection"

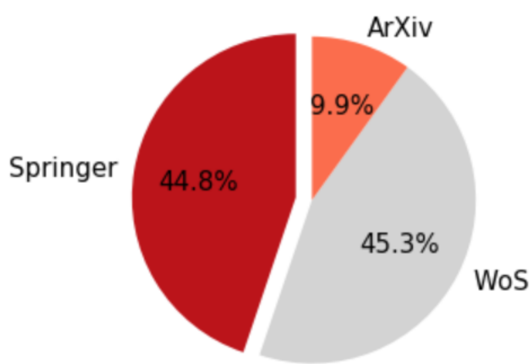

Figure 3: Total amount of reviewed papers distributed over databases.

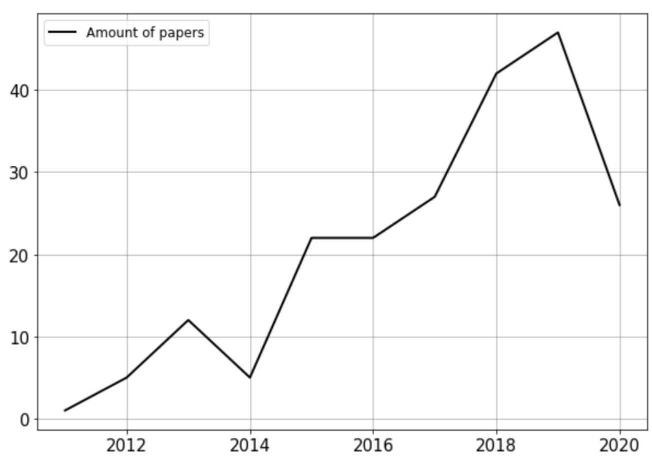

Figure 4: Number of published papers based on all reviewed papers. Note that the results for 2020 are only until May.

Table 3: Summary of Classical Machine Learning Methods Used in the Reviewed Literature.

Method	References
Support Vector Machines	[196], [193], [87], [84], [131], [58], [97], [17], [155], [3], [203], [74], [127], [208], [186], [171], [50], [144], [172], [194], [142]
Random forest	[131], [85], [141], [136], [115], [192], [107], [126], [156], [93], [127], [208], [186], [157], [159], [171], [50], [144]
Decision trees	[116], [26], [164], [118], [56], [133], [208], [186], [214], [144]
iForest	[120], [150], [169], [30]
Fuzzy clustering	[139], [65], [61], [87], [175], [185]
k-means	[158], [17], [1], [21], [93], [187], [4], [216]
DBSCAN	[212], [112], [66], [129], [93], [59], [216]
k-nearest neighbours	[115], [64], [158], [165], [74], [127], [208], [186], [25], [157], [159], [25], [50], [144], [187]
Local outlier factor	[81], [82], [102], [13], [211], [121]
FP-outlier detection	[39]
Invariant mining	[215]
MCOD algorithm	[174]
One-class classification	[22]

3 RESULTS

In this section, we first summarize all anomaly detection methods (3.1) found in the reviewed papers. Afterwards we give an overview in which application areas these methods are mainly used.

3.1 Anomaly Detection Methods

With regard to the methods used in the literature, we have discovered three main clusters: classical machine learning (28.3%), deep learning (19.7%) and statistical approaches (23%). An evaluation of the papers with regard to the methods used (figure 6) also shows that these areas sometimes overlap, as several methods from different areas are somehow combined. This is mostly the case because either different methods are evaluated against each other or a pipeline of methods is used, where different methods build on each other. If several methods build up on each other, we have in principle always assigned the method that forms the core model regarding the detection of anomalies. In other words, if we refer to a method it is always used as the one that learns the difference between a normal and an abnormal data instance, even if this decision is strongly depending on pre-processing techniques (aggregation and feature selection) and/or further processing steps. In addition, almost one third (27.9%) of the papers cannot be assigned to one specific methodical field, since the chosen approaches are often rule-based algorithms designed specifically for one system architecture and are therefore difficult to transfer to other cases. Thus, we will concentrate on the three methodical fields in the following and briefly summarize other interesting methods in the end.

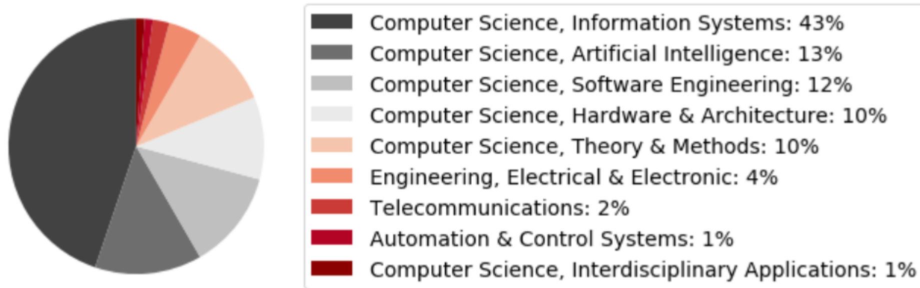

Figure 5: Domain distribution for all reviewed papers which are obtained from WoS.

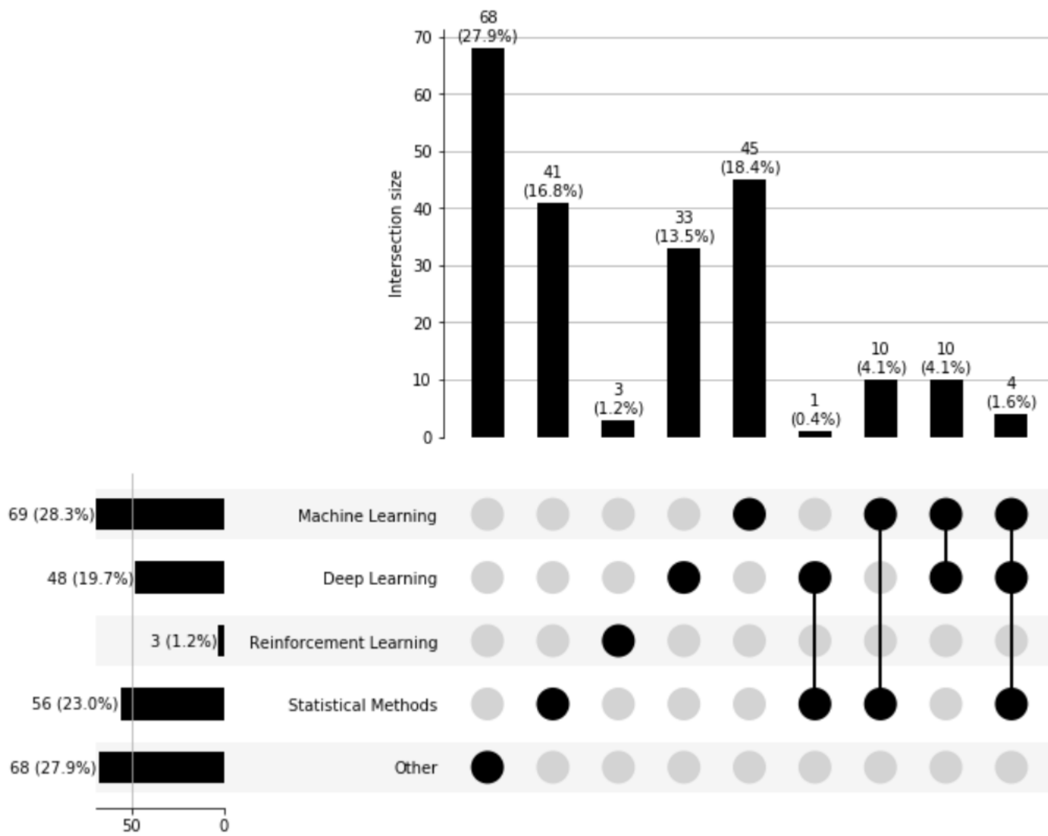

Figure 6: An upset plot showing the distribution of the identified methodological fields (left) over the 215 reviewed papers (horizontal bars). The first five vertical bars on the top show the amount of papers mainly using methods from one field indicated by the black dot. The last four vertical bars show the amount of papers dealing with several methods from different fields.

3.1.1 Classical Machine Learning. Machine learning as a whole includes all methods that are not based on deterministic rules, but rather learn the properties and relationships of data and derive rules from them. As classical machine learning methods we refer in the following to all those methods of machine learning that are not based on neural networks and thus usually have a much smaller number of parameters to be learned. Furthermore, these methods can be divided into unsupervised and supervised methods. Unsupervised learning means that the models are trained without labels and therefore mainly reveal patterns in the data, but focus less on predictions. Representatives of this class often belong to clustering methods. Supervised learning, on the other hand, uses the labels during the training process and has a prediction as its goal, for

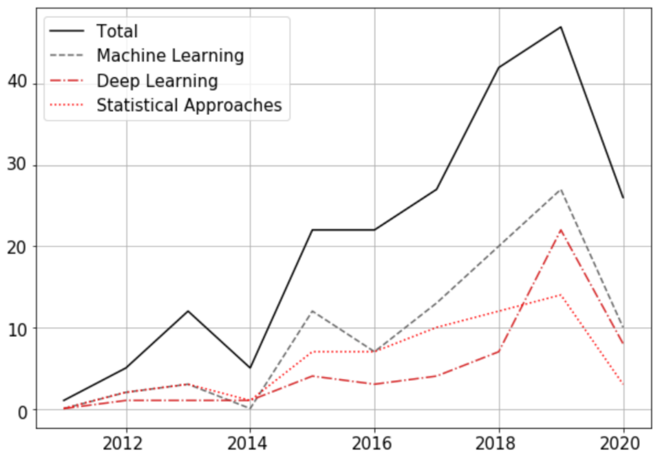

Figure 7: The plot shows the evolution of the three main methodological fields over time. Since 2014, there has been a strong increase in machine learning and deep learning methods, the latter having increased significantly since 2018. The statistical approaches, on the other hand, have tended to flatten out in terms of growth.

which reason the regression and classification methods are particularly included. In the domain of clustering, it becomes apparent that either hard or soft (fuzzy) methods are used. Hard methods allow each data point to belong to only one cluster, whereas fuzzy methods are mostly variations of common clustering algorithms, but allow data points to belong to more than one cluster. For anomaly detection, all clustering algorithms identify those points as anomalies that do not belong to a cluster because they are too far away from other data points. In the reviewed papers, there are mainly three widely used supervised methods that could barely be more different: Support Vector Machines, k-nearest neighbours and tree-based methods such as decision trees and random forests.

All of the reviewed papers that are based on classical machine learning approaches are summarized in table 3.

3.1.2 Deep Learning Methods. Deep learning belongs to the family of machine learning, but unlike classical methods is based on neural networks which use several (≥2) hidden layers. A major source of difficulty in many real-world applications is that it can be very difficult to extract appropriate high-level, abstract features from raw data. Deep learning solves this central problem by introducing representations that are expressed in terms of other, simpler representations. Another perspective on deep learning is that depth enables the computer to learn a multi-step computer program [67]. Each layer of the representation can be thought of as the state of the computer's memory after executing another set of instructions in parallel. Thus, Networks with greater depth can execute more instructions in sequence. In the reviewed papers wie mainly identified three unsupervised methods of deep learning: Autoencoders, recurrent neural networks (RNNs) and self-organizing maps (SOMs). An autoencoder is a combination of an encoder function, which converts the input data into a different representation, and a decoder function, which converts the new representation back into the original format. This architecture can be applied very straightforward for the detection of anomalies: since anomalies are rarely, if ever, included in the training data, the model mainly learns a representation of the normal data behaviour. In the test phase anomalies can be detected by the reconstruction error. Another unsupervised method of deep learning that frequently appears in the reviewed papers RNNs [153]. RNNs are a family of neural networks for processing sequential data and therefore particularly useful for time series data such as performance metrics. However, there are gradient vanishing or exploding problems to RNNs. Considering the weakness of RNNs, long short-term memory (LSTM) was proposed to handle gradient vanishing problem [77]. Most RNN-based models are used for anomaly detection by learning the sequential character of the input data and predicting a probability distribution over the next upcoming values. If, according to the learned probability distribution, the actual next value is an unlikely event, it is labeled as an anomaly. Furthermore, it is also feasible to combine the concepts of autoencoders with those of RNNs in kind of an Encoder-Decoder LSTM architecture [181]. A self-organizing map (SOM) [94] is a special kind of neural network which is able to reduce data dimensions and highlight similarities among data without imposing excessive learning overhead. The central property of the SOM is that it forms a non-linear projection of a high-dimensional data manifold on a regular, low-dimensional (usually two-dimensional) grid. In the display, the clustering of the data space as well as the metric-topological relations of the data items are clearly visible which in turn allows the detection of anomalies.

The most popular class of supervised feed-forward neural networks are multi-layer perceptrons (MLPs). An MLP can be viewed as a logistic regression classifier. Convolutional networks [99], also known as convolutional neural networks, or CNNs, are a specialized kind of supervised neural network for processing data that has a known grid-like topology. Examples include time series data, which can be thought of as a one-dimensional grid taking samples at regular time intervals, and image data, which can be thought of as a two-dimensional grid of pixels. CNNs are most often used as classifiers where the added convolutional layers reduce the feature space and the output is afterwards classified by standard fully connected layers. Extreme Learning Machine (ELM) [80] represents a suit of deep learning techniques (including single-hidden-layer as well as multi-hidden-layer feed-forward networks) in which hidden neurons do not need to be tuned during training. Instead, these hidden nodes are randomly assigned and never updated or can be inherited from their ancestors without being changed. For anomaly detection, it can be used as a regression model as well as a classifier.

All of the reviewed papers that are based on deep learning learning approaches are summarized in table 4.

3.1.3 Statistical Approaches. The field of probability-based methods is by far the most diverse one. Besides other, we have found regression analysis, probabilistic graphical models and above all methods for time series analysis.

Time series analysis often arises when processes are monitored or metrics are tracked over time. It accounts for the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation) that should be accounted

Table 4: Summary of Deep Learning Methods Used in the Reviewed Literature.

Method	References
Multi-layer perceptron	[16], [26], [56], [64], [139], [185], [48], [46], [15], [92], [38], [36], [88], [37], [109], [90], [104], [208], [25], [159], [25], [214], [144]
Autoencoder	[206], [32], [86], [73]
Recurrent neural network	[21], [72], [104], [86]
Convolutional neural network	[180], [62], [2], [104]
Self-organizing map	[179], [27], [177], [187]
Adaptive neuro-fuzzy inference system	[60], [124], [138]
Extreme learning machine	[209], [96], [74]
Probabilistic neural network	[145]
Graph neural network	[53]
Radial basis function network	[190]
Generative adversarial network	[31]
Type-2 fuzzy neural network	[146]
Synergetic neural network	[197]

for. Similar to the LSTM-based approaches, the values that actually occur are then compared to the prediction which allows to decide how rare they are. An anomaly is then a pattern that rarely occurred in the past and that arrives unexpectedly. All Bayesian approaches are based on Bayes rule in one way or another. For example, Bayesian classifiers assign the most likely class to a given example described by its feature vector. Bayesian networks also known as belief networks (or Bayes nets for short), belong to the family of probabilistic graphical models. Anomaly detection based on Bayesian networks can be performed by learning a joint probability distribution and deriving a anomalous states. Regression analysis is a well-known statistical learning technique useful to estimates the relationships between two set of values: the predicted values (independent) and the actual values (dependent). It focuses on the relationship between those two sets of values and helps in understanding how the typical value of the dependent variable changes when any one of the independent variables varies. Since it is based on predicting upcoming values it can be used for anomaly detection by measuring the difference between observed and predicted values.

All found methods from this domain are summarized in table 5.

3.1.4 Other. In addition to the approaches mentioned so far, there are other methods that do not fit into one of the three main categories, but are certainly promising. Since we came across them in our review, we do not want to omit them and summarize them in table 6.

3.2 Application Areas

For a first approximation of the application areas addressed in the reviewed papers, we use a pipeline of text mining approaches: after applying standard pre-processing techniques, we first use term frequency–inverse document frequency (tf-idf) to transfer

Table 5: Summary of Statistical Approaches Used in the Reviewed Literature.

Method	References
Time series analysis	[162], [161], [76], [198], [83], [71], [210], [70]
Bayesian learning	[118], [85], [14], [103], [159], [50], [35], [128], [105], [157]
Principal component analysis	[195], [5], [18], [113], [108]
Regression analysis	[205], [54], [25], [89], [156]
Logistic regression	[131], [115], [26], [207]
Hidden Markov model	[79], [78], [11], [170]
Markov chain	[167], [52], [41]
Gaussian mixture models	[106], [13]
Statistical tests	[173], [114]
Restricted Boltzmann machine	[110], [61]
Dempster-Shafer theory	[65], [125]
KL / Jensen-Shannon divergence	[24], [113]
Linear discriminant analysis	[26]
Matrix sketching	[34]
Maximum likelihood estimation	[178]
Good-Turing smoothing	[178]

Table 6: Selection of Other Approaches Used in the Reviewed Literature.

Method	References
Reinforcement learning	[201], [200], [63],
Wavelet transform	[135], [137], [69]
Game theory	[45], [91]
Artificial bee colony algorithm	[164], [61]
Page rank algorithm	[202]
Particle swarm optimization	[163]
Catastrophe theory	[197]
Phase space analysis	[44]
Chaos theory	[51]

each paper's abstract into a vector representation. Then we apply spectral clustering to identify groups. The high-dimensional result is finally visualized by PCA to get an overview of which titles might be semantically similar to each other. The result is shown by figure 9. This macroscopic view already confirms the insight gained from the related work: Intrusion detection and especially DDoS attacks are a topic of high visibility and are addressed by at least half of all reviewed papers.

Our subsequent manual evaluation shows that intrusion detection is indeed the main focus of approximately 62% of the reviewed papers. In addition, the two topics of performance monitoring and failure detection become visible. An estimated topical distribution is shown in figure 8. In this section we briefly summarize each of these areas.

3.2.1 Intrusion Detection. From a classical point of view, an IDS can be implemented by using signature-based rules. The main assumption for applying anomaly detection here is the fact that signatures of intrusions can change over time, but intrusions might be still

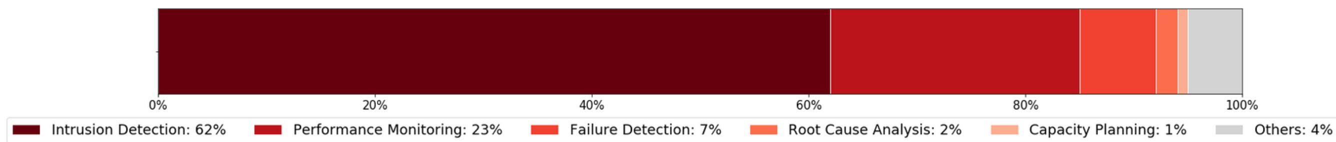

Figure 8: Application area distribution for all reviewed papers.

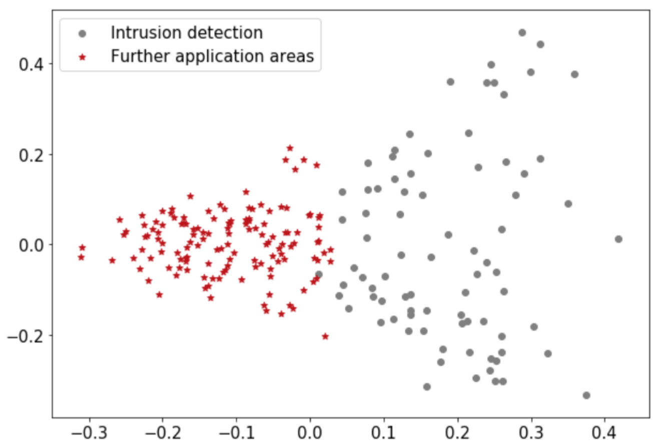

Figure 9: Macroscopic view on application areas identified by tf-idf and spectral clustering. PCA was used for a two-dimensional visualisation. The gray dots represent papers, which, according to their abstract, focus more on intrusion detection (the further to the right the more explicit). The papers represented by the red stars have rather a different focus (the further to the left the more likely).

detectable as anomalies compared to the normal network traffic. IDS are commonly used in traditional enterprise systems, but suffer from a numerous issues in the cloud environment. One issue is the separation of responsibility between the provider and user and the practicality of who and how the IDS should be administered [152]. In the reviewed papers we mainly found the following types of attacks:

- DDoS / DoS: The type of attack that is mostly considered by the related work is the (distributed) Denial of Service attack (DoS). Distributed Denial of Service (DDoS) attacks against cloud providers are a serious threat due to the high impact of availability disruptions, with consequences such as loss of business, loss of reputation and possible ransom demands by the attackers [40]. Douligeris and Mitrokotsa [49] classify DDoS attacks based on the degree of automation, the vulnerability that was exploited, the attack rate dynamics, and whether the impact is disruptive or degrading. The vulnerabilities enumerated are UDP/ICMP flooding attacks, Smurf and Fraggle amplification attacks, protocol exploit attacks and malformed packet attacks.
- Botnets: Botnets are an army of compromised machines that are often under the control and coordination of a single source of (direct/indirect) influence via a remote secure channel. They are generally able to propagate themselves on a network and infect vulnerable machines. Typically they rely either on maintaining contact with the bot master or owner of the botnet for command and control, or on certain modules within the bot code architecture that perform the same function. Over time, however, bot codes could now be engineered to be able to recruit other vulnerable systems as bots into the botnet, report the status of the operations of individual bots in the botnet, and protect the botnet and its member bots from infiltration [132].
- Malware: Malware attacks such as virus, worms and rootkits are threats to VMs in cloud environments. Given the scale of data centers, continuous security monitoring of the virtual assets is essential to detect unexpected (and potentially malicious) behaviour.
- Fraud storms: Cloud computing resources are sometimes hijacked for fraudulent use. A serious fraudulent type is that of fraud storms, which are events of large-scale fraudulent use. These events begin when fraudulent users discover new vulnerabilities in the sign up process and then they exploit in mass. The ability to perform early detection of these storms is a critical component of any cloud-based public computing system.

3.2.2 *Performance Monitoring.* A performance anomaly refers to any sudden degradation of performance that deviates performances which typically results in a decrease in the system efficiency [204]. These changes should be detected by appropriate monitoring techniques. In cloud computing systems, it is not enough to detect outages or other functional anomalies, because those anomalies often cause service interruption and can be resolved by simply restarting or replacing hardware. On the other hand, performance anomalies caused by resource sharing and interference are more worthy of attention to ensure constantly-provided services.

3.2.3 *Failure Detection.* A failure in a cloud environment is said to take place when the provided services by the system do not satisfy the constraints of the customer. Basically, there are three different types of failures in cloud environments: VM failures, software failures and hardware failures. For preventing failures, it is important to accurately predict or detect them and then adopt a suitable strategy to fix them. Failure detection is a method that is used to identify the exact location of an already present failure in the system before it can cause any major damage.

3.2.4 *Root Cause Analysis.* A possible next step after detecting anomalies is to look for their causes so that underlying problems can be resolved and in turn the anomaly detection method improved. This process is often referred to as root cause analysis, and along with anomaly detection, the other important aspect of automated maintenance of cloud environments.

3.3 Benchmark Datasets

Knowing which benchmark datasets are common in a given domain is important to understand the limitations of approaches based on these datasets. For example, most public datasets are limited to certain patterns which are caused by well-defined processes and are furthermore based on specific assumptions that only apply to certain applications. Nevertheless, it is extremely important that scientific datasets exist that are publicly accessible, so that solutions are comparable. During our review we came across the same datasets again and again which we would like to list here (see table 7) in order to enable a discussion in the future to what extent these data already cover certain situations and where they should be supplemented to cover further important aspects of anomaly detection in cloud environments.

4 FUTURE DIRECTIONS

On the methodological side there are mainly approaches of machine learning or deep neural networks as well as statistical techniques that are used. It is worth mentioning that supervised methods still account for at least half of all approaches, which tends to be problematic, as the presence of labeled data is required for a transfer to real-world applications. From what we found we suggest the following directions for future research:

- Reinforcement learning: Besides supervised and unsupervised learning, reinforcement learning is the problem faced by an agent that must learn behaviour through trial-and-error interactions with a dynamic environment. There are two main strategies for solving these kind of problems. The first is to search in the space of behaviours in order to find one that performs well in the environment. The second is to use statistical techniques and dynamic programming methods to estimate the utility of taking actions in states of the world. In our review we found only three reinforcement learning approaches for the detection of anomalies. Since a cloud environment is in principle suitable for interactions with an agent, the problem of optimal resource allocation in particular could be addressed here. However, a suitable simulation of the cloud environment must be ensured for the learning process, which is a challenge in itself.
- Generative Adversarial Networks (GANs): This deep learning architecture consists of two models. A discriminative model that learns to determine whether a sample is from the model distribution or the data distribution and a generative model that can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency [68]. This framework corresponding to a minimax two-player game, allows to learn complex (normal) data distributions and thus can be used for anomaly detection, as shown for exmaple by [101] and [160].
- Attention mechanisms: For LSTM-based encoder-decoders, attention mechanisms [189] have become an integral part in various tasks, allowing modeling of dependencies without regard to their distance in the input or output sequences. It is worth evaluating the advantages of these models also for sequential cloud data (log data, metrics).
- Graph-based approaches: cloud environments possess topological information (servers, VMs, services, communication processes) that can be modeled using a set of nodes and (weighted) links, commonly defined as graphs. This abstraction in turn allows special methods to be applied that can also be used to detect associated anomalies (for example attacks or bottlenecks) [8, 130]. The field of deep learning is also developing more and more possibilities for learning large and complex networks that are evolving over time [213]. However, these are by no means extensively considered for cloud environments so far.
- Active learning: The key idea behind active learning is that a machine learning algorithm can achieve greater accuracy with fewer training labels if it is allowed to choose the data from which it learns [166]. Thus, these systems attempt to overcome the labeling bottleneck by asking queries in the form of unlabeled instances to be labeled by an oracle (for example a human operator). In this way the detection of relevant anomalies can be improved by the domain knowledge of the operator.
- Adversarial learning: Over the last few years, the weak points of neural networks have increasingly come into the focus of the scientific community. Since a neural network learns a complex decision function, it can easily happen that white spots appear, which enable an attacker to provoke a targeted error behaviour of the model. Especially when neural networks are used to detect intrusions, these weak points should be prevented, since otherwise their use as an IDS can not be considered as safe. In order to make neural networks more robust, adversarial learning is used.
- Explainable AI (XAI): Another field of research, which is becoming more and more important, deals with the question of how to make the decision making process of machine learning and especially deep learning models comprehensible and interpretable for the operator. This is also important, if such models are planned to be used in real operations and therefore should be made compliant with existing processes. Therefore not only the performance but also the explainability should be taken into account in the development of future anomaly detection methods.

5 CONCLUSION

In principle, for every complex and rapidly developing area of science, the question arises how exactly a sufficiently good picture of the relevant work can be obtained. We have chosen a systematic approach that not only reflects already widely cited popular publications, but also attaches importance to an overall view that is based on the wide variety of contributions. We analyzed a total of 215 papers that we extracted from three different scientific databases. On the methodological side, we have analyzed three main fields (machine learning, deep learning, statistical methods) and on the application side, we have identified three main areas

Table 7: Public Datasets Used in the Reviewed Literature.

Dataset	Application area	Format	Source
DARPA-KDD 98	Intrusion detection	Tcpdump	[42]
DARPA-KDD 99	Intrusion detection	Tcpdump	[43]
NLS-KDD	Intrusion detection	Tcpdump	[183]
CIDDS-001	Intrusion detection	NetFlow	[151]
UNM	Intrusion detection	Sendmail system call traces	[188]
NAB	Misconfigurations, failures	AWS server metrics	[6]
UNSW-NB15	Intrusion detection	Tcpdump	[123]
HDFS	Performance problems	Hadoop log files	[199]
BoT-IoT	Intrusion detection	Pcap files	[95]
CICIDS 2017	Intrusion detection	Pcap files	[168]
CSE-CIC-IDS2018	Intrusion detection	Pcap files and event logs	[57]

(intrusion detection, performance monitoring, failure detection), which in turn can be divided into sub-areas. This survey may serve all interested scientists as a guidance for further research.

ACKNOWLEDGMENTS

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the framework of "Software Campus 2.0 (TU Berlin)" (project number 01IS17052).

REFERENCES

[1] Mahmoud Abdelsalam, Ram Krishnan, and Ravi Sandhu. 2017. Clustering-based IaaS cloud monitoring. In *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)*. IEEE, 672–679.

[2] Mahmoud Abdelsalam, Ram Krishnan, and Ravi Sandhu. 2019. Online malware detection in cloud auto-scaling systems using shallow convolutional neural networks. In *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 381–397.

[3] Adel Abusitta, Martine Bellaiche, and Michel Dagenais. 2018. An SVM-based framework for detecting DoS attacks in virtualized clouds under changing environment. *Journal of Cloud Computing* 7, 1 (2018), 9.

[4] Ashkan Aghdai, Kang Xi, and H Jonathan Chao. 2019. Intelligent Anomaly Detection and Mitigation in Data Centers. *arXiv preprint arXiv:1906.06388* (2019).

[5] Bikash Agrawal, Tomasz Wiktorski, and Chunming Rong. 2017. Adaptive real-time anomaly detection in cloud infrastructures. *Concurrency and Computation: Practice and Experience* 29, 24 (2017), e4193.

[6] Subutai Ahmad, Alexander Lavin, Scott Purdy, and Zuha Agha. 2017. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing* 262 (2017), 134–147.

[7] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. 2016. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications* 60 (2016), 19–31.

[8] Leman Akoglu, Hanghang Tong, and Danai Koutra. 2015. Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery* 29, 3 (2015), 626–688.

[9] Shadab Alam, Mohammed Shuaib, and Abdus Samad. 2019. A collaborative study of intrusion detection and prevention techniques in cloud computing. In *International Conference on Innovative Computing and Communications*. Springer, 231–240.

[10] Sultan T Alanazi, Mohammed Anbar, Shankar Karuppayah, Ahmed K Al-Ani, and Yousef K Sanjalawe. 2019. Detection techniques for DDoS attacks in cloud environment. In *Intelligent and Interactive Computing*. Springer, 337–354.

[11] Suaad Alarifi and Stephen Wolthusen. 2013. Anomaly detection for ephemeral cloud IaaS virtual machines. In *International Conference on Network and System Security*. Springer, 321–335.

[12] Mohammad Abdelkareem Alarqan, Zarul Fitri Zaaba, and Ammar Almomani. 2019. Detection Mechanisms of DDoS Attack in Cloud Computing Environment: A Survey. In *International Conference on Advances in Cyber Security*. Springer, 138–152.

[13] Osama AlKadi, Nour Moustafa, Benjamin Turnbull, and Kim-Kwang Raymond Choo. 2019. Mixture Localization-Based Outliers Models for securing Data Migration in Cloud Centers. *IEEE Access* 7 (2019), 114607–114618.

[14] Ameen Alkasem, Hongwei Liu, Zuo Decheng, and Yao Zhao. 2015. AFDI: a virtualization-based accelerated fault diagnosis innovation for high availability computing. *arXiv preprint arXiv:1507.08036* (2015).

[15] Ahmad Alnafessah and Giuliano Casale. 2018. A neural-network driven methodology for anomaly detection in apache spark. In *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. IEEE, 201–209.

[16] Ahmad Alnafessah and Giuliano Casale. 2019. Artificial neural networks based techniques for anomaly detection in Apache Spark. *Cluster Computing* (2019), 1–16.

[17] Deepali Arora and Kin Fun Li. 2017. Detecting anomalies in the data residing over the cloud. In *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. IEEE, 541–546.

[18] Hammi Badis, Guillaume Doyen, and Rida Khatoun. 2014. Toward a source detection of botclouds: a pca-based approach. In *IFIP International Conference on Autonomous Infrastructure, Management and Security*. Springer, 105–117.

[19] Omkar P Badve and BB Gupta. 2016. Taxonomy of recent DDoS attack prevention, detection, and response schemes in cloud environment. In *Proceedings of the international conference on recent cognizance in wireless communication & image processing*. Springer, 683–693.

[20] Abhishek Bajpai, Shruti Singh, et al. 2016. A survey on Security Analysis in Cloud computing. In *Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing*. Springer, 249–262.

[21] V Balamurugan and R Saravanan. 2019. Enhanced intrusion detection and prevention system on cloud environment using hybrid classification and OTS generation. *Cluster Computing* 22, 6 (2019), 13027–13039.

[22] Sakil Barbhuiya, Zafeirios Papazachos, Peter Kilpatrick, and Dimitrios S Nikolopoulos. 2018. RADS: Real-time Anomaly Detection System for Cloud Data Centres. *arXiv preprint arXiv:1811.04481* (2018).

[23] Vic Barnett and Toby Lewis. 1984. Outliers in statistical data. *osd* (1984).

[24] Souhila Benmakrelouf, Cédric St-Onge, Nadjia Kara, Hanine Tout, Claes Edstrom, and Yves Lemieux. 2020. Abnormal behavior detection using resource level to service level metrics mapping in virtualized systems. *Future Generation Computer Systems* 102 (2020), 680–700.

[25] Josep Lluís Berral, Nicolas Poggi, David Carrera, Aaron Call, Rob Reinauer, and Daron Green. 2015. ALOJA: a framework for benchmarking and predictive analytics in Hadoop deployments. *IEEE Transactions on Emerging Topics in Computing* 5, 4 (2015), 480–493.

[26] Elham Besharati, Marjan Naderan, and Ehsan Namjoo. 2019. LR-HIDS: logistic regression host-based intrusion detection system for cloud environments. *Journal of Ambient Intelligence and Humanized Computing* 10, 9 (2019), 3669–3692.

[27] Nitesh Bharot, Veenadhari Suraparaju, and Sanjeev Gupta. 2019. DDoS Attack Detection and Clustering of Attacked and Non-attacked VMs Using SOM in Cloud Network. In *International Conference on Advances in Computing and Data Sciences*. Springer, 369–378.

[28] Dhruba Kumar Bhattacharyya and Jugal Kumar Kalita. 2013. *Network anomaly detection: A machine learning perspective.* Crc Press.

[29] Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita. 2013. Network anomaly detection: methods, systems and tools. *Ieee communications surveys & tutorials* 16, 1 (2013), 303–336.

[30] Rodrigo N Calheiros, Kotagiri Ramamohanarao, Rajkumar Buyya, Christopher Leckie, and Steve Versteeg. 2017. On the effectiveness of isolation-based anomaly detection in cloud data centers. *Concurrency and Computation: Practice and Experience* 29, 18 (2017), e4169.

[31] Lelio Campanile, Mauro Iacono, Fabio Martinelli, Fiammetta Marulli, Michele Mastroianni, Francesco Mercaldo, and Antonella Santone. 2020. Towards the

Use of Generative Adversarial Neural Networks to Attack Online Resources. In *Workshops of the International Conference on Advanced Information Networking and Applications*. Springer, 890–901.

[32] Marta Catillo, Massimiliano Rak, and Umberto Villano. 2020. 2L-ZED-IDS: A Two-Level Anomaly Detector for Multiple Attack Classes. In *Workshops of the International Conference on Advanced Information Networking and Applications*. Springer, 687–696.

[33] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. *ACM computing surveys (CSUR)* 41, 3 (2009), 1–58.

[34] Hongyang Chen, Pengfei Chen, and Guangba Yu. 2020. A Framework of Virtual War Room and Matrix Sketch-Based Streaming Anomaly Detection for Microservice Systems. *IEEE Access* 8 (2020), 43413–43426.

[35] Qiang Chen. 2016. Research on Distributed Anomaly Traffic Detection Technology Based on Hadoop Platform. In *International Conference on Bio-Inspired Computing: Theories and Applications*. Springer, 530–535.

[36] Zouhair Chiba, Noreddine Abghour, Khalid Moussaid, Amina El Omri, and Mohamed Rida. 2018. Novel Network IDS in Cloud Environment Based on Optimized BP Neural Network Using Genetic Algorithm. In *Proceedings of the 3rd International Conference on Smart City Applications*. 1–9.

[37] Zouhair Chiba, Noreddine Abghour, Khalid Moussaid, Amina El Omri, and Mohamed Rida. 2019. An Efficient Network IDS for Cloud Environments Based on a Combination of Deep Learning and an Optimized Self-adaptive Heuristic Search Algorithm. In *International Conference on Networked Systems*. Springer, 235–249.

[38] Zouhair Chiba, Noreddine Abghour, Khalid Moussaid, Amina El Omri, and Mohamed Rida. 2019. Smart Approach to Build A Deep Neural Network Based IDS for Cloud Environment Using an Optimized Genetic Algorithm. In *Proceedings of the 2nd International Conference on Networking, Information Systems & Security*. 1–12.

[39] Chien-Yi Chiu, Chi-Tien Yeh, and Yuh-Jye Lee. 2013. Frequent pattern based user behavior anomaly detection for cloud system. In *2013 Conference on Technologies and Applications of Artificial Intelligence*. IEEE, 61–66.

[40] Ashley Chonka, Yang Xiang, Wanlei Zhou, and Alessio Bonti. 2011. Cloud Security Defence to Protect Cloud Computing against HTTP-DoS and XML-DoS Attacks. *J. Netw. Comput. Appl.* 34, 4 (July 2011), 1097–1107. https://doi.org/10.1016/j.jnca.2010.06.004

[41] Domenico Cotroneo, Luigi De Simone, Pietro Liguori, Roberto Natella, and Nematollah Bidokhti. 2019. Enhancing failure propagation analysis in cloud computing systems. In *2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 139–150.

[42] DARPA. 1998. KDD 98. https://kdd.ics.uci.edu/databases/kddcup98/kddcup98.html. Accessed: 2020-10-26.

[43] DARPA. 1999. KDD 99. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. Accessed: 2020-10-26.

[44] Joel A Dawson, Jeffrey T McDonald, Lee Hively, Todd R Andel, Mark Yampolskiy, and Charles Hubbard. 2018. Phase space detection of virtual machine cyber events through hypervisor-level system call analysis. In *2018 1st International Conference on Data Intelligence and Security (ICDIS)*. IEEE, 159–167.

[45] Marcos VO De Assis, Anderson H Hamamoto, Taufik Abrão, and Mario Lemes Proença. 2017. A game theoretical based system using holt-winters and genetic algorithm with fuzzy logic for DoS/DDoS mitigation on SDN networks. *IEEE Access* 5 (2017), 9485–9496.

[46] Marcos VO De Assis, Matheus P Novaes, Cinara B Zerbini, Luiz F Carvalho, Taufik Abrãao, and Mario L Proença. 2018. Fast defense system against attacks in software defined networks. *IEEE Access* 6 (2018), 69620–69639.

[47] Pooja Dehraj and Arun Sharma. 2020. A review on architecture and models for autonomic software systems. *JOURNAL OF SUPERCOMPUTING* (2020).

[48] Frank Doelitzscher, Martin Knahl, Christoph Reich, and Nathan Clarke. 2013. Anomaly detection in iaas clouds. In *2013 IEEE 5th International Conference on Cloud Computing Technology and Science*, Vol. 1. IEEE, 387–394.

[49] Christos Douligeris and Aikaterini Mitrokotsa. 2004. DDoS attacks and defense mechanisms: classification and state-of-the-art. *Computer Networks* 44, 5 (2004), 643 – 666. https://doi.org/10.1016/j.comnet.2003.10.003

[50] Qingfeng Du, Tiandi Xie, and Yu He. 2018. Anomaly detection and diagnosis for container-based microservices with performance monitoring. In *International Conference on Algorithms and Architectures for Parallel Processing*. Springer, 560–572.

[51] Xindong Duan. 2019. Research on abnormal data detection method of web browser in cloud computing environment. *Cluster Computing* 22, 1 (2019), 1229–1238.

[52] Iman El Mir, Abdelkrim Haqiq, and Dong Seong Kim. 2016. Performance analysis and security based on intrusion detection and prevention systems in cloud data centers. In *International Conference on Hybrid Intelligent Systems*. Springer, 456–465.

[53] Marwa A Elsayed and Mohammad Zulkernine. 2020. PredictDeep: Security Analytics as a Service for Anomaly Detection and Prediction. *IEEE Access* 8 (2020), 45184–45197.

[54] Mostafa Farshchi, Jean-Guy Schneider, Ingo Weber, and John Grundy. 2018. Metric selection and anomaly detection for cloud operations using log and metric correlation analysis. *Journal of Systems and Software* 137 (2018), 531–549.

[55] Diogo AB Fernandes, Mário M Freire, Paulo AP Fazendeiro, and Pedro RM Inácio. 2017. Applications of artificial immune systems to computer security: A survey. *Journal of Information Security and Applications* 35 (2017), 138–159.

[56] Jaron Fontaine, Chris Kappler, Adnan Shahid, and Eli De Poorter. 2019. Log-Based Intrusion Detection for Cloud Web Applications Using Machine Learning. In *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*. Springer, 197–210.

[57] Canadian Institute for Cybersecurity. 2018. CSE-CIC-IDS2018. https://www.unb.ca/cic/datasets/ids-2018.html. Accessed: 2020-10-26.

[58] Song Fu, Jianguo Liu, and Husanbir Pannu. 2012. A hybrid anomaly detection framework in cloud computing using one-class and two-class support vector machines. In *International Conference on Advanced Data Mining and Applications*. Springer, 726–738.

[59] Matthias Gander, Michael Felderer, Basel Katt, Adrian Tolbaru, Ruth Breu, and Alessandro Moschitti. 2012. Anomaly detection in the cloud: Detecting security incidents via machine learning. In *International Workshop on Eternal Systems*. Springer, 103–116.

[60] P Ganeshkumar and N Pandeeswari. 2016. Adaptive neuro-fuzzy-based anomaly detection system in cloud. *International Journal of Fuzzy Systems* 18, 3 (2016), 367–378.

[61] Sahil Garg, Kuljeet Kaur, Shalini Batra, Gagangeet Singh Aujla, Graham Morgan, Neeraj Kumar, Albert Y Zomaya, and Rajiv Ranjan. 2020. En-ABC: An ensemble artificial bee colony based anomaly detection scheme for cloud environment. *J. Parallel and Distrib. Comput.* 135 (2020), 219–233.

[62] Sahil Garg, Kuljeet Kaur, Neeraj Kumar, Georges Kaddoum, Albert Y Zomaya, and Rajiv Ranjan. 2019. A hybrid deep learning-based model for anomaly detection in cloud datacenter networks. *IEEE Transactions on Network and Service Management* 16, 3 (2019), 924–935.

[63] Partha Ghosh, Meghna Bardhan, Nilabhra Roy Chowdhury, Santanu Phadikar, et al. 2017. IDS using reinforcement learning Automata for Preserving security in cloud environment. *International Journal of Information System Modeling and Design (IJISMD)* 8, 4 (2017), 21–37.

[64] Partha Ghosh, Abhay Kumar Mandal, and Rupesh Kumar. 2015. An efficient cloud network intrusion detection system. In *Information systems design and intelligent applications*. Springer, 91–99.

[65] Partha Ghosh, Shivam Shakti, and Santanu Phadikar. 2016. A cloud intrusion detection system using novel PRFCM clustering and KNN based dempster-shafer rule. *International Journal of Cloud Applications and Computing (IJCAC)* 6, 4 (2016), 18–35.

[66] Anteneh Girma, Mosses Garuba, and Rajini Goel. 2018. Advanced machine language approach to detect DDoS attack using DBSCAN clustering technology with entropy. In *Information Technology-New Generations*. Springer, 125–131.

[67] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*. Vol. 1. MIT press Cambridge.

[68] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[69] Qiang Guan, Song Fu, Nathan DeBardeleben, and Sean Blanchard. 2013. Exploring time and frequency domains for accurate and automated anomaly detection in cloud computing systems. In *2013 IEEE 19th Pacific Rim International Symposium on Dependable Computing*. IEEE, 196–205.

[70] Fabio Guigou, Pierre Collet, and Pierre Parrend. 2017. Anomaly detection and motif discovery in symbolic representations of time series. *arXiv preprint arXiv:1704.05325* (2017).

[71] Fabio Guigou, Pierre Collet, and Pierre Parrend. 2019. SCHEDA: Lightweight euclidean-like heuristics for anomaly detection in periodic time series. *Applied Soft Computing* 82 (2019), 105594.

[72] Halim Görkem Gülmez, Emrah Tuncel, and Pelin Angin. 2018. A big data analytical approach to cloud intrusion detection. In *International Conference on Cloud Computing*. Springer, 377–388.

[73] Hang Guo, Xun Fan, Anh Cao, Geoff Outhred, and John Heidemann. 2019. Peek Inside the Closed World: Evaluating Autoencoder-Based Detection of DDoS to Cloud. *arXiv preprint arXiv:1912.05590* (2019).

[74] Waqas Haider, Jiankun Hu, and Nour Moustafa. 2017. Designing anomaly detection system for cloud servers by frequency domain features of system call identifiers and machine learning. In *International Conference on Mobile Networks and Management*. Springer, 137–149.

[75] Douglas M Hawkins. 1980. *Identification of outliers*. Vol. 11. Springer.

[76] Jordan Hochenbaum, Owen S Vallis, and Arun Kejariwal. 2017. Automatic anomaly detection in the cloud via statistical learning. *arXiv preprint arXiv:1704.07706* (2017).

[77] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[78] Bin Hong, Yazhou Hu, Fuyang Peng, and Bo Deng. 2015. Distributed state monitoring for IaaS Cloud with continuous observation sequence. In *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*. IEEE, 1037–1042.

[79] Bin Hong, Fuyang Peng, Bo Deng, Yazhou Hu, and Dongxia Wang. 2015. DAC-Hmm: detecting anomaly in cloud systems with hidden Markov models. *Concurrency and Computation: Practice and Experience* 27, 18 (2015), 5749–5764.

[80] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. 2006. Extreme learning machine: theory and applications. *Neurocomputing* 70, 1-3 (2006), 489–501.

[81] Tian Huang, Yongxin Zhu, Yafei Wu, Stéphane Bressan, and Gillian Dobbie. 2016. Anomaly detection and identification scheme for VM live migration in cloud infrastructure. *Future Generation Computer Systems* 56 (2016), 736–745.

[82] Tian Huang, Yan Zhu, Qiannan Zhang, Yongxin Zhu, Dongyang Wang, Meikang Qiu, and Lei Liu. 2013. An lof-based adaptive anomaly detection scheme for cloud computing. In *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*. IEEE, 206–211.

[83] Nurudeen Mahmud Ibrahim and Anazida Zainal. 2019. An Adaptive Intrusion Detection Scheme for Cloud Computing. *International Journal of Swarm Intelligence Research (IJSIR)* 10, 4 (2019), 53–70.

[84] Nurudeen Mahmud Ibrahim and Anazida Zainal. 2020. A Distributed Intrusion Detection Scheme for Cloud Computing. *International Journal of Distributed Systems and Technologies (IJDST)* 11, 1 (2020), 68–82.

[85] Mohamed Idhammad, Karim Afdel, and Mustapha Belouch. 2018. Distributed intrusion detection system for cloud environments based on data mining techniques. *Procedia Computer Science* 127 (2018), 35–41.

[86] Mohammad Islam and Andriy Miranskyy. 2020. Anomaly Detection in Cloud Components.

[87] Aws Naser Jaber and Shafiq Ul Rehman. 2020. FCM–SVM based intrusion detection system for cloud computing environment. *Cluster Computing* (2020), 1–11.

[88] Aws Naser Jaber, Mohamad Fadli Zolkipli, Hasan Awni Shakir, and Mohammed R Jassim. 2017. Host based intrusion detection and prevention model against DDoS attack in cloud computing. In *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*. Springer, 241–252.

[89] Hiranya Jayathilaka, Chandra Krintz, and Rich Wolski. 2017. Performance monitoring and root cause analysis for cloud-hosted web applications. In *Proceedings of the 26th International Conference on World Wide Web*. 469–478.

[90] Priyanka Joshi, Ritu Prasad, Pradeep Mewada, and Praneet Saurabh. 2018. A New Neural Network-Based IDS for Cloud Computing. In *Progress in Computing, Analytics and Networking*. Springer, 161–170.

[91] El Mehdi Kandoussi, Iman El Mir, Mohamed Hanini, and Abdelkrim Haqiq. 2017. Modeling an anomaly-based intrusion prevention system using game theory. In *International conference on innovations in bio-inspired computing and applications*. Springer, 266–276.

[92] Hisham A Kholidy. 2019. Correlation-based sequence alignment models for detecting masquerades in cloud computing. *IET Information Security* 14, 1 (2019), 39–50.

[93] Hyunjoo Kim, Jonghyun Kim, Youngsoo Kim, Ikkyun Kim, and Kuinam J Kim. 2019. Design of network threat detection and classification based on machine learning on cloud computing. *Cluster Computing* 22, 1 (2019), 2341–2350.

[94] Teuvo Kohonen. 2012. *Self-organizing maps*. Vol. 30. Springer Science & Business Media.

[95] Nickolaos Koroniotis, Nour Moustafa, Elena Sitnikova, and Benjamin Turnbull. 2019. Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *Future Generation Computer Systems* 100 (2019), 779–796.

[96] Rafał Kozik, Michał Choraś, Witold Hołubowicz, and Rafał Renk. 2016. Extreme learning machines for web layer anomaly detection. In *International Conference on Image Processing and Communications*. Springer, 226–233.

[97] S Krishnaveni, Palani Vigneshwar, S Kishore, B Jothi, and S Sivamohan. 2020. Anomaly-Based Intrusion Detection System Using Support Vector Machine. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*. Springer, 723–731.

[98] Ram Shankar Siva Kumar, Andrew Wicker, and Matt Swann. 2017. Practical machine learning for cloud intrusion detection: challenges and the way forward. In *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*. 81–90.

[99] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1, 4 (1989), 541–551.

[100] Alexander Lenk, Markus Klems, Jens Nimis, Stefan Tai, and Thomas Sandholm. 2009. What's inside the Cloud? An architectural map of the Cloud landscape. In *2009 ICSE workshop on software engineering challenges of cloud computing*. IEEE, 23–31.

[101] Dan Li, Dacheng Chen, Jonathan Goh, and See-kiong Ng. 2018. Anomaly detection with generative adversarial networks for multivariate time series. *arXiv preprint arXiv:1809.04758* (2018).

[102] Mingwei Lin, Zhiqiang Yao, Fei Gao, and Yang Li. 2015. Toward anomaly detection in iaas cloud computing platforms. *International Journal of Security and Its Applications* 9, 12 (2015), 175–188.

[103] Mingwei Lin, Zhiqiang Yao, Fei Gao, and Yang Li. 2016. A Virtual Machine Instance Anomaly Detection System for IaaS Cloud Computing. *International Journal of Future Generation Communication and Networking* 9, 3 (2016), 255–268.

[104] Jiaxin Liu, Xucheng Song, Yingjie Zhou, Xi Peng, Yanru Zhang, Pei Liu, and Dapeng Wu. 2019. Deep Anomaly Detection in Packet Payload. *arXiv preprint arXiv:1912.02549* (2019).

[105] Yuan Liu and Ruhui Ma. 2013. Network anomaly detection based on BQPSO-BN algorithm. *IETE Journal of Research* 59, 4 (2013), 334–342.

[106] Tania Lorido-Botran, Sergio Huerta, Luis Tomás, Johan Tordsson, and Borja Sanz. 2017. An unsupervised approach to online noisy-neighbor detection in cloud data centers. *Expert Systems with Applications* 89 (2017), 188–204.

[107] Widad Mirghani Makki, Maheyzah MD Siraj, and Nurudeen Mahmud Ibrahim. 2019. A Harmony Search-Based Feature Selection Technique for Cloud Intrusion Detection. In *International Conference of Reliable Information and Communication Technology*. Springer, 779–788.

[108] Haroon Malik, Ian J Davis, Michael W Godfrey, Douglas Neuse, and Serge Manskovskii. 2016. Connecting the dots: anomaly and discontinuity detection in large-scale systems. *Journal of Ambient Intelligence and Humanized Computing* 7, 4 (2016), 509–522.

[109] M Manickam and SP Rajagopalan. 2019. A hybrid multi-layer intrusion detection system in cloud. *Cluster Computing* 22, 2 (2019), 3961–3969.

[110] M Mayuranathan, M Murugan, and V Dhanakoti. 2019. Best features based intrusion detection system by RBM model for detecting DDoS in cloud environment. *Journal of Ambient Intelligence and Humanized Computing* (2019), 1–11.

[111] Peter Mell, Tim Grance, et al. 2011. The NIST definition of cloud computing. (2011).

[112] Fan Jing Meng, Xiao Zhang, Pengfei Chen, and Jing Min Xu. 2017. Driftinsight: detecting anomalous behaviors in large-scale cloud platform. In *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)*. IEEE, 230–237.

[113] HaiBo Mi, HuaiMin Wang, YangFan Zhou, Michael R Lyu, and Hua Cai. 2012. Localizing root causes of performance anomalies in cloud computing systems by analyzing request trace logs. *Science China Information Sciences* 55, 12 (2012), 2757–2773.

[114] Haibo Mi, Huaimin Wang, Yangfan Zhou, Michael Rung-Tsong Lyu, Hua Cai, and Gang Yin. 2013. An online service-oriented performance profiling tool for cloud computing systems. *Frontiers of Computer Science* 7, 3 (2013), 431–445.

[115] Preeti Mishra, Akash Negi, ES Pilli, and RC Joshi. 2019. VMProtector: Malign Process Detection for Protecting Virtual Machines in Cloud Environment. In *International Conference on Advances in Computing and Data Sciences*. Springer, 360–369.

[116] Preeti Mishra, Emmanuel S Pilli, Vijay Varadharajan, and Udaya Tupakula. 2016. Securing virtual machines from anomalies using program-behavior analysis in cloud environment. In *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 991–998.

[117] Preeti Mishra, Emmanuel S Pilli, Vijay Varadharajan, and Udaya Tupakula. 2017. Intrusion detection techniques in cloud environment: A survey. *Journal of Network and Computer Applications* 77 (2017), 18–47.

[118] Chirag Modi and Dhiren Patel. 2018. A feasible approach to intrusion detection in virtual network layer of Cloud computing. *Sādhanā* 43, 7 (2018), 114.

[119] Chirag N Modi and Kamatchi Acha. 2017. Virtualization layer security challenges and intrusion detection/prevention systems in cloud computing: a comprehensive review. *the Journal of Supercomputing* 73, 3 (2017), 1192–1234.

[120] Sara Kardani Moghaddam, Rajkumar Buyya, and Kotagiri Ramamohanarao. 2019. ACAS: An anomaly-based cause aware auto-scaling framework for clouds. *J. Parallel and Distrib. Comput.* 126 (2019), 107–120.

[121] Nour Moustafa, Gideon Creech, Elena Sitnikova, and Marwa Keshk. 2017. Collaborative anomaly detection framework for handling big data of cloud computing. In *2017 Military Communications and Information Systems Conference (MilCIS)*. IEEE, 1–6.

[122] Nour Moustafa, Jiankun Hu, and Jill Slay. 2019. A holistic review of network anomaly detection systems: A comprehensive survey. *Journal of Network and Computer Applications* 128 (2019), 33–55.

[123] Nour Moustafa and Jill Slay. 2015. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 military communications and information systems conference (MilCIS)*. IEEE, 1–6.

[124] Pandeeswari Nagarajan and Ganeshkumar Perumal. 2015. A neuro fuzzy based intrusion detection system for a cloud data center using adaptive learning.

Cybernetics and Information Technologies 15, 3 (2015), 88–103.
[125] T Nathiya and G Suseendran. 2019. An Effective Hybrid Intrusion Detection System for Use in Security Monitoring in the Virtual Network Layer of Cloud Computing Technology. In *Data Management, Analytics and Innovation*. Springer, 483–497.
[126] Anjum Nazir and Rizwan Ahmed Khan. 2019. Combinatorial Optimization based Feature Selection Method: A study on Network Intrusion Detection. *arXiv preprint arXiv:1906.04494* (2019).
[127] Hani Neuvirth, Yehuda Finkelstein, Amit Hilbuch, Shai Nahum, Daniel Alon, and Elad Yom-Tov. 2015. Early detection of fraud storms in the cloud. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 53–67.
[128] Laisen Nie, Dingde Jiang, and Zhihan Lv. 2017. Modeling network traffic for traffic matrix estimation and anomaly detection based on Bayesian network in cloud computing networks. *Annals of Telecommunications* 72, 5-6 (2017), 297–305.
[129] Roman Nikiforov. 2018. Clustering-based Anomaly Detection for microservices. *arXiv preprint arXiv:1810.02762* (2018).
[130] Caleb C Noble and Diane J Cook. 2003. Graph-based anomaly detection. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. 631–636.
[131] Onyekachi Nwamuo, Paulo Magella de Faria Quinan, Issa Traore, Isaac Woungang, and Abdulaziz Aldribi. 2019. Arguments Against Using the 1998 DARPA Dataset for Cloud IDS Design and Evaluation and Some Alternative. In *International Conference on Machine Learning for Networking*. Springer, 315–332.
[132] Emmanuel C. Ogu, Olusegun A. Ojesanmi, Oludele Awodele, and 'Shade Kuyoro. 2019. A Botnets Circumspection: The Current Threat Landscape, and What We Know So Far. *Inf.* 10, 11 (2019), 337. https://doi.org/10.3390/info10110337
[133] Opeyemi Osanaiye, Haibin Cai, Kim-Kwang Raymond Choo, Ali Dehghantanha, Zheng Xu, and Mqhele Dlodlo. 2016. Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. *EURASIP Journal on Wireless Communications and Networking* 2016, 1 (2016), 130.
[134] Opeyemi Osanaiye, Kim-Kwang Raymond Choo, and Mqhele Dlodlo. 2016. Distributed denial of service (DDoS) resilience in cloud: review and conceptual cloud DDoS mitigation framework. *Journal of Network and Computer Applications* 67 (2016), 147–165.
[135] David O'Shea, Vincent C Emeakaroha, John Pendlebury, Neil Cafferkey, John P Morrison, and Theo Lynn. 2016. A Wavelet-inspired Anomaly Detection Framework for Cloud Platforms. In *CLOSER (1)*. 106–117.
[136] Cemile Diler Özdemir, Mehmet Tahir Sandıkkaya, and Yusuf Yaslan. 2018. Malicious Behavior Classification in PaaS. In *International Conference on Cloud Computing and Services Science*. Springer, 215–232.
[137] David O'Shea, Vincent C Emeakaroha, Neil Cafferkey, John P Morrison, and Theo Lynn. 2016. Detecting Anomaly in Cloud Platforms Using a Wavelet-Based Framework. In *International Conference on Cloud Computing and Services Science*. Springer, 131–150.
[138] N Pandeeswari and R Karuppathal. 2017. Hypervisor Based Anomaly Detection System in Cloud Computing Using ANFIS. 路技刊 18, 6 (2017), 1335–1344.
[139] N Pandeeswari and Ganesh Kumar. 2016. Anomaly detection system in cloud environment using fuzzy clustering based ANN. *Mobile Networks and Applications* 21, 3 (2016), 494–505.
[140] Ahmed Patel, Mona Taghavi, Kaveh Bakhtiyari, and Joaquim Celestino Júnior. 2012. Taxonomy and proposed architecture of intrusion detection and prevention systems for cloud computing. In *International Symposium on Cyberspace Safety and Security*. Springer, 441–458.
[141] Rajendra Patil, Harsha Dudeja, and Chirag Modi. 2020. Designing in-VM-assisted lightweight agent-based malware detection framework for securing virtual machines in cloud computing. *International Journal of Information Security* 19, 2 (2020), 147–162.
[142] Ady Wahyudi Paundu, Takeshi Okuda, Youki Kadobayashi, and Suguru Yamaguchi. 2015. Leveraging static probe instrumentation for vm-based anomaly detection system. In *International Conference on Information and Communications Security*. Springer, 320–334.
[143] Damai Jessica Prathyusha, Shaik Naseera, DJ Anusha, and K Alisha. 2020. A Review of Biologically Inspired Algorithms in a Cloud Environment to Combat DDoS Attacks. In *Smart Intelligent Computing and Applications*. Springer, 59–68.
[144] Juan Qiu, Qingfeng Du, Yu He, YiQun Lin, Jiaye Zhu, and Kanglin Yin. 2018. Performance anomaly detection models of virtual machines for network function virtualization infrastructure with machine learning. In *International Conference on Artificial Neural Networks*. Springer, 479–488.
[145] Mahdi Rabbani, Yong Li Wang, Reza Khoshkangini, Hamed Jelodar, Ruxin Zhao, and Peng Hu. 2020. A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing. *Journal of Network and Computer Applications* 151 (2020), 102507.
[146] Sivakami Raja and Saravanan Ramaiah. 2017. An efficient fuzzy-based hybrid system to cloud intrusion detection. *International Journal of Fuzzy Systems* 19, 1 (2017), 62–77.
[147] Sutharshan Rajasegarar, Christopher Leckie, and Marimuthu Palaniswami. 2008. Anomaly detection in wireless sensor networks. *IEEE Wireless Communications* 15, 4 (2008), 34–40.
[148] Divya Rajput and Ankit Thakkar. 2019. A Survey on Different Network Intrusion Detection Systems and CounterMeasure. In *Emerging Research in Computing, Information, Communication and Applications*. Springer, 497–506.
[149] AC Ramachandra, Subhrajit Bhattacharya, et al. 2020. Literature Survey on Log-Based Anomaly Detection Framework in Cloud. In *Computational Intelligence in Pattern Recognition*. Springer, 143–153.
[150] Rui Ren, Jinheng Li, Lei Wang, Jianfeng Zhan, and Zheng Cao. 2018. Anomaly Analysis for Co-located Datacenter Workloads in the Alibaba Cluster. *arXiv preprint arXiv:1811.06901* (2018).
[151] Markus Ring, Sarah Wunderlich, Dominik Grüdl, Dieter Landes, and Andreas Hotho. 2017. Flow-based benchmark data sets for intrusion detection. In *Proceedings of the 16th European conference on cyber warfare and security*. 361–369.
[152] S. Roschke, F. Cheng, and C. Meinel. 2009. Intrusion Detection in the Cloud. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*. 729–734. https://doi.org/10.1109/DASC.2009.94
[153] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. 1986. Learning representations by back-propagating errors. *nature* 323, 6088 (1986), 533–536.
[154] Mikail Mohammed Salim, Shailendra Rathore, and Jong Hyuk Park. 2019. Distributed denial of service attacks and its defenses in IoT: a survey. *The Journal of Supercomputing* (2019), 1–44.
[155] Mohd Rafiz Salji, Nur Izura Udzir, Mohd Izuan Hafez Ninggal, Nor Fazlida Mohd Sani, and Hamidah Ibrahim. 2016. An Anomaly Detection Algorithm based on Online Learning Lagrangian SVM for Cloud Computing Environment. *International Journal of Security and Its Applications* 10, 12 (2016), 173–186.
[156] Tara Salman, Deval Bhamare, Aiman Erbad, Raj Jain, and Mohammed Samaka. 2017. Machine learning for anomaly detection and categorization in multi-cloud environments. In *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*. IEEE, 97–103.
[157] Mehmet Tahir Sandıkkaya, Yusuf Yaslan, and Cemile Diler Özdemir. 2019. DeMETER in clouds: detection of malicious external thread execution in runtime with machine learning in PaaS clouds. *Cluster Computing* (2019), 1–14.
[158] S Sandosh, V Govindasamy, and G Akila. 2020. Enhanced intrusion detection system via agent clustering and classification based on outlier detection. *Peer-to-Peer Networking and Applications* (2020), 1–8.
[159] Carla Sauvanaud, Mohamed Kaâniche, Karama Kanoun, Kahina Lazri, and Guthemberg Da Silva Silvestre. 2018. Anomaly detection and diagnosis for cloud services: Practical experiments and lessons learned. *Journal of Systems and Software* 139 (2018), 84–106.
[160] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. 2019. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis* 54 (2019), 30–44.
[161] Florian Schmidt, Florian Suri-Payer, Anton Gulenko, Marcel Wallschläger, Alexander Acker, and Odej Kao. 2018. Unsupervised anomaly event detection for cloud monitoring using online arima. In *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*. IEEE, 71–76.
[162] Florian Schmidt, Florian Suri-Payer, Anton Gulenko, Marcel Wallschläger, Alexander Acker, and Odej Kao. 2018. Unsupervised Anomaly Event Detection for VNF Service Monitoring Using Multivariate Online Arima. In *2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*. IEEE, 278–283.
[163] Aravinthkumar Selvaraj, Rizwan Patan, Amir H Gandomi, Ganesh Gopal Deverajan, and Manjula Pushparaj. 2019. Optimal virtual machine selection for anomaly detection using a swarm intelligence approach. *Applied soft computing* 84 (2019), 105686.
[164] Jitendra Kumar Seth and Satish Chandra. 2018. An Effective DOS Attack Detection Model in Cloud Using Artificial Bee Colony Optimization. *3D Research* 9, 3 (2018), 44.
[165] Jitendra Kumar Seth and Satish Chandra. 2018. MIDS: Metaheuristic Based Intrusion Detection System for Cloud Using k-NN and MGWO. In *International Conference on Advances in Computing and Data Sciences*. Springer, 411–420.
[166] Burr Settles. 2009. *Active learning literature survey*. Technical Report. University of Wisconsin-Madison Department of Computer Sciences.
[167] Wenyao Sha, Yongxin Zhu, Min Chen, and Tian Huang. 2015. Statistical learning for anomaly detection in cloud server systems: A multi-order Markov chain framework. *IEEE transactions on cloud computing* 6, 2 (2015), 401–413.
[168] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A Ghorbani. 2018. Toward generating a new intrusion detection dataset and intrusion traffic characterization.. In *ICISSP*. 108–116.
[169] Vishal Sharma, Vinay Verma, and Anand Sharma. 2019. Detection of DDoS Attacks Using Machine Learning in Cloud Computing. In *International Conference on Advanced Informatics for Computing Research*. Springer, 260–273.
[170] Chaochen Shi and Jiangshan Yu. 2019. A Hidden Markov Model-Based Method for Virtual Machine Anomaly Detection. In *International Conference on Provable Security*. Springer, 372–380.

[171] Guthemberg Silvestre, Carla Sauvanaud, Mohamed Kaâniche, and Karama Kanoun. 2015. Tejo: A supervised anomaly detection scheme for newSQL databases. In *International Workshop on Software Engineering for Resilient Systems*. Springer, 114–127.

[172] Steven Simpson, Simon Oechsner, Andreas Mauthe, David Hutchison, et al. 2015. A framework for resilience management in the cloud. *e & i Elektrotechnik und Informationstechnik* 132, 2 (2015), 122–132.

[173] GS Smrithy and Ramadoss Balakrishnan. 2016. A statistical technique for online anomaly detection for big data streams in cloud collaborative environment. In *2016 IEEE International Conference on Computer and Information Technology (CIT)*. IEEE, 108–111.

[174] Imen Souiden, Zaki Brahmi, and Lamine Lafi. 2017. Data stream mining based-outlier prediction for cloud computing. In *International Conference on Digital Economy*. Springer, 131–142.

[175] T Raja Sree and S Mary Saira Bhanu. 2019. Detection of HTTP flooding attacks in cloud using fuzzy bat clustering. *Neural Computing and Applications* (2019), 1–17.

[176] Karthik Srinivasan, Azath Mubarakali, Abdulrahman Saad Alqahtani, and A Dinesh Kumar. 2019. A Survey on the Impact of DDoS Attacks in Cloud Computing: Prevention, Detection and Mitigation Techniques. In *Intelligent Communication Technologies and Virtual Mobile Networks*. Springer, 252–270.

[177] Madhan Kumar Srinivasan, K Sarukesi, Ashima Keshava, and P Revathy. 2012. ecloudids tier-1 ux-engine subsystem design and implementation using self-organizing map (som) for secure cloud computing environment. In *International Conference on Security in Computer Networks and Distributed Systems*. Springer, 432–443.

[178] Siddharth Srinivasan, Akshay Kumar, Manik Mahajan, Dinkar Sitaram, and Sanchika Gupta. 2018. Probabilistic real-time intrusion detection system for docker containers. In *International Symposium on Security in Computing and Communication*. Springer, 336–347.

[179] Ioannis M Stephanakis, Ioannis P Chochliouros, Evangelos Sfakianakis, Syed Noorulhassan Shirazi, and David Hutchison. 2019. Hybrid self-organizing feature map (SOM) for anomaly detection in cloud infrastructures using granular clustering based upon value-difference metrics. *Information Sciences* 494 (2019), 247–277.

[180] EK Subramanian and Latha Tamilselvan. 2019. A focus on future cloud: machine learning-based cloud security. *Service Oriented Computing and Applications* 13, 3 (2019), 237–249.

[181] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.

[182] Hassan Takabi, James BD Joshi, and Gail-Joon Ahn. 2010. Security and privacy challenges in cloud computing environments. *IEEE Security & Privacy* 8, 6 (2010), 24–31.

[183] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A Ghorbani. 2009. A detailed analysis of the KDD CUP 99 data set. In *2009 IEEE symposium on computational intelligence for security and defense applications*. IEEE, 1–6.

[184] Marina Thottan and Chuanyi Ji. 2003. Anomaly detection in IP networks. *IEEE Transactions on signal processing* 51, 8 (2003), 2191–2204.

[185] Siva Rama Krishna Tummalapalli and ASN Chakravarthy. 2020. Intrusion detection system for cloud forensics using bayesian fuzzy clustering and optimization based SVNN. *Evolutionary Intelligence* (2020), 1–11.

[186] Ozan Tuncer, Emre Ates, Yijia Zhang, Ata Turk, Jim Brandt, Vitus J Leung, Manuel Egele, and Ayse K Coskun. 2017. Diagnosing performance variations in HPC applications using machine learning. In *International Supercomputing Conference*. Springer, 355–373.

[187] Olufogorehan Tunde-Onadele, Jingzhu He, Ting Dai, and Xiaohui Gu. 2019. A study on container vulnerability exploit detection. In *2019 IEEE International Conference on Cloud Engineering (IC2E)*. IEEE, 121–127.

[188] UNM. 1998. Sequence-based Intrusion Detection. https://www.cs.unm.edu/~immsec/systemcalls.htm. Accessed: 2020-10-26.

[189] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[190] S Velliangiri and J Premalatha. 2019. Intrusion detection of distributed denial of service attack in cloud. *Cluster Computing* 22, 5 (2019), 10615–10623.

[191] Toby Velte, Anthony Velte, and Robert Elsenpeter. 2009. *Cloud computing, a practical approach*. McGraw-Hill, Inc.

[192] Priyanka Verma, Shashikala Tapaswi, and W Wilfred Godfrey. 2020. An Adaptive Threshold-Based Attribute Selection to Classify Requests Under DDoS Attack in Cloud-Based Systems. *Arabian Journal for Science and Engineering* 45, 4 (2020), 2813–2834.

[193] GuiPing Wang and JiaWei Wang. 2016. An anomaly detection framework for detecting anomalous virtual machines under cloud computing environment. *International Journal of Security and Its Applications* 10, 1 (2016), 75–86.

[194] GuiPing Wang, JianXi Yang, and Ren Li. 2019. UFKLDA: An unsupervised feature extraction algorithm for anomaly detection under cloud environment. *ETRI Journal* 41, 5 (2019), 684–695.

[195] Tao Wang, Jiwei Xu, Wenbo Zhang, Zeyu Gu, and Hua Zhong. 2018. Self-adaptive cloud monitoring with online anomaly detection. *Future Generation Computer Systems* 80 (2018), 89–101.

[196] Michael R Watson, Angelos K Marnerides, Andreas Mauthe, David Hutchison, et al. 2015. Malware detection in cloud computing infrastructures. *IEEE Transactions on Dependable and Secure Computing* 13, 2 (2015), 192–205.

[197] Wei Xiong, Hanping Hu, Naixue Xiong, Laurence T Yang, Wen-Chih Peng, Xiaofei Wang, and Yanzhen Qu. 2014. Anomaly secure detection methods by analyzing dynamic characteristics of the network traffic in cloud communications. *Information Sciences* 258 (2014), 403–415.

[198] Ke Xu, Yun Wang, Leni Yang, Yifang Wang, Bo Qiao, Si Qin, Yong Xu, Haidong Zhang, and Huamin Qu. 2019. CloudDet: Interactive Visual Analysis of Anomalous Performances in Cloud Computing Systems. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 1107–1117.

[199] Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael I Jordan. 2009. Detecting large-scale system problems by mining console logs. In *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles*. 117–132.

[200] Youchang Xu, Ningjiang Chen, Ruwei Huang, and Hanlin Zhang. 2018. KPI Data Anomaly Detection Strategy for Intelligent Operation and Maintenance Under Cloud Environment. In *International Conference on Intelligent Information Processing*. Springer, 311–320.

[201] Youchang Xu, Ningjiang Chen, Hanlin Zhang, and Birui Liang. 2018. Adaptive Anomaly Detection Strategy Based on Reinforcement Learning. In *International Conference of Pioneering Computer Scientists, Engineers and Educators*. Springer, 493–504.

[202] Xiaoben Yan, Wei Zhou, Yun Gao, Zhang Zhang, Jizhong Han, and Ge Fu. 2014. Padm: Page rank-based anomaly detection method of log sequences by graph computing. In *2014 IEEE 6th International Conference on Cloud Computing Technology and Science*. IEEE, 700–703.

[203] Chen Yang. 2019. Anomaly network traffic detection algorithm based on information entropy measurement under the cloud computing environment. *Cluster Computing* 22, 4 (2019), 8309–8317.

[204] Kejiang Ye. 2017. Anomaly Detection in Clouds: Challenges and Practice. In *Proceedings of the First Workshop on Emerging Technologies for Software-Defined and Reconfigurable Hardware-Accelerated Cloud Datacenters* (Xi'an, China) *(ETCD'17)*. Association for Computing Machinery, New York, NY, USA, Article 6, 2 pages. https://doi.org/10.1145/3129457.3129497

[205] Kejiang Ye, Yangyang Liu, Guoyao Xu, and Cheng-Zhong Xu. 2018. Fault injection and detection for artificial intelligence applications in container-based clouds. In *International Conference on Cloud Computing*. Springer, 112–127.

[206] Chuan Yin, Canlin Pan, and Pengquan Zhang. 2020. Deep neural network combined with MapReduce for abnormal data mining and detection in cloud storage. *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING* (2020).

[207] Yue Yuan, Anuhan Torgonshar, Wenchang Shi, Bin Liang, and Bo Qin. 2018. Digging Evidence for Violation of Cloud Security Compliance with Knowledge Learned from Logs. In *Chinese Conference on Trusted Computing and Information Security*. Springer, 318–337.

[208] Mattia Zago, Manuel Gil Pérez, and Gregorio Martínez Pérez. 2019. Scalable detection of botnets based on DGA. *Soft Computing* (2019), 1–21.

[209] Jing Zhang. 2019. Anomaly detecting and ranking of the cloud computing platform by multi-view learning. *Multimedia Tools and Applications* 78, 21 (2019), 30923–30942.

[210] Jian Zhang, Yawei Zhang, Pin Liu, and Jianbiao He. 2016. A spark-based DDoS attack detection model in cloud services. In *International Conference on Information Security Practice and Experience*. Springer, 48–64.

[211] Qiannan Zhang, Yafei Wu, Tian Huang, and Yongxin Zhu. 2013. An intelligent anomaly detection and reasoning scheme for VM live migration via cloud data mining. In *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*. IEEE, 412–419.

[212] Xiao Zhang, Fanjing Meng, and Jingmin Xu. 2018. Perfinsight: A robust clustering-based abnormal behavior detection system for large-scale cloud. In *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*. IEEE, 896–899.

[213] Ziwei Zhang, Peng Cui, and Wenwu Zhu. 2020. Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[214] Beilei Zheng, Jianan Gu, and Chuliang Weng. 2019. CBA-Detector: An Accurate Detector Against Cache-Based Attacks Using HPCs and Pintools. In *International Symposium on Advanced Parallel Processing Technologies*. Springer, 109–122.

[215] Jingwen Zhou, Zhenbang Chen, Ji Wang, Zibin Zheng, and Michael R Lyu. 2015. A Data Set for User Request Trace-Oriented Monitoring and its Applications. *IEEE Transactions on Services Computing* 11, 4 (2015), 699–712.

[216] Mikhail Zolotukhin, Elena Ivannikova, and Timo Hämäläinen. 2017. On detection of network-based co-residence verification attacks in sdn-driven clouds. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Springer, 235–246.