A Review of Cloud-Based Malware Detection System: Opportunities, Advances and Challenges

Ömer Aslan, Merve Ozkan-Okay, and Deepti Gupta

***Abstract* — Cloud computing has an important role in all aspects of storing information and providing services online. It brings several advantages over traditional storing and sharing schema such as an easy access, on-request storage, scalability and decreasing cost. Using its rapidly developing technologies can bring many advantages to the protection of Internet of Things (IoT), Cyber-Physical Systems (CPS) from a variety of cyber-attacks, where IoT, CPS provides facilities to humans in their daily lives. Since malicious software (malware) is increasing exponentially and there is no well-known approach to detecting malware, the usage of cloud environments to detect malware can be a promising method. A new generation of malware is using advanced obfuscation and packing techniques to escape from detection systems. This situation makes almost impossible to detect complex malware by using a traditional detection approach. The paper presents an extensive review of cloud-based malware detection approach and provides a vision to understand the benefit of cloud for protection of IoT, CPS from cyber-attack. This research explains advantages and disadvantages of cloud environments in detecting malware and also proposes a cloud-based malware detection framework, which uses a hybrid approach to detect malware.**

***Index Terms* — Cloud computing, cloud malware detection, cyber-physical system, malware detection.**

I. Introduction

In recent years, cyber-related attacks to the word economy have been increasing exponentially. According to Steve Morgan, cyber-attacks damage the world economy about $6 trillion in 2021 [1]. According to the researchers, these days approximately more than 1 million malicious software files are created every day [2] and the cost of the malware especially to cyber-physical systems (CPS) [3] and critical systems is rising as well. McAfee report shows that there is an outrageous increase in backdoors, banking Trojans, and fake applications for mobile devices [4].

Malware is one of the biggest threats in terms of launching a cyber-attack in the information security realm. Malware is any software such as virus, worm, rootkit, backdoors and ransomware, which performs malicious activities on the victim's machine [2] with or without consent of the system owner. Over the past decade, malware has been rising at an alarming rate and there is no well-known approach to detecting all malware in the wild. This is a new generation of malware using advanced obfuscation and packing techniques to escape from detection systems. This makes it nearly impossible to detect complex malware with a traditional approach.

Malware detection is the process of detecting the presence of malware by analyzing program executable. There have been proposed several different approaches to detect malware including traditional and advanced techniques. Traditional techniques have been used more than a decade including signature-, heuristic-, behavior-, and model checking-based detection approaches. Advanced techniques are based on various approaches including machine learning, deep learning, edge computing and cloud computing. It is known that signature-based detection approach performs well in terms of time and memory usage, but it fails to detect unknown malware. Even though heuristic, behavior, and model checking-based approaches can detect a significant portion of the malware, these approaches cannot detect some portion of zero-day malware. Deep learning and edge computing (mobile devices) based detection approaches use similar methods, which are used in the signature, heuristic, and behavior-based ones, however these approaches also fail to detect complex and zero-day malware.

Malware detection schema direction is changing from traditional to a new one. One of the most effective new detection approaches is cloud-based detection. It includes two sides – client and server on cloud computing, client submits a suspicious file over the internet, and server performs the analysis and specifies whether the given suspicious file is malware or not. During the analysis process, the server uses different detection agents to improve the performance. During the feature extraction phase strings, system calls, static and dynamic features, API traces, application traces and hybrid features are used. Recent studies present that cloud-based detection approach enhances the detection rate for known and unknown malware [5], [6] and provides more detailed analysis for each malware sample.

Cloud-based detection approach brings many advantages over traditional approaches. Cloud environment provides more computational power and much bigger databases for malware detection. Multiple execution traces of the same malware can be gathered [5]. It also improves the detection performance for personal machines, mobile devices and CPS. On the other hand, there are some drawbacks such as loss of control over data, overhead between client and server, lack of real time monitoring, and limited usage of

Submitted on February 13, 2021.
Published on March 10, 2021.
Ömer Aslan, Department of Computer Engineering, University of Siirt, Turkey.
(e-mail: omer.aslan@siirt.edu.tr)
Merve Ozkan-Okay, Department of Computer Engineering, University of Ankara, Turkey.
(e-mail: merveozkan@Ankara.edu.tr)
Deepti Gupta, Department of Computer Science, University of Texas at San Antonio, USA.
(e-mail: deepti.mrt@gmail.com).

infrastructure. This review paper presents a detailed review of cloud-based malware detection approach and makes the following contributions:

- Provides a summary of the current academic studies on cloud-based malware detection approach.
- Presents a vision to understand the benefit of cloud for protection of cyber-physical systems from malware.
- Explains the trends in creation of malware and hiding techniques.
- Discusses the current challenges and suggest new techniques for malware detection.
- Presents a cloud-based malware detection framework, which is based on signature-, behavior-, deep learning-, and heuristic-based approaches.

The rest of this paper is organized as follows. Section II describes trends in malware creation and hiding techniques. Section III explains an overview of cloud-based malware detection systems. Related work on cloud-based malware detection approach is summarized in section IV. Discussion and evaluation of cloud-based malware detection approach is presented in section V. Section VI presents the proposed framework of our approach. Finally, conclusion and future work is given in section VII.

II. Trends in Malware Creation and Hiding Technologies

Software which performs malicious actions on the victim machine is defined as malware [5]. There are different sorts of malware include: virus, worm, backdoor, rootkits and ransomware. Different types of malware, primary characteristics and well-known malware families can be seen in Table I and II. Hackers are launching cyber-related attacks by using malware, which exploits errors, vulnerabilities, and failures in the existing systems such as buffer overflow, security misconfiguration and computer networks protocols flaws. These days, since a lot of malware instances present the characteristics of multiple classes in the same time [5], the classification of malware is becoming more difficult.

Virus is the first malware type which has appeared in the wild. Timely, other types of malware that appear in the computer systems. At the beginning, malware was created for simple purposes such as hacking friends' computers or some simple financial gain. However, in the process of time, it was replaced by a complicated malware that damaged large organizations, industries, and governments assets. Malware can be categorized in two ways: Traditional and next generation malware. Traditional malware is usual malware that is easy to detect and remove from the computer systems, However, next generation malware is more destructive, difficult to detect and remove from the computer systems. Besides, next generation malware can easily bypass the protection software that is running in kernel mode and hide themselves in the computer systems.

By using next generation malware persistent and targeted cyber-attacks can be launched. During the attacks different types of malware are used. Most of the time, the next generation malware uses common obfuscation techniques to disguise itself from the detection systems. Common obfuscation techniques and their explanation can be seen in Table III. Practically it is almost impossible to detect next generation malware with a single detection approach. Thus, there is an urgent necessity to use different approaches and more computational power to detect malware.

TABLE I: Types of Malware and Primary Characteristics

Malware Types	Main Characteristics
Virus	Common and well-recognized malware
Worm	Spreads by using networks Allows unauthorized access to CPS systems
Trojan Horse	Appears to be a normal software Sends secret information to other parties
Backdoor	Bypasses security systems Opens systems to remote access
Rootkits	Provide privileged access Hide their suspicious codes from the host system
Ransomware	Encrypts the data on infected system
Obfuscated malware	Uses concealing techniques to hide itself in the systems

III. Overview of Cloud-Based Malware Detection System

Cloud computing has been rapidly emerging as a new paradigm for accessing various services including storage, compute, data management, messaging, media services, machine learning and AI, developer tools, and security. Fig. 1 shows the different types of cloud deployment models along with various services and users. Cloud computing services allow access to the data anytime, anywhere and at any device. However, this access advantage can become a serious threat, which allows easy access to malware too. Cloud malware can lead to several serious security issues including siphoning out sensitive data, login credentials, virtual machine hijack, damage systems and identity theft etc. In the cloud, malware can run inside virtual machines, and can be the cause of stealing a user's personal information. In cloud environment signature, behavior and machine-learning-based approaches have been proposed to detect malware. Dominating cloud service providers (CSPs) including Amazon Web Services (AWS), Azure, and Google [7] also provides security services to detect malware at cloud level. Amazon GuardDuty is a threat detection service, which analyzes the unauthorized behavior to detect the malware, Google also announced a new tool to detect the modern threats.

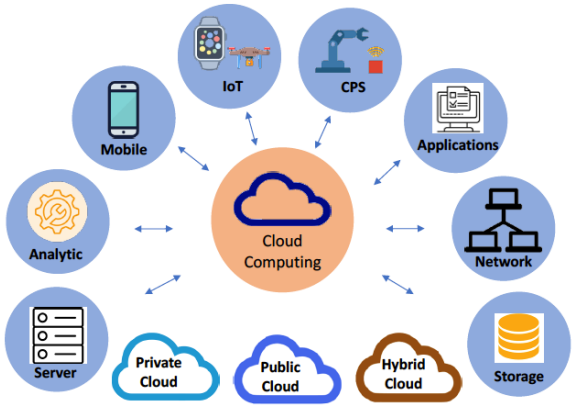

Fig. 1. Cloud Computing technologies and deployment models.

Fig. 2 shows the flow of malware detection at cloud level. Multiple users including personal computers, mobiles, and IoT receive files via email, HTTP, media, IM, and P2P etc. These users upload their files to the cloud and receive the report about the file. In cloud, signature-based detection approaches recognize malware based on the pattern comparison, these patterns are stored on the cloud.

Signature-based approach is considered quite fast and accurate for known malware, however, fails to detect new malware. Anomaly-based detection approaches detect malware based on their behavior, in addition this approach identifies a new malware, however, generates false alarm too. In the past, machine learning based detection approaches have been investigated to detect the malware, and these approaches show the promising results in the terms of performance and efficiency. Machine learning based malware detection approaches use various algorithms including decision tree, support vector machines, LSTM [8] and others. This approach works effectively only if sufficient data and computation power is available to train the models. However, machine learning based malware detection approaches also face the scalability issue.

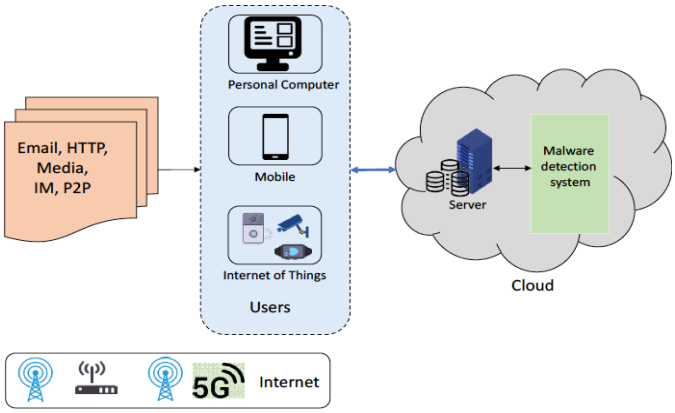

Fig. 2. Cloud-based malware detection system.

TABLE II: WELL-KNOWN MALWARE FAMILIES

Malware MD5	Malware Type	Malware Family
f3c0c179d69ea7bcede7d908ca2c4cd0	Worm	Worm.Win32, Worm.Vobfus.Gen
f3c1a983e85b56ed3c7304e7446fc510	Dropper	Trojan.Win32,TrojanDropper
f3c68eb51119fabbb0538b3ab6711c43	Adware	Generic Adware
f3cbd0045a27c7736f78788c2321d66c	Backdoor	Hacktool.Win32, Backdoor.Win32
f3d3a9cdbb62f8d9bb5e15eaa414597d	Ransomware	Trojan.Win32, Ransom:Win32
87f88e2e48c0f45ba32b9535e8a9c2ab	Injector	Trojan.Generic, W32.Trojan
54bc2102bbfa0cd23d30b086082887f3	Packed Malware	Gen:Heur.Cridex, TR/Crypt.XPACK.Gen
9085a7dff20d6a5c287d3056d3ed1cc4	Rootkit	Dropper.Generic*r .AC,Win32 : Rootkit - gen*
48cd89827939b3a8976d9bb0993bc338	Spyware	Win.Spyware.Zbot, Gen:Variant.Razy
28cb0c8083f6a41e7b04137ab166c580	Packed Malware	Gen:Packer.PESpin, Trojan.Win32.Crypt

TABLE III: COMMON OBFUSCATION TECHNIQUES

Malware Types	Main Characteristics
Encryption	It hides malicious code blocks in its entire code
Oligomorphic	It uses different keys while encrypting and decrypting
Polymorphic	It uses layered encryption and encrypted portion contains various decoder copy
Metamorphic	It hides malicious code blocks and change malware for every iteration Each version of malware is different
Stealth	It performs various hiding techniques in order to escape from the detection systems
Packaging	It compresses malware to disguise from the detection systems To analyze correctly, malware must be unpacked

IV. RELATED WORKS ON CLOUD-BASED MALWARE DETECTION APPROACH

Recently, cloud computing technology has been used in many different areas. One of the most recently used area is detecting malware. A lot of academic studies have been done on this area. Cloud-based detection approach and the methods that are used in the clouds bring many advantages over other detection approaches. Cloud-based detection methods are reviewed based on main idea, feature extraction method, and algorithm types. Before explaining each cloud-based detection method in detail, some methods in the literature are summarized in Table IV.

Ye et al. [9] propose a Valkyrie system that is based on a semi parametric classification model which combines file content and file relations together for malware detection. This system is included in Comodo's Anti-Malware products. The authors claim that their results show that the accuracy of the proposed system outperforms other popular antivirus scanners such as Kaspersky, McAfee, VirusScan and Bitfender. However, file associations and content have different properties. Therefore, directly combining these features can reduce the quality of data such as correlation and consistency.

Yan [10] proposes a malware detection technique to extract correlation signatures from advanced malware families in the cloud. The paper presents CAS, a framework for large-scale and cross-family malware analysis. CAS uses a new method for Advanced Persistent Threats (APTs) correlation. According to the author, the testing of the proposed method shows that CAS detects a high amount of malware samples effectively by using malware correlation signatures at inline speed. Advanced malware includes packers, PE malware, scripts, mobile malware, and non PE malware. To improve the study, several issues related to protecting antivirus cloud service remain to be addressed.

Masud et al. [11] propose a generalized multi-partition, multichunk ensemble technique named as EMPC. The aim of this study is to significantly reduce the expected classification error compared to the current single-partition, single-chunk ensemble method. They formulate the malware detection problem as a data stream classification problem and identify the drawbacks of traditional malicious code detection techniques over data mining approaches. They propose a solution using the cloud computing framework and applying recommended techniques to synthetically

generated data, real botnet traffic, and real malicious executables. The authors claim that their technique achieves better detection accuracy than other stream data classification techniques. The classification accuracy could potentially be improved by using supervised dimensionality reduction techniques for improved feature selection. Additionally, the run time of the proposed approach can be improved by taking advantage of the additional parallelism found in cloud computing architecture.

Sun et al. [6] proposes a cloud-based anti-malware system called CloudEyes that provides efficient and reliable security services for resource-constrained devices. CloudEyes offers suspect bucket cross-filtering, a new signature detection mechanism based on a reversible sketch structure for the cloud server. The authors claim that the obtained results show that the systems in CloudEyes are efficient and functional, and their systems can be over other current systems with less time and communication consumption. However, detection performance can be improved by combining signature and other detection algorithms.

Kumar et al. [12] propose a malware detection system based on the clustering and trend micro-locality sensitive hashing. First, they use the Cuckoo sandbox that provides dynamic analysis reports by running files in an isolated environment. Then, they propose a new feature extraction approach to extract features from the reports. The most important features are selected by using principal component analysis (PCA), random forest (RF) and Chi-square methods. According to the authors, the results obtained better accuracy, FPR and AUC than the non-clustering approach. However, the proposed approach should be tested using a larger data set.

Mirza et al. [13] provides an energy-efficient hosting model for an intelligent malware detection framework. The proposed model combines different components of Amazon's cloud services to develop a dynamically scalable hosting model. According to the authors, the test results show respectable energy efficiency in terms of CPU usage. However, the proposed model can be over enhanced by integrating the intrusion detection system which is powered by the cloud-based detection engine.

Agrawal and Wahie [14] study cloud as a provider for the advent of ‘in-the-cloud’ anti-malware service. In the paper, vulnerabilities and limitations of cloud-based antivirus detection systems have been discussed. The authors conclude that the work with a proposal for utilizing the best of cloud services with a robust form of antivirus solution instead of the ‘much strived’ for lightweight solution can further enhance the detection in the cloud. This work should be enhanced by examining more in-depth research such as data mining, machine learning, deep learning as well as real implementation for cloud-based detection systems.

A distributed malware detection system that uses an additional screening step before signature matching has been proposed in [15]. Two-stage scanning of the proposed system provides quick and memory efficient detection while reducing the amount of storage. According to the paper, the proposed SplitScreen method is implemented as an extension of ClamAV that increases scanning performance by half memory usage. In addition, as the number of signatures increases, the memory usage and run time decreases. In the proposed system, a single server is used for detection in the cloud side. Using more servers as well as putting some burdens on the client side can increase detection performance further.

Abdelselam et al. [16] proposed a malware detection approach based on the Convolutional Neural Network (CNN) for the cloud computing environment. First, they use a standard 2d CNN structure by training on the metadata which is available for each process in a virtual machine. Then, they improve classifier accuracy using a new 3d CNN that helps reduce mislabeled samples. Their experiments are applied on data collected by executing several malware types on virtual machines. They selected malware randomly in experiments.

Extensive survey paper on malware detection approaches and their methods are summarized in our previous study [5]. The paper first discusses the probability of detecting malware in theory and practice. Then, further discusses available malware detection dataset, features extractions, detection approaches and methods. Detection approaches are divided into eight categories including signature-, behavior-, heuristic-, model checking-, deep learning-, mobile-, IoT-, and cloud-based. According to paper, none of them can detect complex and intelligent malware with high percentages. However, especially behavior-, model checking, deep learning-, and cloud-based detection approaches are promising solutions among others. Combining these approaches with appropriate technologies can create more advanced detection systems.

The proposed 2d CNN model achieves an accuracy of 79%, and their 3d CNN model considerably improves the accuracy by up to 90%. This work can be advanced by increasing the scale of their tests by using more malware samples.

Yadav [17] presents a unified WFCM-AANN malware detection technique that identifies the malware in the system. The proposed work consists of clustering and classification modules. In the clustering module, the input dataset is collected in clusters using the Weighted Fuzzy C-mean clustering (MFCM) algorithm. In the classification module, the centroid from the clusters is given to the intermittent Auto-Associative Neural Network which is used to define whether the information is intruding or not. The authors claim that the proposed model identifies malware with high detection sensitivity, thus performing better than current classifiers. However, the proposed system performance can be improved further near the future.

Penning et al. [18] summarize mobile malware threats, attacks and cybercriminal motivations behind malware. They discuss in more detail current prevention methods, their limitations, and difficulties encountered in preventing malware on mobile devices. In addition, they propose a cloud based framework for mobile malware detection. The proposed framework requires a collaboration between mobile subscribers, app stores, and IT security professionals. According to authors the cloud-based malware detection is a potential approach to mobile security. This work should be developed by examining more studies and how to benefit cloud services and collaborations.

TABLE IV: SUMMARY OF RELATED WORKS ON CLOUD-BASED MALWARE DETECTION APPROACH AND IT METHODS

Paper	Proposed Method	Goal/Success	Year
Ye et al. [9]	Valkyrie system that based on a semiparametric classification	Accuracy of system outperforms other popular anti-malware software	2011
Masud et al. [11]	Multi-partition, multichunk ensemble technique named as EMPC	It achieves better detection accuracy than other stream data classification techniques	2011
Cha et al. [15]	SplitScreen uses an additional screening step before the signature matching phase	The run time and memory usage of SplitScreen decreases as the number of signatures increases	2011
Yan [10]	Extract correlation signatures from different malware families in cloud environment	It indicates that CAS can detect high amount of malware samples efficiently at inline speed	2012
Penning et al. [18]	Mobile malware detection framework using cloud	It is a promising method for mobile security	2014
Sun et al. [6]	CloudEyes, providing effective and reliable security services for devices with limited resources	It outperforms other systems with less time and communication consumption	2016
Agrawal and Wahie [14]	Cloud-based anti-malware service	It investigates the vulnerabilities and limitations of the cloud	2016
Xiao et al. [19]	Malware detection scheme with Q-learning	It increases detection accuracy, reduces the detection delay	2017
Abdelselam et al. [16]	Cloud-based detection system using CNN	The 2d CNN model achieves 79% accuracy, and 3d CNN 90%	2018
Mirza et al. [13]	Energy efficient model in the cloud environment	It demonstrates considerable energy efficiency for CPU utilization	2018
Yadav [17]	Consolidated WFCM-AANN malware detection technique	It identifies malware with high detection sensitivity, thus performing better than current classifiers	2019
Kumar et al. [12]	Detection framework based upon the concept of clustering and trend micro locality sensitive hashing	It achieves better performance results when it is compared to the non-clustering approach in terms of accuracy, FPR, and AUC	2020

Investigation of cloud-based malware detection game is presented in [19]. They derived Nash equilibrium static malware detection game which shows how mobile devices share their application traces into security servers by using access points. To get optimal offloading rate for a mobile device, Q-learning is used. In order to improve the performance and accelerate the reinforcement learning process, Dyna architecture and post-decision state learning are used. When proposed Q-learning based detection compared with the benchmark study with 100 mobile devices, detection accuracy and utility of mobile devices are increased while detection delay is decreased. In game theory, the collaborative deep learning game [20] and cooperative smart farming game [21] have been proposed to enforce IoT devices and smart farms to be part of collaborative deep learning, and cooperative smart farming respectively.

When existing studies are examined, it is observed that different data mining techniques such as feature extraction methods, preprocessing phase, and feature reduction methods have been used to create dataset. Machine learning algorithms are used on the dataset created in the previous phase to detect malware attacks. When the techniques in these researches are evaluated, it is seen that preprocessing and feature reduction before implementing machine learning methods improves the performance. Using the same techniques and methods in the cloud environment also improves the detection rate [22]. In addition, some machine learning methods can perform better than others depending on the number of features that have been used, the distribution and size of the data that have been evaluated. As a result, when scientific studies are examined, it seems that malware attacks still cannot be detected efficiently with high accuracy. Besides, the cloud-based approach is still at an early stage and there needs to be more studies to see effects of cloud computing in malware detection [23].

V. EVALUATING OF CLOUD-BASED MALWARE DETECTION APPROACH

In the literature review, cloud-based malware detection approach and its related methods are discussed. The development of cloud-based detection approach by years can be seen in Fig. 3. This figure shows the average cloud-based studies' performance by years when detecting malware. Before 2010, there were not many studies which used the cloud environment for malware detection. The average performance of the cloud-based approach is changing between 82% to 94%. The performance of the cloud-based approach's performance depends on the cloud environment resources, the algorithm, and malware datasets that are used. It can be seen that the detection rate approximately increases only 1% percent by year. This means that either there are not enough studies which use cloud environments or the methods proposed to detect malware are not efficient.

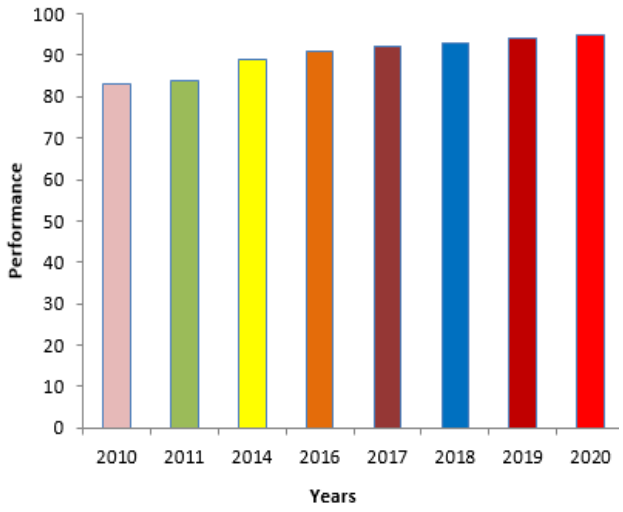

Fig. 3. Development of cloud-based detection approach performance by years.

Generally, a cloud-based approach consists of multiple detection agents which can perform signature, behavior, and heuristic approaches at the same time in the cloud. Using multiple detection agents at the same time improves the performance. The cloud environment provides much bigger database and intensive computational resources with easy installations, configurations and updating regularly. Even though the cloud-based methods can detect various forms of known and unknown malware, they cannot detect all new generations of malware. This is because code obfuscation techniques prevent malicious code blocks from being analyzed correctly [24]. This raises the false positives and negatives rate immensely. Besides, models that are used in the clouds are not resistant to evasion attacks. For instance, crafted inputs lead to deceive machine learning and deep learning models which result in classifications. There is huge demand for more scientific works to cover shortcomings of cloud based existing methods. The use of new techniques along with the use of machine learning, data mining, and deep learning algorithms [25] in the cloud environment will improve the performance.

Although each detection approach has its own advantages and performs better for different features sets, none of them successfully detect all malware. When complexity of malware increases over time, the detection rate decreases for all detection approaches [5]. It can be said that behavior, model checking, deep learning and cloud-based approaches outperform than signature, heuristic, and mobile devices-based ones. The cloud-based detection approach is still at an early stage and implementing a well algorithm in the cloud will produce better results among others.

Even though malware detection schemas are being improved day by day, the following research challenges are still remaining an open question in the cloud-based and other detection approaches:

- Obfuscating and packing techniques are used by malware to escape from detection systems. Cloud-based and other approaches are not resistant to all obfuscation techniques.
- Client needs to upload a suspicious file to the cloud which may disclose some sensitive information.
- Real-time monitoring is a challenging task. Most of the studies about malware detection are using dataset ,which are not suitable for real-time monitoring.
- Cloud-based and many other malware detection approaches are prone to false positives and negatives.
- Learning algorithms are prone to bias and overfitting. This situation decreases detection rates and increases false positives rates.
- No cloud-based detection method can effectively detect all zero-day malware.
- There are no well-known accepted datasets for everyone to evaluate the performance of cloud-based detection approaches.

VI. Proposed Framework

In this survey paper, we propose a cloud-based malware detection framework. Proposed framework can be seen in Fig. 4. The proposed framework consists of malware detection agents which uses 4 detection algorithms in different virtual machines on the cloud: Virtual machine 1 (VM1- SBD/signature based algorithm), virtual machine 2 (VM2- BBD/ behavior based algorithm), virtual machine 3 (VM3- DLBD/ deep learning based algorithm), and virtual machine 4 (VM4- HBD/ heuristic-based detection). The proposed framework works as follows:

1) Client sends suspicious file over the networks to the server.
2) Server receive the suspicious file and performs signature based detection algorithm in VM1-SBD.
3) In the same time, behavior-based and deep learning-based algorithms are running in VM2-BBD and VM3-DLBD.
4) If the signature-based detection schema can identify the suspicious file with high guess as malware or benign.
   (a) The file is marked and the result sent to the client machine immediately.
   (b) The algorithms that are running in VM2-BBD and VM3- DLBD stops and detection processes will finish
5) If the suspicious file could not be identified by using a signature based detection algorithm in VM1-SBD, the detection process continues with behavior and deep learning-based detection in VM2-BBD and VM3-DLBD.
6) If the behavior-based detection schema can identify the suspicious file with high guess as malware or benign.
   (a) The file is marked and the result sent to the client machine immediately.
   (b) The algorithms that are running in VM3-DLBD stops and the detection process will finish.
7) If the suspicious file could not be identified by using a behavior based detection algorithm in VM2-BBD, the detection process continues with deep learning based-detection in VM3- DLBD.
8) If the deep learning-based detection schema can identify the suspicious file with high guess as malware or benign.
   (a) The file is marked and the result sent to the client machine immediately.
   (b) The detection process will finish.
9) If the suspicious file could not identify by using 3 detection agents.
   (a) The properties that gathered in VM1-SBD, VM2-BBD and VM3-DLBD are combined and the heuristic algorithm makes best guesses to identify the suspicious file.
   (b) The result is sent back to the client and the detection process will complete.

For signature-based detection, usual antivirus scanner algorithm will be used. For behavior-based detection algorithm, our previous algorithm entitled subtractive center behavioral model which is explained in [2] will be used, and for deep learning-based detection algorithm sequential model which is using Windows API Calls [26] will be used. For heuristic detection, the features extracted from VM1-SBD, VM2-BBD and VM3-DLBD will be combined and based on the specified rules, suspicious file will be determined to be malware or benign. The proposed framework will run fast and efficiently. Because each detection algorithm will run on different VMs on the cloud,

and when one detection algorithm identifies a suspicious file as malware or benign with over 90% assurance, the detection process will stop.

The proposed framework is efficient because it uses a hybrid approach which consists of signature-, behavior-, deep learning-, and heuristic-based approaches. If one algorithm will miss the identified suspicious file correctly, the other algorithm will identify correctly. The proposed framework will be implemented in the next study.

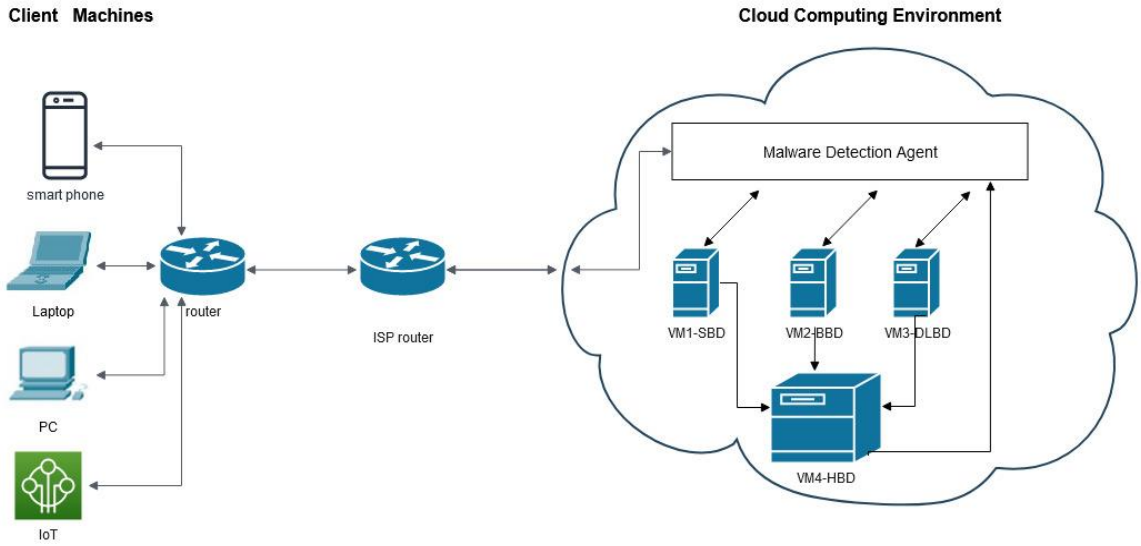

Fig. 4. Proposed framework.

VII. Conclusion and Future Work

These days, the number, sophistication and severity of malware attacks are increasing and there is no well-known approach that can detect all malware. Although a number of different methods have been proposed in various detection approaches and cloud-based, no detection method could effectively detect all new generation malware. Cloud-based detection approach brings many advantages over traditional approaches. Cloud environment provides more computational power and much bigger databases for malware detection. It also improves the detection performance for personal machines, mobile devices and CPS. However, there are some drawbacks on the cloud side such as loss of control over data, lack of real time monitoring, limited usage of infrastructure over different customers, and there is overhead between the client and server. Reducing these deficiencies on the cloud side will improve the performance.

Although the trends in malware creation techniques and detection approaches are changing in time, this survey paper still can be considered as one of the key references for the computer scientist and developers who work in this field. It can be concluded that building an effective approach to detect malware is a very challenging task and cloud-based approach is still at an early stage. Applying different detection approaches as well as using block chain and big data in the cloud environment will be promising for the future.

References

[1] Steve Morgan, “cybersecurity almanac: 100 facts, figures, predictions and statistics,” Cybercrime Magazine Cisco and Cybersecurity Ventures, 2019.

[2] Ömer Aslan, Refik Samet, and Ömer Özgür Tanrıöver, “Using a Subtractive Center Behavioral Model to Detect Malware,” Security and Communication Networks 2020, 2020.

[3] Ajeet Singh and Anurag Jain, “Study of cyber-attacks on cyber-physical system,” In Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT). 26–27, 2018.

[4] R Samani and G Davis, “McAfee Mobile Threat Report Q1,” 2019. https://www.mcafee.com/enterprise/en-us/assets/reports/rpmobile-threat-report-2019.pdf.

[5] Ömer Aslan and Refik Samet, “A comprehensive review on malware detection approaches,” IEEE Access 8, 6249–6271, 2020.

[6] Hao Sun, Xiaofeng Wang, Rajkumar Buyya, and Jinshu Su, “CloudEyes: Cloud-based malware detection with reversible sketch for resource-constrained internet of things (IoT) devices,” Software: Practice and Experience 47(3), 421–441, 2017.

[7] Deepti Gupta, Smriti Bhatt, Maanak Gupta, Olumide Kayode, and Ali Saman Tosun, “Access control model for google cloud iot. In 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity),” IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). IEEE, 198–208, 2020.

[8] Olumide Kayode, Deepti Gupta, and Ali Saman Tosun, “Towards a distributed estimator in smart home environment,” In 2020 IEEE 6th World Forum on Internet of Things (WF-IoT). IEEE, 1–6, 2020.

[9] Yanfang Ye, Tao Li, Shenghuo Zhu, Weiwei Zhuang, Egemen Tas, Umesh Gupta, and Melih Abdulhayoglu, “Combining file content and file relations for cloud based malware detection,” In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. 222–230, 2011.

[10] Wei Yan, “CAS: A framework of online detecting advance malware families for cloud-based security,” In 2012 1st IEEE International Conference on Communications in China (ICCC). IEEE, 220–225, 2012.

[11] Mohammad M Masud, Tahseen M Al-Khateeb, Kevin W Hamlen, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham, “Cloud-based malware detection for evolving data streams,” ACM transactions on management information systems (TMIS) 2(3), 1–27, 2011.

[12] Rahul Kumar, Kamalakanta Sethi, Nishant Prajapati, Rashmi Ranjan Rout, and Padmalochan Ber, “Machine Learning based Malware Detection in Cloud Environment using Clustering Approach,” In 2020

11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 1–7, 2020.
[13] Qublai K Ali Mirza, Irfan Awan, and Muhammad Younas, "A Cloud-Based Energy Efficient Hosting Model for Malware Detection Framework," In 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 1–6, 2018.
[14] Aditya Agrawal and Karan Wahie, "Analyzing and optimizing cloud-based antivirus paradigm," In 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH). IEEE, 203–207, 2016.
[15] Sang Kil Cha, Iulian Moraru, Jiyong Jang, John Truelove, David Brumley, and David G Andersen, "SplitScreen: Enabling efficient, distributed malware detection," Journal of Communications and Networks 13(2), 187–200, 2011.
[16] Mahmoud Abdelsalam, Ram Krishnan, Yufei Huang, and Ravi Sandhu, "Malware detection in cloud infrastructures using convolutional neural networks," In 2018 IEEE 11th International Conference on Cloud Computing (CLOUD). IEEE, 162–169, 2018.
[17] Ram Mahesh Yadav,"Effective analysis of malware detection in cloud computing," Computers & Security 83, 14–21, 2019.
[18] Nicholas Penning, Michael Hoffman, Jason Nikolai, and Yong Wang, "Mobile malware security challeges and cloud-based detection," In 2014 International Conference on Collaboration Technologies and Systems (CTS). IEEE, 181–188, 2014.
[19] Liang Xiao, Yanda Li, Xueli Huang, and XiaoJiang Du, "Cloud-based malware detection game for mobile devices with offloading," IEEE Transactions on Mobile Computing 16(10), 2742–2750, 2017.
[20] Deepti Gupta, Olumide Kayode, Smriti Bhatt, Maanak Gupta, and Ali Saman Tosun, "Learner's Dilemma: IoT Devices Training Strategies in Collaborative Deep Learning," In 2020 IEEE 6th World Forum on Internet of Things (WF-IoT). IEEE, 1–6, 2020.
[21] Deepti Gupta, Paras Bhatt, and Smriti Bhatt, "A Game Theoretic Analysis for Cooperative Smart Farming," arXiv preprint arXiv:2011.11098, 2020.
[22] Deyannis, D., Papadogiannaki, E., Kalivianakis, G., Vasiliadis, G., & Ioannidis, S. "Trustav: Practical and privacy preserving malware analysis in the cloud," In Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy (pp. 39-48), 2020.
[23] Mishra, P., Aggarwal, P., Vidyarthi, A., Singh, P., Khan, B., Alhelou, H. H., & Siano, P. "VMShield: Memory Introspection-based Malware Detection to Secure Cloud-based Services against Stealthy Attacks," IEEE Transactions on Industrial Informatics, 2021.
[24] Ömer Aslan and Refik Samet. "Investigation of possibilities to detect malware using existing tools." IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 2017.
[25] Ren, Z., Wu, H., Ning, Q., Hussain, I., & Chen, B. "End-to-end malware detection for android IoT devices using deep learning," Ad Hoc Networks, 101, 102098, 2020.
[26] Yazı A. F. Elezaj O. Ahmed J Catak, F. O, "Deep learning based Sequential model for malware analysis using Windows exe API Calls," PeerJ Computer Science 6, e285, 2020.

ÖMER ASLAN is a Dr. research assistant in computer engineering department at the university of Siirt, Turkey. He received his PhD in cyber security field 2020 from university of Ankara, Turkey, MSc in information security field in 2014 from university of Texas at San Antonio, United States of America, and BSc in computer engineering department in 2009 at university of Trakya, Turkey. He is working on computer systems, information security, cyber security, malware analysis and cloud computing. He has published several papers on international journals and conferences.

Merve Ozkan-Okay received B.S. and M.S. degree in computer engineering from Ankara University, in 2014 and 2016 respectively. She is a Research Assistant and doing Ph.D. in Department of Computer Engineering, Ankara University. Her current research interests include cyber security, cloud-based systems, machine learning and image processing. She has published several papers on international journals and conferences.

Deepti Gupta is pursuing Ph.D. in computer science at University of Texas at San Antonio (UTSA). She received B.S. and M.S. degree in mathematics from Chaudhary Charan Singh University, India, the M.Tech. degree in computer engineering from Shobhit University, India and the M.S. degree in computer science from The University of Texas at San Antonio (UTSA). She has worked as an Adjunct Faculty in Department of Computer Science at St. Edward University, Austin. Her primary area of research includes security and privacy in cloud computing and Internet of Things, security models, and deep learning. She has also served as reviewer and committee member in conferences.