PCP Framework to Expose Malware in Devices

Rahul Nigam
Dept. of Computer Science and Engineering
Centre for Advanced Studies
Dr. A.P.J. Abdul Kalam Technical University
Lucknow, Uttar Pradesh-226031
rnigam392@gmail.com

Rohit Kumar Pathak
Dept. of Computer Science and Engineering
Centre for Advanced Studies
Dr. A.P.J. Abdul Kalam Technical University
Lucknow, Uttar Pradesh-226031
rohit18197@gmail.com

Arun Kumar
Dept. of Computer Science and Engineering
Centre for Advanced Studies
Dr. A.P.J. Abdul Kalam Technical University
Lucknow, Uttar Pradesh-226031
drarun@cas.res.in

Dr. Shiv Prakash
Dept. of Computer Science and Engineering
JK Institute of Applied Physics and
Technology, Prayagraj – 211002
shivprakash.cse@gmail.com

***Abstract*—Cybersecurity is the biggest threat to the world economy. Nowadays, the internet is the main cause of increase the cybercrime and help the attacker to act to harm the victim system. The attacks and exploits are becoming more sophisticated and harder to detect. The attacker uses new techniques and processes to steal important information from the system. A cyber attacker creates malware to damage system and gains unauthorized access. The signature-based and pattern analysis of malicious code is not effective and efficient for malware detection. In this paper, proposed a new framework to detect the malware in devices using the combination of blockchain technology and machine learning algorithm. The main objective of this research is to improve the false-positive and false-negative rate to increase the accuracy of malware detection in devices. The other significance of this paper is to identify the new type of malware which cannot be identified yet using the signature method and pattern method. The experimental result shows better accuracy, precision for detecting the malware.**

***Index Terms*—blockchain, malware, ethereum**

I. INTRODUCTION

In recent days, the security of the mobile, computer, network devices are the major issue faced by the organization. Attackers use a different type of modern technology to attack the victim system that can't be detected easily. The attacker uses the malicious code to inject it and uses the system to steal all personal information of the user. The malware is the biggest threat in security and cannot easily be detected by t h e antivirus [1]. The malware is a malicious code or software which can harm the system, network or mobile devices, browser, etc. The uses of mobile devices are effectively increasing the problem of getting unsecured applications and data over the internet [2]. The vulnerable application also creates a problem with the devices by installing them on mobile devices and system. It happens due to a lack of design, codes, and other weak flaws. Different types of malware cause security incidents on the system, mobile devices, and network. The broad Adaptively of Android devices and their ability to get to critical and classified data has brought about these devices being focused by malware engineers.

Existing Android malware examination systems can be extensively ordered into static and dynamic analysis [3]. Vulnerability scoring systems have many disadvantages and do not provide an exact view of the risks related to software vulnerabilities [4]. The vulnerability and malware over a network have a greater impact on the mobile devices, network and create a thread against security [5].

In this paper, the new blockchain framework was proposed that helps to detect the malicious activity and malware in the devices which can occur from any cause of reason. It helps to detect the new type of malware in devices and also improve the accuracy of the proposed method from the previously existing techniques.

The blockchain is a decentralized, distributed public ledger that helps for the transaction [6]. It contains the chain of the blocks which are interconnected to each other, each block contains address, data, timestamp, and nonce. All are in the hash format which can be more secure as compared to the normal text. The Hash value is irreversible and if the attacker gets some hash value it cannot regenerate the text easier to get information. The machine learning classifying algorithm helps to classify the data based on training and testing of data [7]. Many classification algorithms exist in machine learning like SVM, Naive Bayes, KNN classifier etc, which help to classify the data from the dataset. Training and testing of define classifying algorithm model for prediction of any malicious information detection with the higher accuracy and precision [8].

A. Blockchain

Blockchain is decentralized, distributed public ledger that used to record the transaction across different Systems. Blockchain is a sort of distributed record for keeping up a lasting and carefully designed value-based information [9]. Blockchain limits as a decentralized database that is supervised by PCs having a spot with distributed peer to peer

(P2P) organize [10]. Every one of the PCs in the dispersed system keeps up a duplicate of the record to anticipate a solitary purpose of disappointment(SPOF). All duplicates are refreshed and approved simultaneously. A blockchain record comprises two kinds of records, singular exchanges and squares [2]. The principal square comprises a header and information that relates to exchanges occurring inside a set timeframe. The square's timestamp is utilized to help make an alphanumeric string called a hash.

1) Types of Blockchain: An public blockchain [1], which can be accessed by the end-user. There is no access restriction. Anyone can send the transaction and also validate the transaction over the internet. The private blockchain is restricted to a few people. The validation is done within the organization. The validation is done from a specific block.

Consortium or Federated Blockchain. This kind of blockchain attempts to expel the sole self-rule which gets vested in only one element by utilizing private blockchains. This block contains its data information and a hash pointer of the previous block in the blockchain [11]. Given blockchain capacities dependent on the check of a hash and advanced marks. Each square contains a block header with the quantity of the square, a timestamp of the exchange just as the hash of the past square which contains the nonce [1]. Based on it, can develop certificate stores, digital property protection and many other applications.

The hybrid blockchain is a combination of a private and public blockchain, it is flexible that the multiple public blocks join the private block on the network. Every blockchain has its profit and loss and many consensus algorithms like proof of work, proof of stake etc. used for the making the block and also help in validating the node across the network

2) Ethereum: Ethereum is an open-source, public, decentralized platform for blockchain and money based type of application. It uses the Proof of work. Each ethereum account consists of some transaction value known as ether, which can be communicated between two accounts [12]. It acts as a third party between the two blocks for the transaction.

3) Ganache: Ganache helps to set up the ethereum blockchain to test the smart contract transaction over the network. The metal mask is a browser extension tool for ethereum for the transaction. It acts as a bridge that allows visiting the distributed web in the browser [13].

4) Malware: Malware is a type of program or software that harm the computer, server, client or Network. During the last 10 years [14], mobile devices technology have grown rapidly due to the daily increase in the number of users and facilities, give a synopsis of portable malware including Trojans, Back doors, Ransomware, Botnets, and Spyware [15].

5) Zeus: It is a banking Trojan [16] utilizes keystroke logging to bargain victim credentials when the person visits a Banking site.

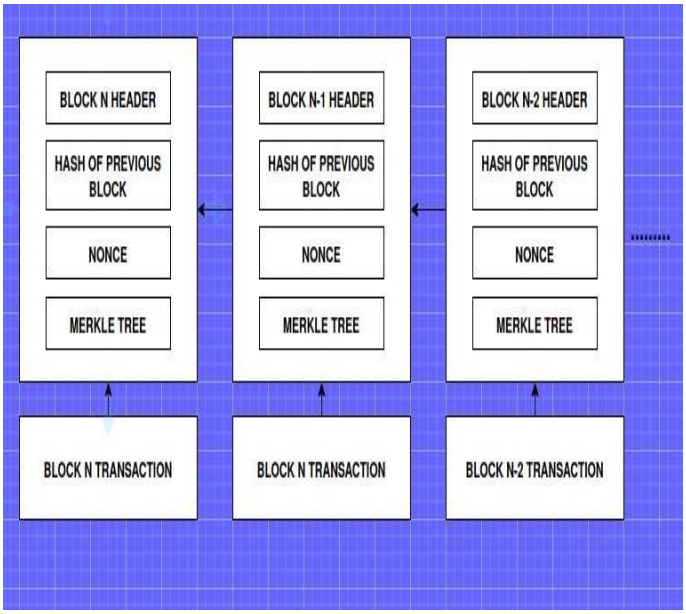

Fig. 1. Blockchain Architecture

6) Emotet: It is a type of malware. It tends to be conveyed through either malicious download attachments or links.

7) Nanocore: NanoCore is a Remote Access Trojan spread utilizing malspam as a nasty Excel XLS spreadsheet.

8) CoinMiner: It is a cryptocurrency miner where it spreads through other malware.

This section introduces the malware, blockchain, and give a brief introduction about the topic which uses in this paper, section 2 contains information about the Design part of the proposed model, section 3 contains information about the Implementation part and section 4 contains information about the Result and comparison part, section 5 contains information about the Conclusion and future work, while last part of the section has the reference information. The significance of each section has its importance and own work which can be summarized in the paragraph.

II. METHODOLOGY

Implementation of a new framework and uses the dataset for the comparison also finds the key feature related to blockchain technology which can help in getting the result and compare it with the others. The different sections in Methodology and research design define its method and proper design of the solution of the thesis. The key features must be taken are Address, Signature,data-value, timestamp, etc.
The process of malware detection firstly, get the malware

dataset then extract the feature using the static and dynamic analysis and after extracting the important feature from the dataset and then doing the feature selection from the dataset and then apply a machine learning classifier to classify the malware and non-malware based on training and testing of the dataset [17]. Testing dataset helps to predict and classify the malware based on the training of classifying algorithms. Training helps to create a model for predicting malware.

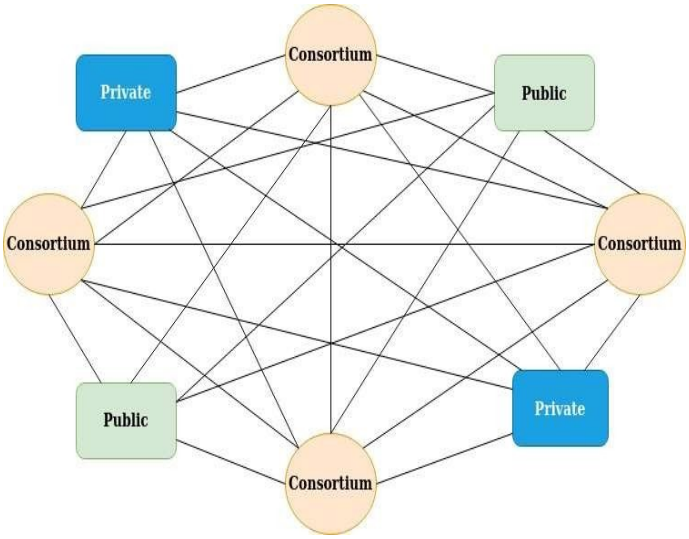

Fig. 2. Blockchain Framework

Fig 2 shows the blockchain framework that defines CB as a consortium blockchain, the PB define as a private blockchain, the public is a pubic blockchain the nodes are interconnected to each other, the validation and verification should be done for the devices based on consensus algorithm. Types of blockchain helpful for an efficient way to detection of an attack and malicious activity happen then these help in securing the transaction.

Blockchain PCP (Public Connect Proof) framework, a public blockchain which is accessed by any person, anyone sends the transaction and also acts as a validator node. While the private blockchain can access within the organization and have restrictions within the network. The consortium blockchain works across different organizations. The consortium acts as a validator node while private and public blockchain is work as a member node for the transaction. The combination of these types of blockchain help to validate the block if there is any suspicious activity happening in any block then other connected blocks get the alarm and information and get ready for any attack. The hash value data which can be encoded in the block are more secure, the decryption of the hash value cannot be easily possible.

Fig. 3 shows the process to detect the malware in devices, using the dataset which has the following information like hash value, the virtual size etc. The dataset is collected from

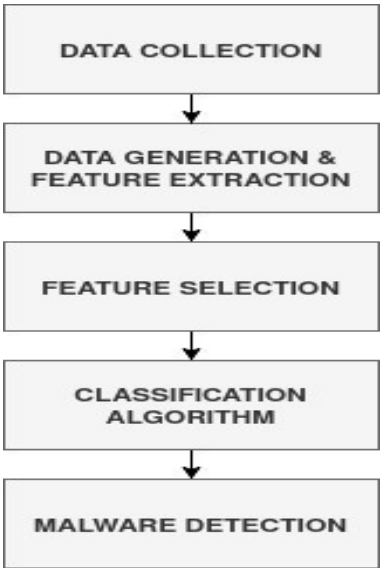

Fig. 3. Detection Process

the Kaggle. Extracting the feature from the dataset using the tools. After extracting important features from dataset then feature a selection of the malware taking the column having information like hash value, virtual address and virtual size. The machine learning classification algorithm implemented to classify the malware node or non-malware based on the feature of the dataset. 70 % of the dataset to train the model for the dataset and rest 30 % are used for testing of the model for prediction and detection of malware.

The proposed design for the implementation of the new blockchain framework for malware detection in devices uses the combination of Consortium, public and private blockchain all has its importance and its functionality. Using these types help to detect using validator and member nodes. The working of the proposed framework is as, there are a validator and member node, Validator node work is to initiate, receive transaction as well as validate the transaction while member node only initializes or receive a transaction.

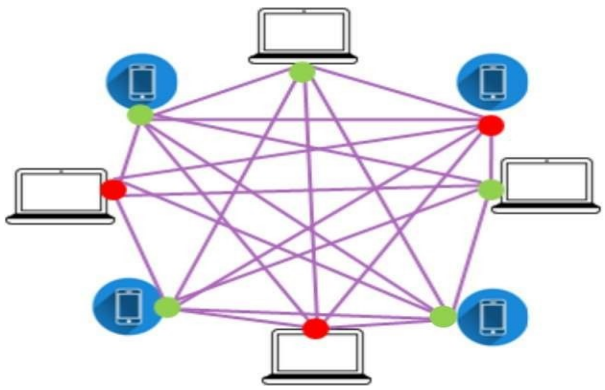

Fig. 4. Transaction among devices

The Fig.4 shows the working of the node on the blockchain network, the PC, Mobile are interconnected, there is a green circle as a member node and red circle as a validator node, the validator node PC and mobile users to validate the transaction across the network among the other member PC and a mobile node. The member node consists of the transaction value and it's validated by the validator node when there is any transaction happening between the nodes across the network.

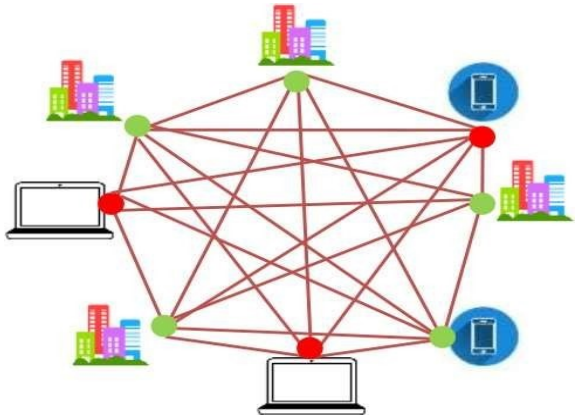

Fig. 5. Working of Nodes within the organization

The Fig.5 shows the working of the node on the blockchain network, the PC, Mobile, organization are interconnected, there is a green circle as a member node and red circle as a validator node, the validator node PC and mobile users to validate the transaction across the network among the other member like organization, PC and a mobile node. The organization works as a consortium blockchain in the network that acts as a private as well as public block.

III. RESULTS

In this, the block has to be implemented and create a blockchain for the transaction. A block can be created with a unique hash value and all the blocks are interrelated to each other and initial '0000' indicates that it uses POW. It would like to give the transaction address to the nearest node and checking the validity of connecting chain code is valid or not.

The detection of the malware evaluates based on different metrics like false- positive, True-positive, False-negative, True-negative. The result obtained from the experiment based on parameters. By the help of the confusion matrix obtained from the result, plot the ROC curve, to find out the accuracy, precision, and Recall values for detection of malware as compared to other models.

Fig.6 shows the mining of the block among the transaction. The block is created using the hash algorithm which has

```
Trying to Mine block 1...
Block Mined!!! : 000007fbe2a0af47f6554a4fc208f155c6d47d7016a5b8c076f6db220e40d7d8
Trying to Mine block 2...
Block Mined!!! : 00000d0ed3a5bda132fc48b12a75a5058c728b4d45ec20564a838f7db2a24873
Trying to Mine block 3...
Block Mined!!! : 00000ef730f0785018691f41e35c13b142b5f24079fb17fa620d74bde58898e4

Blockchain is Valid: true
```

Fig. 6. Block Mining

hash value, nonce and timestamp and information, each new block verify first and then add to the previously created block and create a chain of blocks. The following results show the validating new transaction and record the data on the blockchain. It shows the POW of the block in the network.

The detection of the malware evaluates based on different metrics like false-positive, True-positive, False-negative, True-negative. The result obtained from the experiment based on parameters to plot the ROC curve. To find out the accuracy, precision, and Recall to get the detection of malware.

For the finding of the false-positive, true-positive, false-negative, false-positive ratio apply the graph-based approach and make the following result obtained. The following help to find the curve area under the curve using the dataset.

There is a confusion matrix generated from the given dataset which is taken into two parts for result generation. The first dataset, identify the 340 nodes as a malware, and the other dataset identifies the 331 nodes as malware and non-malware also identifying efficiently from the dataset.

TRUE-POSITIVE	FALSE-NEGATIVE
331	4
FALSE-POSITIVE	**TRUE-NEGATIVE**
11	54

Fig. 7. Confusion Matrix for dataset 1

Fig.7shows, the find of the false-positive, true- positive, false- negative,false-positive ratio apply the graph-based approach and make the following result obtained. The following help to find the curve area under the curve using the dataset. There is

a confusion matrix generated from the given dataset which is taken into two parts for result generation. The first dataset, identify the 340 nodes as malware, and the other dataset identifies the
331 node as malware and non-malware also identifying efficiently from the dataset.

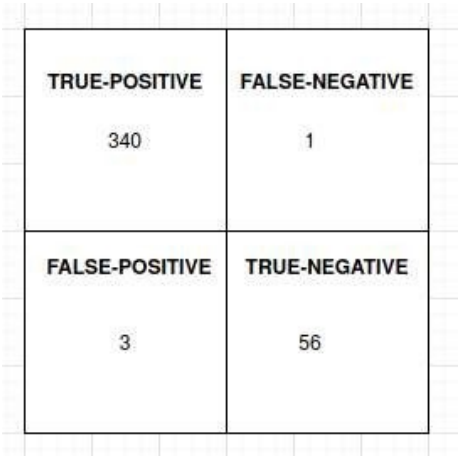

Fig. 8. Confusion Matrix for dataset 2

Fig.8 shows the confusion matrix generated from the dataset 2, it shows that our proposed model successfully identify 340 the malicious node which is a malware, and 56 non-malicious nodes also identify that it is a non-malicious node and rest 4 node does not predict between malicious and non-malicious nodes.

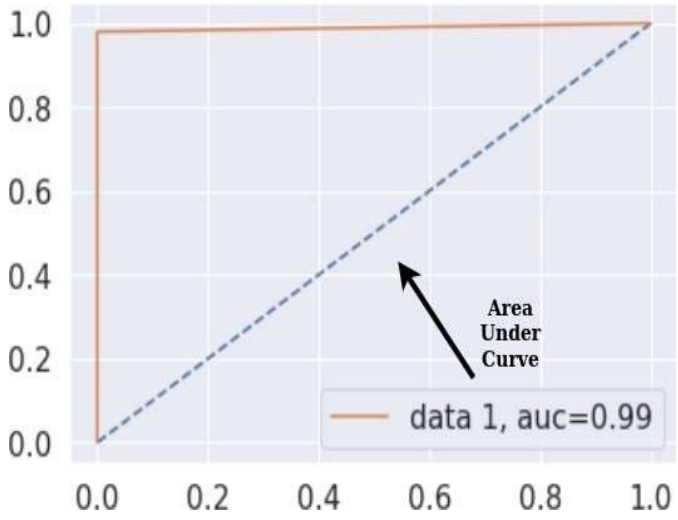

Fig. 9. ROC Curve

Fig 9 shows the Analysis, based on accuracy, Recall and a precision factor which can be obtained from ROC Curve draw based on false-positive, false-negative ratio, By value of the true positive and true negative ratio and value to the finding, the accuracy, Recall and precision and compare it with the previous existing technique. The result obtained from the given method gives a more accurate result and has a better detection factor to find the malware in devices as compared to the previously existing method. The proposed method has 0.9775 accuracies, and 0.963 Precision and recall be 0.957.

TABLE I
COMPARISION TABLE

Parameters	Proposed Method	Decision Tree	Logistic Regression
Accuracy	0.9775	0.960	0.928
Precision	0.963	0.937	0.913
Recall	0.957	0.935	0.909

Table 1 shows, comparison table defines the comparison of the different parameters based on the previous existing machine learning algorithm method used in different papers for malware detection with the proposed method defined in this paper. The Comparative analysis defines that the proposed method has better accuracy, Precision and Recall factor from the Previous existing method.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, Proposed a new blockchain framework PCP (public connect proof) to detect the malware in the devices using the blockchain technology. Classify the different malware features and detecting malicious activity on the block. The malware analysis dataset taken from Kaggle used to detect the malware detection, the combination of the blockchain technology and machine learning classifying algorithm helps to identify the malware. Utilizing this technique can label all addresses node as malicious or non-malicious and have to identify the malicious behavior based on feature defined in dataset. It is concluded that get better results and better efficiency to find out the malware detection. The accuracy and precision and recall value of the proposed method is better from the previous methods. The new type of malware also is detected easily from this process. In future, Large datasets will be used for malware detection and also add more features related to malware and blockchain and improve the accuracy and precision value of malware detection.

REFERENCES

[1] D. Evices, "M ALWARE D ETECTION T ECHNIQUES F OR M OBILE," vol. 7, no. 4, pp. 1–10, 2017.

[2] B. Anderson, D. Quist, J. Neil, C. Storlie, and T. Lane, "Graph-based malware detection using dynamic analysis Graph-based malware detection using dynamic analysis," no. February 2014, 2011, doi: 10.1007/s11416-011-0152-x.

[3] K. R. Machine and E. Engineering, "This is a repository copy of Machine learning aided Android malware classification . White Rose Research Online URL for this paper : Version : Accepted Version Article : Article available under the terms of the CC-BY-NC-ND licence Machine learning aided malware classification of Android applications," 2017.

[4] M. Petraityte, A. Dehghantanha, and G. Epiphaniou, "A Model for Android and iOS Applications Risk Calculation : CVSS Analysis and Enhancement Using Case-Control Studies," vol. 70, 2018.

[5] A. Cimitile, F. Mercaldo, F. Martinelli, V. Nardone, A. Santone, and G. Vaglini, "Model Checking for Mobile Android Malware Evolution."

[6] J. Gu, B. Sun, X. Du, and S. Member, "Consortium Blockchain-Based Malware Detection in Mobile Devices," *IEEE Access*, vol. 6, pp. 12118–12128, 2018, doi: 10.1109/ACCESS.2018.2805783.

[7] S. Y. Yerima, "Longitudinal performance analysis of machine learning based Android malware detectors," 2019.

[8] S. Rana and A. H. Sung, "Malware Analysis on Android Using

Supervised Machine Learning Techniques Malware Analysis on Android Using Supervised Machine Learning Techniques," vol. 7, pp. 178–188, 2018, doi: 10.17706/ijcce.2018.7.4.178-188.

[9] J. Moubarak and E. Filiol, "Developing a K-ary malware using Blockchain," *NOMS 2018 - 2018 IEEE/IFIP Netw. Oper. Manag. Symp.*, pp. 1–4, doi: 10.1109/NOMS.2018.8406331.

[10] S. Homayoun, A. Dehghantanha, K. R. Choo, and S. Antonio, "A Blockchain-based Framework for Detecting Malicious Mobile Applications in App Stores," no. Ccece, 2019.

[11] I. Lin and T. Liao, "A Survey of Blockchain Security Issues and Challenges," vol. 19, no. 5, pp. 653–659, 2017, doi: 10.6633/IJNS.201709.19(5).01.

[12] N. Elisa, L. Yang, F. Chao, Y. Cao, and N. Elisa, "A framework of blockchain-based secure and privacy-preserving E-government system," *Wirel. Networks*, vol. 0, 2018, doi: 10.1007/s11276-018-1883-0.

[13] A. Tseng, Y. Chen, Y. Kao, and T. Lin, "Deep Learning for Ransomware Detection."

[14] S. Saad, W. Briguglio, and H. Elmiligi, "The Curious Case of Machine Learning In Malware Detection."

[15] J. Hu, L. Yeh, S. Liao, and C. Yang, "Autonomous and malware-proof blockchain-based firmware update platform with efficient batch verification for Internet of Things," *Comput. & Secur.*, vol. 86, no. 2019, pp. 238–252, 2020, doi: 10.1016/j.cose.2019.06.008.

[16] M. Turkanović, M. Hölbl, and K. Košič, "EduCTX: A Blockchain-Based Higher Education Credit Platform," vol. 6, 2018, doi: 10.1109/ACCESS.2018.2789929.

[17] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An Overview of Blockchain Technology : Architecture , Consensus , and Future Trends," 2017, doi: 10.1109/BigDataCongress.2017.85.