DETECTION OF MALICIOUS ACTIVITIES (DOMAIN, CODE, URL, USERS BEHAVIOUR ANALYSIS) BY USING J48

Ms.R. S. Maldhure, [2]Dr.Mrs. S. S. Sherekar, [3]Dr.V.M.Thakare

[1]Student ME, [2]Professor, [3]Professor

[1] PG Department of Computer Science and Engineering,
[1]SGBAU, Amravati, India

Abstract : Malware is malicious software consist of malignant program code which are synchronized by attackers. It has becomes essential task for detecting and analysing malicious activity for domain, code ,URL, User behavior in large scale social network.Research work have discovered around 1million new malware samples per quarter. Hence the paper has proposed the framework to combine "runtime behavior" with "static structures" to detect malware variants.Committing financial fraud and propagating malware and spam are very common criminal actions for people engaged in accessing uniform resource locators (URLs) and online social networks (OSN).
OSNs also open the door for harmful activities and behaviors. URLs are malicious which redirects users to phishing (malicious) websites Thus to stop such activity a spam and malicious URLs detection system is required by removing spam content and malicious URLs in Email. This paper proposed methodology for "Detection of malicious activities (domain, code, URL, users behaviour analysis) by using J48".To solve the problem, this paper proposed the integrated machine learning methods and framework consisting of J48 in detecting the malware accurately**.**

IndexTerms - malicious domain, malware, user behaviors, malicious URLS, malicious activity, spam detection, Network Service Provider, social network,J48.

I. INTRODUCTION

The world wide web and information technologies route new path for e-commerce. But at the same time they also create opportunities for hackers or attacker[1].The standard detection method consists of identifying malware by searching some patterns in the code (a signature)an Android malware detection system that combines "static logic structures" and "dynamic runtime information"[1].By using supervised classification, which improves the systems accuracy and detects more amounts of spam and malicious URLs, There are two types of in-device malware detection systems. The first one is to perform static malware detection[2].This type of systems uses static such as API calling information and control flow graphs to generate signatures for detection[2].Today one of the major challenges facing such networks is the creation of false online identities, most commonly known as Sybil accounts where a user can create a large number of fake accounts and use them for malicious purposes. CAPTCHA, is the use of an automated system to verify if a request is from a real user[3].The Domain Name System is a hierarchical naming system for computers, services, or any resource connected to the Internet. For detection of the Malicious Domains processthe network traffic, particularly DNS traffic. And then analyze all DNS requests and match the query with the blacklist[4]. A number of online social network activities have greatly expanded in both scope and volume, opening new opportunities for public exposure In addition social networks consists of context-sensitive and relational data while also including a considerable amount of malicious content[5].

This paper has proposed system to implement the integrated machine learning methods consist of J48 for detecting the malware accurately. This proposed method can gives improved accuracy including discrete and continuous attributes handled by this algorithm.

II. BACKGROUND

Recent research on malware malicious activity detection has proposed numbers of approaches that leverage the various activity, also discuss the intent interface and binder mechanism, which are important knowledge needed to designour interception techniques.
To circumvent detection and to quickly deploy malware, hackers usually do not develop new malware from scratch, but rather improve existing logic or add new malicious logic into existing malware[1].They also repackage malware using disassembled tools to disassemble a benign app, and inject it with malicious logic, then repackage it as a new but malicious app. And call a set of malware with similar logic as a malware family[1]. Georgios Paliouras et al., have presented learning method to filter spam email. The two machine learning algorithm are considered for anti-spam filtering such as Naïve Bayesian and Memory based learning approach and they are compared concerning performance. So, that in both methods spam filtering accuracy has improved and keyword based filter are used widely for email [2].
Currently, research is being conducted on the development ofdifferent mechanisms to detect malicious activities in an OSN[3].Three main feature depending on the strategy leveraged against such malicious behaviors i.euser-generated content,social graph connections, and user profile activities[3]. To detect malicious domains, previous approaches makeuse of passive DNS analysis, active DNS probing, and WHOIS information DNS traffic collected from a number of ISP networks with the aim of detecting

malicious Fast-Flux services [4].Detecting Malicious Links in Online Social Networks through User Behavior will mainly focus on their features in relation to malicious URL detection.[5].

The subsequent structure of this paper is as follows:

Section I Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V**discusses attributes and parameters and how these are affected on detection of malicious activity. **Section VI** proposed method and outcome result possible. Finally **section VII** Conclude this review paper

III. PREVIOUS WORK DONE

In research literature, many features and classification technique have been studied to provide various detection of malicious activities in social network and improve the performance in terms of fault prediction, effectiveness, accuracy ,utility of the proposed system and detect more amount of spam and malicious URL,s.

Mingshen Sun et al. (2017) [1] has proposed a framework to combine "runtime behavior" with "static structures" to detect malware variants. Present the design and implementation of MONET, which has a client and a backend server module. The client app consists of three main components, (1) SBG generator, (2) runtime information collector and (3) RBG and SSS generator

Sunil B. Rathod et al. (2015) [2], have presented learning method to filter spam email. The two machine learning algorithm are considered for anti-spam filtering such as Naïve Bayesian and decision tree approach and they are compared concerning performance. So, that in both methods spam filtering accuracy has improved.

Muhammad Al-Qurishi et al. (2018) [3] presents an integrated social media content analysis platform that leverages three levels of features, i.e., user-generated content, social graph connections, and user profile activities, to analyze and detect anomalous behaviors that deviate significantly From the norm in large-scale social networks

Ibrahim Ghafiret al. (2015) [4] have worked on process of the network traffic, particularly DNS traffic. And analyzes all DNS requests and match the query with the blacklist. The blacklist of malicious domains is automatically updated each day and the detection is in the realtime**.**And also applied methodology on campus live traffic and showed that it can detect malicious domain connections in the real time**.**

Bandar Alghamdi et al. (2016) [5] have proposed research aims to understand the state of literature on detecting malicious URLs in OSNs, with a focus on two major aspects: URL and OSN objects.

IV. EXISTING METHODOLOGIES

Many approaches have been implemented for detection of malicious activity over the last several decades. There are different methodologies that are implemented for different malicious detection activity in social network.

A. *User Oriented Behavior-Based Malware Variants Detection System by using MONET*:

MONET consists of a client module and a backend server module. Android malware detection system
that combines "static logic structures" and "dynamic runtime information" MONET uses the following four steps to extract runtime information to perform malware detection: (1) static behavior graph generation, (2) runtime information collection, (3) runtime behavior signature generation, and (4) signature detection. In this method Malware detection algorithm will execute to check whether given suspicious app is malware or not by using some detection algorithm like graph decoupling, malware signature generation and signature matching[1].

B. *Evaluation of Content Based Spam and Malicious URL Detection in E-mail by using Bayesian Classifier and decision tree*:

Naïve bayes classifier is statistical classifier famous for Email filtering, Spam emails are identified by classification method. Bayesian Classifier use following method for classification.

1) Phishtank Dataset and DMOZ : Dataset Phishtank is source of blacklisted phishing URLs which admits user input and they are verified by users.2) URL Preprocessing: IP addresses and hexadecimal characters are used to hide the actual URLs. b) Hexadecimal Character the URL can also be represented using hexadecimal base values with a '%' symbol. It may represent any special characters Spoof guard identified the '@' and '-' symbol most
Prominent in phishing URLs.3) Performance Measurement: As combination classification model builds of Bayesian andDecision Tree C4.5, It is essential to derive performance on the basis of parameters such as Accuracy classified Error ,precisionand Recall are evaluated

$$\text{Accuracy} = (TN + TP) / (TN + TP + FN + FP)$$

$$\text{Error} = 100- (\text{Accuracy})$$

$$\text{Precision} = (TP) / (TP+FP)$$

$$\text{Recall} = (TP) / (TP + FN)$$

also used data mining approach like supervised classification which improvesthe systems accuracy[2].

C. *Analysis of User Behavior to Identity Malicious Activities in Large-Scale Social Networks*:

Proposed system is based on multiple layers that facilitate simple scaling and upgrading to fit any need and to detect behavior pattern structure. A. Social Sensing Layer: The first layer in this architecture acts as an interface between the social interactions of end users and data acquisition layer. This system allows users to enter and monitor the data collection process as it unfolds. B. Data

Acquisition and Preparation Layer: layerimpact on the quality of the data analysis carried out in a research study.A component called the task concurrent controller monitors all parallel tasks (requests, responses, and data manipulation processes) running through the system at a given time.C. Data Storage Management Layer*:* The main role of this subsystem is to establish the links betweendata storage and the active elements of the solution using a Hadoop framework[3].

***D. DNS Traffic Analysis for Malicious Domains Detection by using methodology of malicious domain detection*:**

This detection method is based on a blacklist of malicious domains. As it is shown in Figure 1 first process the network traffic, particularly DNS traffic. And then analyze all DNS requests and match the query with the blacklist. The blacklist of malicious domains is automatically updated each day and the detection is in the real time.This system also implemented Bro Intelligence Framework, this framework enables you to consume data from different data sources and make it available for matching.

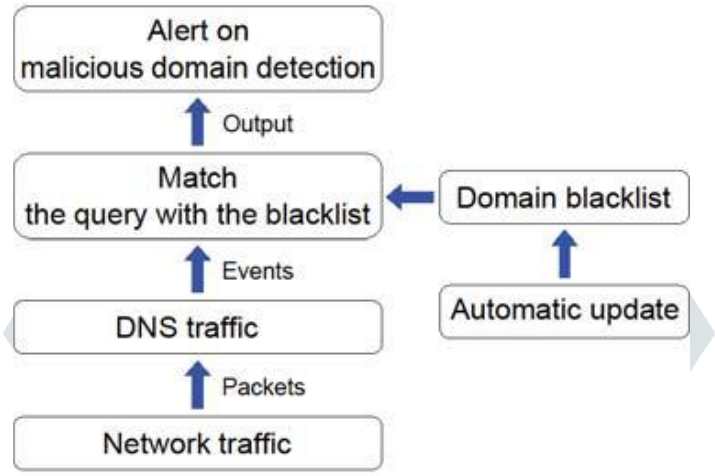

Fig.1. Methodology of malicious domain detection

In this detection method configured the intelligence framework to monitor all the domains which are seen in DNS query requests traffic [4].

E. Detecting Malicious Links in Online Social Networks through User Behavior by using url classification with OSN feature:

A) Host-based Features: The classifiers used to distinguish malicious URLs from legitimate URLs are more accurate when the most relevant features are extracted. It contains rich information about the website that hosts the URL. B) Domain-based Features: Based on the domain information such as IP, domain age and some DNS queries, a wide range of blacklist lookup services can be used to detect malicious URLs.C) User Profile-based Features: The profile is a page that contains the personal information of a particular user's account such as name, gender, photo, etc. A profile can be used to differentiate individual users in OSNs [5].

V. ANALYSIS AND DISCUSSION

User oriented behavior based malware variant detection system uses runtime behavior" with "static structures" to detect malware variants show that MONET can accurately detect malware variants and defend against transformation attacks with only a minimal performance and battery overhead and.[1].Evaluation of content based spam and malicious URL detection using Bayesian classifier and decision tree show how to detect malicious URL and also used data mining approach like supervised classification which improve the system accuracy and detect more amount of spam and malicious URLs.[2].Analysis of User Behavior to Identify Malicious Activities in Large-Scale Social Networks shows that how to social interactions on data consisting of textual content before reaching to an assumption of the activity by the user account to be real or malicious and the results of the experiments conducted using a classification engine applying five machine-learning algorithms, i.e. RF, decision tree J48, Cls Reg, SVM, and OIR,[3].DNS Traffic Analysis for Malicious Domains used method that based on blacklist of malicious domains show that how to process DNS traffic and list of malicious domain is automatically updated and detection in real time[4].Detecting Malicious Links in Online Social Networks use URL feature it focus on URL features in relation to malicious domain detection using classification methods show that how to identifying malicious users in a real dynamic environment and to improve classifications that can scale efficiently and handle URLs in OSNs [5].

TABLE 1:Comparison between different mobility schemes.

Prediction models and approach	Advantages	Disadvantages
User oriented behavior-based malware variants detection system by using MONET	This proposed method Accurately represent the runtime behavior of a malware. MONET can Achieve around 99% accuracy in detecting malware variants.	The drawback of this method MONET has a low impact on the battery resource with minimal Performance.
A comparative performance Evaluation of Content Based Spam and Malicious URL Detection in E-mail.	This method use data mining approach like supervised classification which improves the systems accuracy. In this Bayesian Classifier and Decision Tree C4.5 gives 95.54 % accuracy	The drawback of this method Bayesian classifier make a strong assumption on the data distribution .
Leveraging analysis of user behavior to identify malicious activities in large-scale social networks.	This proposed method provide performance By using three level of feature i.e.user-generated content, social graph connections, and user profile activities to analyze user behavior	The drawback of this method open to security issues (hacking, viruses, etc.) in OSN. research is being conducted on the development of different mechanisms to detect malicious activities in an OSN.
DNS Traffic Analysis for Malicious Domains Detection	This method detect Malicious domain connections in the real time provide better performance.	domain name has to be known before they can be added to the blacklist because Domain blacklist is not generally effective at detecting new or previously unknown malicious domains.
Toward Detecting Malicious Links in Online Social Networks through User Behavior	In this approach, Lexical features successfully improved the overall accuracy and OSNs also have many advantages in identifying malicious users in a real dynamic environment.	In this method increase in the percentage of attacked domains will also result in the same percentage of classification failures

VI. PROPOSED METHODOLOGY

Detection of malicious activities (domain, code URL, users behaviour analysis) is important and difficult task to analyse and discuss about various methods based on different parameters i.e. accuracy, quality, cost, time, flexibility, effectiveness, etc for different detection method .There are still problems which trouble in this field. Some approach using machine learning algorithm like J48. The evolution of malware possesses serious threat ever since the concept of malware took root in the technology industry. The malicious software which is specifically designed to disrupt, damage, or gain authorized access to a computer system has made a lot of researchers try to develop a new and better technique to detect malware but it is still inaccurate in distinguishing the malware activities and ineffective. To solve the problem, this paper proposed the integrated machine learning methods consist of J48 in detecting the malware accurately. The integrated classifier algorithm applied to examine, classify and generate rules of the pattern and program behaviour of system call information.

Algorithm for J48 :

i)In case the instances belong to the same class the tree represents a leaf so the leaf is returned by labeling with the same class.

ii)The potential information is calculated for every attribute, given by a test on the attribute.Then the gain in information is calculated that would result from a test on the attribute.

(iii)Then the best attribute is found on the basis of the present selection criterion and that attribute selected for branching

Diagrammatic representation of proposed method is shown as follows:

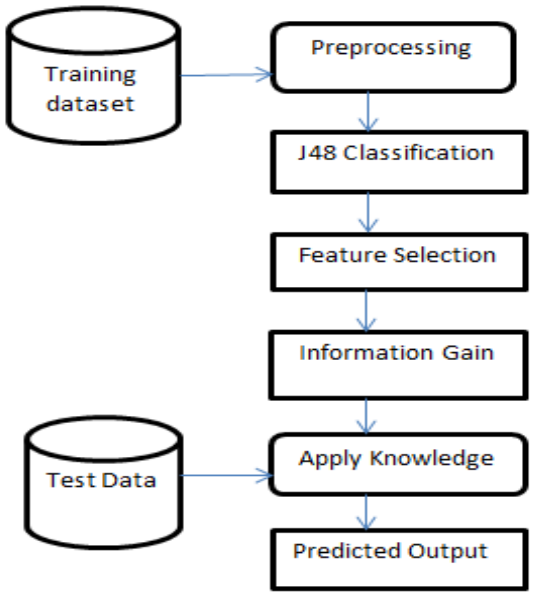

Fig 2: J48 Algorithm for Malware Detection

VII.OUTCOMES AND POSSIBLE RESULT

In this way the proposed method is performed for the detection of malicious activity usingJ48algorithm. With the help of the classification algorithm of the proposed method accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. All the algorithms other than proposed algorithm have accuracy rate below than 78% but the accuracy of the proposed algorithm is 99.87%.

VIII. CONCLUSION

This paper focused on the study of various approach i.e..User Oriented Behavior-Based Malware Variants Detection System by using MONET, Evaluation of Content Based Spam and Malicious URL Detection in E-mail by using Bayesian Classifier and decision tree, Analysis of User Behavior to Identity Malicious Activities in Large-Scale Social Networks, DNS Traffic Analysis for Malicious Domains Detection by using methodology of malicious domain detection, Detecting Malicious Links in Online Social Networks through User Behavior by using URL classification with OSN feature .But there are some problems in accuracy and performance so to improve this different classification algorithm are used like J48. This proposed method giveaccuracy99.87%.and both the discrete and continuous attributes are handled by this algorithm.

IX. FUTURE SCOPE

From observations of the proposed method the future work will include conducting additional experiments to tackle common problem, the modified J48 classifier has been used to increase the accuracy rate of the data mining procedure. In near future will use some more data sets to validate the proposed algorithm. The propose method can be added more efficient methods that will develop more validate data.

REFERENCES

[1]Mingshen Sun, Xiaolei Li, John C. S. Lui, *Fellow, IEEE*, Richard T. B. Ma, and Zhenkai Liang "Monet: A user-oriented behavior-based malware variants detection System for android",IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 12, NO. 5, MAY 2017.

[2] Sunil B. Rathod " Comparative Performance Evaluation of Content Based Spam and Malicious URL Detection in E-mail", 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS). Vol. Issue page no 4673-7437,august 2015.

[3] Muhammad Al-Qurishi, *Student Member, IEEE*, M. Shamim Hossain , *Senior Member, IEEE*, Majed Alrubaian, *Member, IEEE*, Sk Md Mizanur Rahman, *Member, IEEE*, and Atif Alamri, *Member, IEEE "*Leveraging Analysis of User Behavior to Identify Malicious Activities in Large-Scale Social Networks*" IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, VOL. 14, NO. 2, FEBRUARY 2018*

[4]. Ibrahim Ghafir and Vaclav Prenosil "DNS Traffic Analysis for Malicious Domains Detection*", 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)"*Vol. Issue page no 4799-5991,2015.

[5]Bandar Alghamdi, Jason Watson, Yue Xu "Toward Detecting Malicious Links in Online Social Networks through User Behavior", *2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops*page no. 7695-6039
September 2016 .