Application of Data Mining Methods in the Analysis of Different Attacks on Network

Shilpa Srivastava
M.Tech. Research Scholar, CSED, BERI, Bhopal, India
Dr.Mohit Gangwar
Principal/Professor (CSED), BERI, Bhopal, India
E-mail: mohitgangwar@gmail.com

Abstract- Many of us connect to Wireless Fidelity (Wi-Fi) without knowing what specific threats one is vulnerable. The list of vulnerabilities is large by nature, and most of these ignored by users. Computer networking has made collaboration necessary to both attackers and defenders. Phishing attacks combine technology and social engineering to gain access to restricted information. The most common phishing attacks today send mass email directing the victim to a web site of some perceived authority. This paper is focused on wireless network and phishing attacks. To analysis attacks on network signal we are applying different data mining algorithms like J48, random Forest and random Tree algorithms on network dataset of 3 years with 6 different attribute name "Company", "Data Provider", "Data Used", "Date", "Class", & "Signal" from different telecom companies to achieve 95 to 99% accuracy with a false positive rate of 0.5-1.5% and modest false negatives. Thus, the comparative views shows that J48 algorithm for phishing detection achieves better performance as compared to random Forest and random Tree algorithm.

Keywords- Phishing Attacks, J48, Random Forest, Random Tree, Confusion Matrix, Network Signal.

1. INTRODUCTION

Phishing attacks combine technology and social engineering to gain access to restricted information. The most common phishing attacks today send mass email directing the victim to a web site of some perceived authority. These web sites typically spoof online banks, government agencies, electronic payment firms, and virtual marketplaces. The fraudulent web page collects information from the victim under the guise of "authentication," "security," or "account update." Some of these compromised hosts simply download malware onto clients rather than collect information directly [1]. Phishing is a type of deception designed to steal your valuable personal data, such as credit card numbers, passwords, account data, or other information.

Figure 1: Working Process of Phishing Attacks [1]

1.1. TYPES OF PHISHING ATTACK

Type of Phishing attacks are as follow:

1.1.1 DECEPTIVE PHISHING

Today most common method for phishing is deceptive email message. Some scam emails links are received by the recipients though email. The user not having awareness about that scam they click and signing on that website from where the scammer collect all confidential information of that user [1, 4, 8].

1.1.2 MALWARE-BASED PHISHING

It is scams that involve running malicious software on users' PCs. They also allow copying all sensitive information and echoed to other software. This can be entered via email attachment, and downloadable files from website. This kind of phishing generates with those users who are not always update their software application [1, 4, 9, 10].

1.1.3 KEY LOGGERS AND SCREEN LOGGERS

This is a variety of malware that track keyboard input and send sensitive information to the hacker via the Internet. This malware generate itself into user's browsers as small utility programs that run automatically when browser is started [1, 4, 9, 10].

1.1.4 SESSION HIJACKING

It describes an attack where users' activities are monitored until they sign in to a target account or transaction. At that point the malicious software takes over and can undertake unauthorized actions, such as transferring funds, without the user's knowledge [1, 4, 9, 10].

1.1.5 WEB TROJANS

It generate invisibly pop up when users logged in. They collect all sensitive information of user and transmit them to the phisher [1, 4, 9, 10].

1.1.6 HOSTS FILE POISONING

When a user types a URL to visit a website it must first be translated into an IP address before it's transmitted over the Internet. By "poisoning" the hosts file, hackers have a fake address transmitted, taking the user unintentionally to a fake website where their information can be stolen [1, 4, 9, 10].

1.1.7 SYSTEM RECONFIGURATION ATTACKS

It modifies settings on a user's PC for malicious purposes. For example it modify favorites website to 'look alike' website for example: a bank website URL may be changed from "bankofabc.com" to "bancofabc.com" [1, 4, 9, 10].

1.1.8. DATA THEFT

Data theft is a widely used approach to business following. By stealing confidential communications, design documents, legal opinions, and employee related records, etc., thieves profit from selling to those who may want to embarrass or cause economic damage or to competitors [1, 4, 9, 10].

1.1.9. DNS-BASED PHISHING

Pharming is the term given to hosts file modification or Domain Name System (DNS)-based phishing. With a pharming scheme, hackers change the host's files or domain name system so that requests for URLs or name service return a fake address and similar communications are directed to a fake site. The result: users are not aware that the website where they are entering confidential information is controlled by hackers and is probably not even in the same country as the legal website [1, 4, 9, 10].

1.1.10. CONTENT-INJECTION PHISHING

It is describing the situation where hackers replace part of the content of a legal site with false content which is designed to mislead or misdirect the user into giving up their confidential information to the hacker [1,4,9,10].

1.1.11. MAN-IN-THE-MIDDLE PHISHING

It is harder to detect than many other forms of phishing. In these attacks hackers is present between the user and the legal website or system. They record the information being entered but continue to pass it on so that users' transactions are not affected. Later they use the information or credentials collected when the user is not active on the system [1, 4, 9, 10].

1.1.12. SEARCH ENGINE PHISHING

Phishers create websites with some attractive look and indexed it in search engine as a legal website. Users find the sites while searching normally for products or services and are fooled into giving up their information. For example, scammers provide false banking sites which offering lower credit costs or better interest rates than other banks. Victims who use these sites to save or make more from interest charges are encouraged to transfer existing accounts and deceived into giving up their details [1, 4, 9, 10].

1.2. ORGANIZATION OF PAPER RELATED WORK

Section 1: Introduction-In this section we give brief introduction of phishing attacks and their classification. Section 2: Related Work-In this section we studied various papers related to the previous classification method and algorithm. Section 3: Problem Identification & Problem Description-In this section we have described the problems of phishing attacks and its description. Section 4: Proposed Data mining Models-In this section we have applied different data mining models to analysis comparative results. Section 5: Simulation and Implementation-In this section we have discussed the implementation of applied data mining models, Section 6: Results & Discussion-In this section we have discuss analysis and results. Section 7: Conclusion-In this section we describe the conclusion.

2. RELATED WORK

One of the first mass attacks on embedded software was performed by the Chernobyl virus in [5]. The objective of this malware is purely obliteration. It attempts to erase the hard disk and overwrite the BIOS (Basic input and output service) at specified dates. Cell phones have also become targets for worms with the first reports in the wild in [6], the same author in [7] predicted infectious malware for the Linksys line of home routers, switches and wireless access points. Adelstein, Stillerman and Kozen identify nondestructive malware in Open Firmware boot platforms as a threat. To assure portability, parts of the boot software are written in the stack based language, Forth, and these scripts are executed via an interpreter. They propose a code analyzer the checks for malicious code at load time and prevent aged code from running. Arbaugh, Farber, and Smith implement a cryptographic access control system, AEGIS, to ensure that only sanctioned bootstrapping firmware can be installed on the host platform [3].

This study explores a variant of email based phishing, where distribution occurs through online market places and hardware is "spoofed" by maliciously compromising its embedded software. Our central example, the malicious home network router, steals information not only by passive eavesdropping, but by Pharming or DNS (Domain name Server or System) spoofing [2].

Browser toolbars at potential phishing web sites using a mixture of link analysis, content analysis, reputation databases, and IP (Internet Protocol) address information. Spoof Guard does two rounds of checks. If either of these tests fails, a second round examines images and form boxes to determine if the page semantically represents a request for information (e.g. login, credit card, etc.) Another system, PwdHash, generates per site passwords by hashing domain name concatenated to the user password. When the domain names differ, the resulting string does not reveal a usable passphrase. Pharming attacks defeat both of these tactics because they assume correct name resolution. The Net craft toolbar claims defends against Pharming attacks since it reveals the geographic location of the

server. While this can raise suspicion, it does not provide a strong defense. Criminal networks have commoditized zombie machines with prices ranging from $0.02 to $0.10 per unit; attackers can choose plausible locations for their hosts if this method ever becomes an effective defense [7].

Phishing is some kind of criminal activity employing in both technical and social engineering to steal personal information by surfing and visiting fake web pages which is look like as same as a legal website of bank and company and ask the user to enter personal information like user name, password, credit card number, etc. This paper main goal is to investigate the potential of data mining technique to detecting the complex problem of phishing website in order to help all users being hacked by stealing their personal information and password. Experimentations against phishing data sets and using different common associative classification algorithms (MCAR and CBA) and traditional learning approaches have been conducted with reference to classification accuracy. The results show that the MCAR and CBA algorithms outperformed SVM and algorithms [30].

Phishing attacks are one of the trending cyber-attacks that apply socially engineered messages that are communicated to people from professional hackers aiming at fooling users to reveal their sensitive information; the most popular communication channel to those messages is through users' emails. This paper presents an intelligent classification model for detecting phishing emails using knowledge discovery, data mining and text processing techniques. The pre-processing phase is enhanced by applying text stemming and WordNet ontology to enrich the model with word synonyms. The model applied the knowledge discovery procedures using five popular classification algorithms and achieved a notable enhancement in classification accuracy; 99.1% accuracy was achieved using the Random Forest algorithm and 98.4% using J48, which is to our knowledge, the highest accuracy rate for an accredited dataset. This paper presents a comparative study with similar proposed classification techniques [21].

3. PROBLEM DESCRIPTION AND PROBLEM IDENTIFICATION

On the basis of previous paper and related works networks are vulnerable for attacks. So, we have identified problem on the bases of vulnerabilities from previous papers are as follow [22]:

When discussing network security, the three common terms used are as follows:

■ Vulnerability- It is having a possibility to attack or harmed network or device. This includes routers, switches, desktops, servers, and even security devices themselves [22].

■ Threats- A threats is something which has potential to harm or damage the computer system or network [4, 22].

■ Attacks- The threats use a variety of tools, scripts, and programs to launch attacks against networks and network devices [22, 27].

3.1. PROBLEM DESCRIPTION

On the basis of previous paper and identification of different attacks for networks, Networks are not secure and easily accessible by the attackers, various data mining techniques are applied in detection and analysis of mobile networks [26-29]. In this work we are collected three year of network data from different telecom companies to identify the passive attacks on network signals either the network is active or not active. Table 1 & 2 represents different network provider companies with different parameters name Data provided in GB, Data used in GB, Data provided duration (1 month), Class (Active & Not Active) and Signal (Yes & No).

Network provider companies are:

TABLE 1: DIFFERENT MOBILE SERVICE PROVIDERS

Sr. No.	Mobile Service Provider
1.	Idea [31]
2.	Docomo [32]
3.	Airtel [33]
4.	Bsnl [34]
5.	Jio [35]
6.	Vodaphone [36]
7.	Relaince [37]

In Table 2 shows demonstration of dataset in that row represents the Network signal companies and the column represents their respective parameters.

TABLE 2: NETWORK DATA SET

Company	Data Provided (GB)	Data used (GB)	Date	Class	Signal
Idea	25	20	15-Jan-2015	active	Yes
Docomo	30	20	15-Jan-2015	Inactive	No
Airtel	28	20	15-Jan-2015	active	Yes
Bsnl	30	11	15-Jan-2015	Inactive	No
Jio	59	57	15-Jan-2015	active	Yes
Vodafone	39	34	15-Jan-2015	active	No
Reliance	28	25	15-Jan-2015	active	Yes
Idea	25	17	15-Feb-2015	Inactive	No

4. PROPOSED DATA MINING MODELS

In this Research Work various supervised classification algorithm techniques were applied to create different data mining models for detection and analysis of network signal data.

1. J48 Algorithm [23]
2. Random Forest Algorithm [24]
3. Random Tree Algorithm [25]

4.1. J48 ALGORITHM

J48 algorithm examine the normalization information gain that result to choose the attribute for splitting data, This splitting is stop if all instances in a subset belong to the same class. The first level of tree is a single header node. It is just a pointer node to its children. The second level of the tree has 2 sub trees labeled from 1 to2.

4.2. RANDOM FOREST

First Random Forest algorithm is a supervised classification algorithm, we can see it from its name, which is to create a forest by some way and make it random. There is a direct relationship between the numbers of trees in the forest by some way and make it random. There is a direct relationship between the number of trees in the forest and the result it can get the larger the number of trees, the more accurate the result. But one thing to note is that creating the forest is not the same as constructing the decision with information gain or gain index approach [24].

4.3. RANDOM TREE

It can be used with data in a distribute environment and requires that you have a connection to analytic Server. Using this node, you build an ensemble model that consists of multiple decision trees. The random Tree hub can be utilized with the information in an appropriated domain to construct a gathering model that comprises of various choices trees [25].

4.4. PROPOSED MODELS

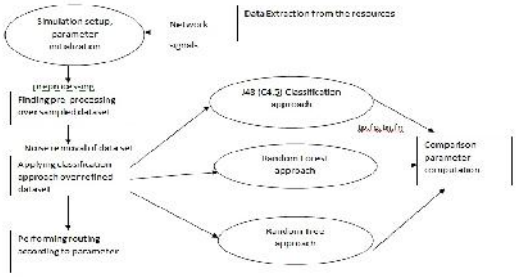


Figure 2: Flow Diagram for the Proposed Work Flow Which Is Executed Over Weka Tool

In Figure 2 shown the diagram for the proposed work flow which executed over Weka tool in this network signal data is extracted from the resources in next step parameter is initialized after preprocessing noise is removed from data after this applying classification approach (J48,Ramdom Forest and Random Tree) over refined dataset the comprised parameter is computed.

4.5. ACCURACY AND EFFICIENCY CALCULATION

Now, simply diagonal elements of the confusion matrix represent the true positive values and the rest of elements represent false positive values. Different operative characteristics are defined as follows:

True Positive (TP) = When test outcome is positive and condition is also positive then the situation is called as True Positive Values.

False Positive (FP) = When test outcome is positive and condition is negative then the situation is called as False Positive Values.

True Negative (TN) = When test outcome is negative and condition is also negative then the situation is called as False negative Values.

False Negative (FN) = When test outcome is negative and condition is positive then the situation is called as False positive Values.

4.6. ANALYSIS OF PRECISION, RECALL, ACCURACY BY USING DIFFERENT DATA MINING MODELS

Analysis, Precision, Recall, F-Measure, & Accuracy is calculated using following formulas [21].

- Precision = True Positive / (True Positive +False Positive)
- Recall = True Positive / (True Positive +False Positive)
- F1 = 2(Precision*Recall)/ (Precision + Recall)
- Accuracy = $\frac{\text{(True Positive +True Negative)}}{\text{(True Positive +True Negative +False Positive + False Negative)}}$

4.7. TOOLS AND SYSTEM DESCRIPTION

We are using creative models and to perform analysis of WEKA to perform our analysis. We required some hardware and software interface for the purpose of our simulation. Brief description is as shown in Table 3.

TABLE 3: REQUIRED HARDWARE AND SOFTWARE

Hardware Required	
System	Intel core I3
Hard disk	1Tb
RAM	4 GB
Software Required	
Operating System	Windows7
Tool	WEKA
Data Sheet	Excel File

4.8. WEKA TOOL

In the work WEKA tool is used for the analysis and evaluation.

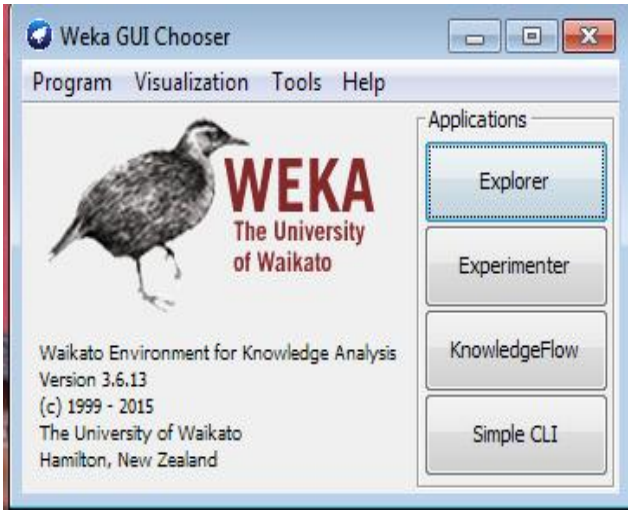


FIGURE 3: WEKA TOOL INITIALIZATION: STARTING WEKA GUI

In the above Figure 3, a WEKA tool Initialization is presented where Explorer, Experimenter, and other Knowledge framework is shown. This figure shows the start page of weka tool, which we have used to analysis.

5. SIMULATION AND IMPLEMENTATION

Here we collect three year of network data from different telecom companies to identify the passive attacks on network signals either the network is active or not active. In table 7 shows demonstration of dataset in that row represents the Network signal companies and the column represents their respective parameters.

TABLE 4: NETWORK DATA SET

Company	DataProvided (GB)	Dataused (GB)	Date	Class	Signal
Idea	25	20	15-Jan-2015	active	Yes
Docomo	30	20	15-Jan-2015	Inactive	No
Airtel	28	20	15-Jan-2015	active	Yes
Bsnl	30	11	15-Jan-2015	Inactive	No
Jio	59	57	15-Jan-2015	active	Yes
Vodafone	39	34	15-Jan-2015	active	No
Reliance	28	25	15-Jan-2015	active	Yes
Idea	25	17	15-Feb-2015	Inactive	No

Above Table 4 showing three year of network data from different telecom companies to identify the passive attacks on network signals either the network is active or not active. The table represents different network provider companies with different parameters name Data provided in GB, Data used in GB, Data provided duration (1 month), Class (Active & Not Active) and Signal (Yes & No).

TABLE 5: ATTRIBUTES AND ITS DATA TYPES

Attributes	Data Type
Idea	Categorical(Active, Not Active)
Docomo	Categorical(Active, Not Active)
Airtel	Categorical(Active, Not Active)
Bsnl	Categorical(Active, Not Active)
Jio	Categorical(Active, Not Active)
Vodaphone	Categorical(Active, Not Active)
Relaince	Categorical(Active, Not Active)

5.1. J48 Analysis Visualization

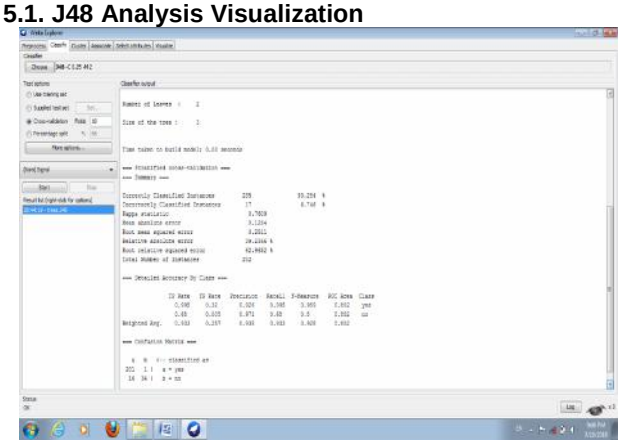


FIGURE 4: THE ANALYSIS VISUALIZATION BY J48 OF DETAILED ACCURACY BY CLASS

Information gain is the mathematical tool that algorithm J48 has used to decide, in each node, which variable fits better in terms variable prediction. In this algorithm correctly classified instances are 235 out of 252 instances and incorrectly classified instances are 17 out of 252 instances. The percentage of accuracy in J48 is 93.26% to detect network signals. TP, FP, TN using active and not active signals.

TABLE 6: CONFUSION MATRIX FOR RANDOM FOREST

A	B	Classified as
202	1	Yes
16	24	No

From the above Table 6 confusion matrix true positive for class a=yes is 201 while false positive is 1 where as class b= No is 34 while false negative is 16.

5.2. RANDOM FOREST ANALYSIS VISUALIZATION

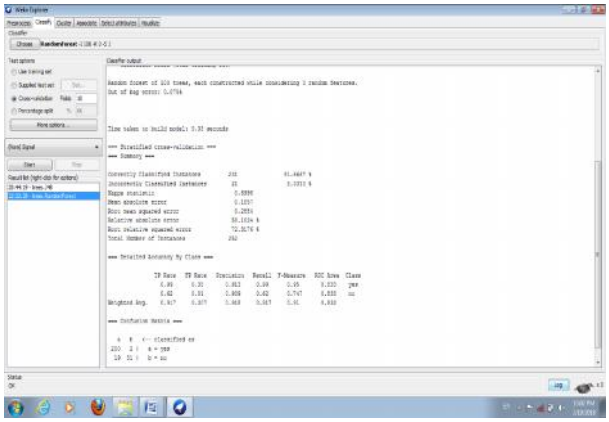


FIGURE 5: APPLYING RANDOM FOREST VISUALIZATION ON CLASSIFY ATTRIBUTE

TABLE 7: CONFUSION MATRIX FOR RANDOM FOREST

A	B	Classified as
200	2	Yes
19	31	No

From the above Table 7 confusion matrix true positive for class a=yes is 200 while false positive is 2 where as class b= No is 31 while false negative is 19.

5.3. RANDOM TREE ANALYSIS VISUALIZATION

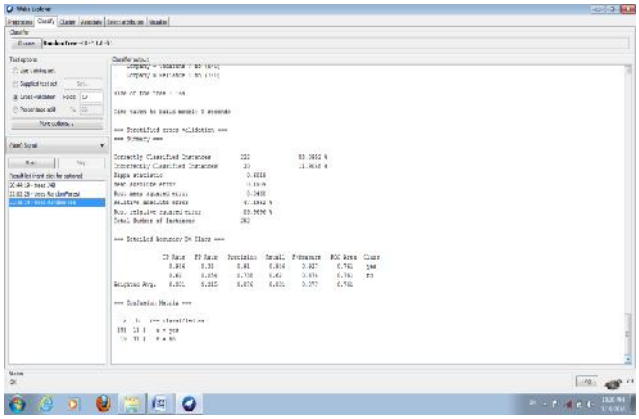


TABLE 8: CONFUSION MATRIX FOR RANDOM TREE

A	B	Classified as
191	11	Yes
19	31	No

6. RESULTS AND DISCUSSIONS

This section presents the results that the proposed classification model achieved by applying the three proposed classification algorithms to the features extracted from the data set of 252 network signal dataset. The generated features were fed to the three classifiers, namely J48, Random Forest& Random Tree. To avoid over fitting, we used 10-fold cross validation technique which uses 0.9 of the training data set as data for training the algorithm and the remaining 0.1 of training data set for testing purposes, and repeat this division of the data set for training and testing for 10 times. The experiments were conducted using the open source WEKA data mining software. The results were evaluated using the performance metrics discussed in the previous section. Table 9 depicts the weighted average of classification results for each of the algorithms. The results show that our model achieves high accuracy rates in classifying phishing on network signals, and outperforms similar proposed classification schemes as we will explain in the next section, thanks to the proposed pre-processing phase and feature reduction and evaluation process in the proposed model. The best results were achieved by the J48 is 93.26% to detect network signals.

TABLE 9: CLASSIFICATION OF DATA MINING MODELS

D.M Models	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
J48	0.933	0.257	0.935	0.933	0.928	0.802
Random Forest	0.917	0.307	0.918	0.917	0.91	0.833
Random Tree	0.881	0.315	0.876	0.881	0.877	0.761

6.1. COMPARATIVE ANALYSIS OF PROPOSED ALGORITHM

Information gain is the mathematical tool that algorithm J48 has used to decide, in each node, which variable fits better in terms variable prediction. In this algorithm correctly classified instances are 235 out of 252 instances and incorrectly classified instances are 17 out of 252 instances. The percentage of accuracy in J48 is 93.26% to detect network signals TP, FP, TN & FN using active and not active signals. Table 10 is showing comparative results of applied algorithms.

Dataset	Algorithm	Precision %	Recall %	F-Measure%	Accuracy%
Network Signals Number of instances is 252	J48	92.6	99.5	95.9	93.25
	RF	91.3	98.6	94.5	91.67
	RT	90.5	94.6	92.7	88.09

TABLE 10: COMPARATIVE ANALYSIS BETWEEN EXISTING AND PROPOSED ALGORITHM IN PERCENTAGE

6.2 A Graphical Comparison Analysis Of Classification Approach

In the Figure 7, a graphical analysis of computation parameter, with the data set is computed. Comparative view of result shows that in terms of accuracy J48-93.25% is the best.

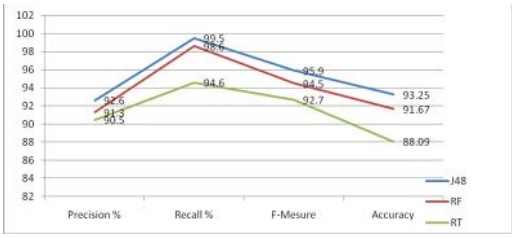


FIGURE 7: SHOWING THE GRAPHICAL REPRESENTATION OF ANALYSIS. RESULTS FOR APPLIED ALGORITHMS

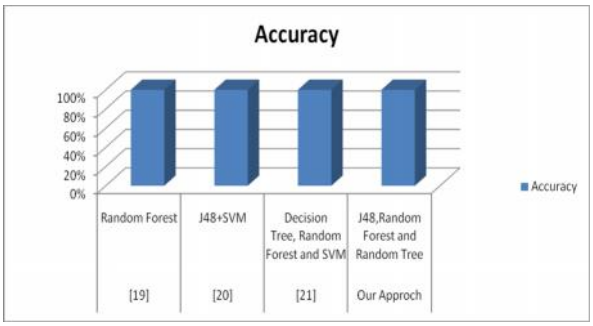
6.3. Comparative Analysis

A set of proposed studies are found in the literature of phishing email detection using data mining techniques, in this section we compare our proposed model with a set of previously proposed models for phishing detection. Table 11 summarizes a set of three previous related works along with the classification algorithm(s) used and the accuracy of the classification results, the results are visualized in Figure 8.

TABLE 11: COMPARISON OF OUR APPROACH WITH PREVIOUS WORK

Paper Reference	Classification	Accuracy
[19]	Random Forest	0.87
[20]	J48+SVM	0.89
[21]	Decision Tree, Random Forest and SVM	0.91
Our Approach	J48,Random Forest and Random Tree	0.93

FIGURE 8: COMPARISON OF OUR APPROACH ACCURACY WITH RELATED WORK



The study in [19] used a feature vector of 47 features extracted from the same data sets of Nazario and Spam Assassin corpus, using Random Forest algorithm for training the classification model. Their model achieved 0.87 accuracy. Our model outperforms their model in accuracy rate with less feature set. The study in [20] applied both J48 and SVM for classifying emails using a feature set of 30 features and yielded an accuracy rate of 0.88, our approach outperforms this result using the same classification algorithm J48 with a classification accuracy of 0.89. The study in [21] achieved high rate of accuracy in classifying phishing emails, it used a group of classification algorithms including Random Forest, SVM and decision trees. However, this study was built on a small and not verified phishing data set. In our approach we are applied J48 Random forest and Random tree approach different attacks on networks Our model achieved 0.93 accuracy. Our model outperforms their model in accuracy rate with less feature set.

7. CONCLUSION

Cybercriminals are continually finding new ways to avoid detection and develop techniques to and manipulate communications and improve the success rates of phishing attack. Phishing attacks combine technology and social engineering to gain access to restricted information. The most common phishing attacks today send mass email directing the victim to a web site of some perceived authority. This paper is focused on wireless network and phishing attacks. To analysis attacks on network signal we are applying different data mining algorithms like J48, random Forest and random Tree algorithms on network dataset of 3 years with 6 different attribute name “Company”, “Data Provider”, “Data Used”, “Date”, “Class”, & “Signal” from different telecom companies to achieve 95 to 99% accuracy with a false positive rate of 0.5-1.5% and modest false negatives. Thus, the comparative views shows that J48 algorithm for phishing detection achieves better performance as compared to random Forest and random Tree algorithm.

REFERENCES

[1] A Review on Phishin Attacks and Various Anti Phishing Techniques

[2] Stajano, F. and Wilson, P. Understanding scam victims: Seven principles for systems security. Commun. ACM 54, 3 (Mar. 2011), 70–75.

[3] Jagatic, T.N., Johnson, N.A., Jakobsson, M., and Menczer, F. Social phishing. Commun. ACM 50, 10 (Oct. 2007), 94–100.

[4] Hong, J. Why have there been so many security breaches recently? Blog@CACM.

[5] CERT. Incident note IN-99-03.http://www.cert.org/incident notes/IN-99-03.html, April 1999.

[6] Ivan Arce. The shellcode generation. IEEE Security & Privacy, September/October 2004.

[7] Ivan Arce. The rise of the gadgets. IEEE Security & Privacy, September/October 2003.

[8] Engin Kirda and Christopher Kruegel 2005 ,” Protecting Users Against Phishing Attacks with AntiPhish” Computer Software and Applications Conference, COMPSAC 2005. 29th Annual International (Volume: 1).

[9] Madhusudhanan Chandrasekaran Ramkumar Chinchani Shambhu Upadhyaya,” PHONEY: Mimicking User Response to Detect Phishing Attacks”, WOWMOM '06 Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks, Pages668-672, IEEE Computer Society Washington.

[10] Ying Pan, Xuhua Ding 2006”Anomaly BasedWeb Phishing Page Detection” Proceedings of the 22nd Annual Computer Security Applications Conference (ACSAC'06).

[11] Craig M. McRae Rayford B.Vaughn2007 ,”Phighting the Phisher:Using Web Bugs and Honeytokens to Investigate the Source of Phishing Attack“, Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07).

[12] Alireza Saberi, Mojtaba Vahidi, Behrouz Minaei Bidgoli 2007, “Learn To Detect Phishing Scams Using Learning and Ensemble Methods”, Proceedings of the 2007 IEEE/WIC/ACM.

[13] Eric Medvet, Engin Kirda, Christopher Kruegel 2008,”Visual-Similarity-Based Phishing Detection” Proceedings of the 4th international conference on Security and privacy in communication networks.

[14] Maher Aburrous, M. A. Hossain, Keshav Dahal, Fadi Thabatah 2009,”Modelling Intelligent Phishing Detection System for e-Banking using Fuzzy Data Mining” CyberWorlds, 2009. CW '09.

[15] Divya James and Mintu Philip2012,”A NOVEL ANTI PHISHING FRAMEWORK BASED ON VISUAL CRYPTOGRAPHY” International Conference on Power, Signals, Controls and Computation (EPSCICON).

[16] Mohd Mahmood Ali and Lakshmi Rajamani2012,”APD: ARM Deceptive Phishing Detector System Phishing Detection in Instant Messengers Using Data Mining Approach” Springer Berlin Heidelberg.

[17] Isredza Rahmi A HAMID, Jemal ABAWAJY, Tai-hoon KIM2013,“Using Feature Selection and Classification Scheme for Automating Phishing Email Detection” Studies in Informatics and Control22(1):61-70· March 2013.

[18] Moh'd Iqbal AL Ajlouni1, Wa'el Hadi,Jaber Alwedyan2013,”Detecting Phishing Websites Using

Associative Classification"European Journal of Business and Management www.iiste.org ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online) Vol.5, No.23, 2013.
[19] International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.4, July 2016.
[20] Phishing Activity Trends Report, http://docs.apwg.org/reports/apwg_trends_report_q1-q3_2015.pdf, Accessed June 2016.
[21] https://security.googleblog.com/2014/11/behind-enemy-lines-in-our-war-against.html , Accessed June 2016.
[22] Network security vulnerabilities threats and attacks pdf.
[23] Gaganjot kaur department of computer science and engineering GNDUAmritsa (pb.), India has introduced J48 Algorithm.
[24] Random Forest Algorithm Explained by Wikipedia- The free Encyclopaedia.
[25] Random Forest Tree Algorithm Explained by Wikipedia-The free Encyclopaedia.
[26] CAPEC-164: Mobile Phishing. https://capec.mitre.org/data/definitions/164.html
[27] Ashford, W. (2014) Phishing Attacks Track Mobile Adoption, Research Shows. http://www.computerweekly.com/news/2240215873/Phishing-attacks-track-mobile-adoption-research-shows.
[28] Kessem, L. (2012) Rogue Mobile Apps, Phishing, Malware and Fraud.https://blogs.rsa.com/rogue-mobile-apps-phishing-malware-and-fraud.
[29] Klein, A. (2010) The Golden Hour of Phishing Attacks. http://www.trusteer.com/blog/golden-hour-phishing-attacks.
[30] European Journal of Business and Management www.iiste.org (Paper) ISSN 2222-2839 (Online) Vol.5, No.23, 2013.
[31] https://www.ideacellular.com/.
[32] https://www.tatadocomo.com/
[33] https://www.airtel.in/.
[34] www.bsnl.co.in/
[35] https://www.jio.com/