International Journal for Modern Trends in Science and Technology, 8(07): 18-23, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0807003
Available online at: http://www.ijmtst.com/vol8issue07.html

Real Time Phishing Attack Detection using Machine Learning

Dr.M.Dhasaratham[1] | V.Pranavi[2] | I.Satya Janaki[2] | V.Pradeep[2] | Dr.Rajaram Jatothu[1]

[1]Professor, Dept of CSE,Teegala Krishna Reddy Engineering College,Hyderabad,Telangana,India.
[2]Dept of CSE,Teegala Krishna Reddy Engineering College,Hyderabad,Telangana,India.

To Cite this Article
Dr.M.Dhasaratham, V.Pranavi, I.Satya Janaki, V.Pradeep and Dr.Rajaram Jatothu. Real Time Phishing Attack Detection using Machine Learning. International Journal for Modern Trends in Science and Technology 2022, 8(07), pp. 18-23. https://doi.org/10.46501/IJMTST0807003

Article Info
Received: 25 May 2022; Accepted: 24 June 2022; Published: 27 June 2022.

ABSTRACT

Today, there is an exponential growth of e-services requiring the exchange of personal and sensible information over the Internet. Phishing techniques are emerging as the easiest solution to break the weakest link of the security chain: the end user. Social engineering attacks are deployed by financial/cyber criminals at a very low cost to induce naïve Internet users to reveal user ID, passwords, bank account and credit card numbers. This problem needs to be addressed in the mobile field as well, due to the large diffusion of mobile platforms such as smartphones, tablet, etc. To overcome this problem we propose a framework for phishing detection in Android mobile devices which, on the one hand exploits well-known techniques already implemented by popular web browsers plug-in, such as public blacklist search, and, on the other hand, implement a machine learning based engine to ensure zero-hour protection from new phishing campaigns.

KEY WORDS: Machine learning,Google API

1. INTRODUCTION

Motivation:

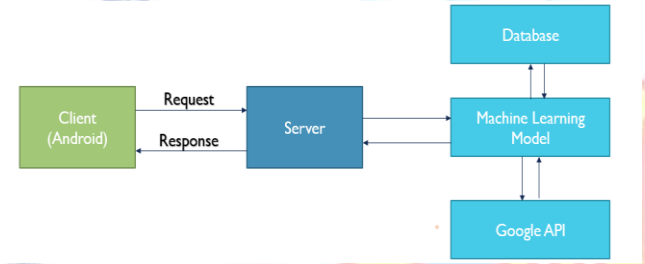
Social Engineering based attack leverages psychological manipulation of people, tricked into performing actions or disclosing confidential information. Phishing is one of the more known social engineering attack and aims at exploiting weaknesses in system processes caused by users' behavior. Indeed, a system can be secure enough against password theft(e.g. the client-server communication channel is encrypted),but nothing can be done against a naive user threatening the security of the system by revealing her/his password to a fake Web site reached, for example, via an email-embedded HTTP link.

Objective of Project:

Many of the approaches proposed in literature, regardless of their effectiveness, still have a strong verticality and focus on specific aspects such as: attack techniques; existing security context; systems and protocols used to capture data; methodological approaches used for phishing detection (black lists, heuristics, machine learning, etc.); devices on which deploy the developed solution. To the best of our knowledge, none of the solutions proposed so far, shown a unified approach across different environments (such as mobile and desktop) and across subsets of the above mentioned aspects. This is mainly due to the fact that, each solution often needs a number

of prerequisites that are: i) difficult to remove, in case some detection technique is covered by other third party software (e.g. antivirus); ii) hard to adapt to other contexts, enabling the exploitation of the same detection technique for other threats (e.g. botnets); Therefore, in this project, we propose a unified reference model and present the Real-time Phishing Attack Detection using Machine Learning (RPAD-ML) framework that implements it for the Android platform. Therefore, in this project, we propose a unified reference model and present the Real-time Phishing Attack Detection using Machine Learning
(RPAD-ML) framework that implements it for the Android platform.

2. ARCHITECTURE



An architecture for describing real time phishing attack detection using machine learning(RPADML)

3. LITERATURE SURVEY

Literature survey is the most important step in the software development process. The huge literature on phishing detection techniques is almost completely oriented toward methodologies tailored for desktop/laptop environments (see [4] for an extensive survey). Considering the literature related to the mobile environment that, as mentioned before, is not exhaustive, we can distinguish between work on "traditional" phishing [8-13] (resulting in no particular research trend) and work on Malware-based phishing detection [14]. The lack of literature on this specific subject is a consequence of a variety of reasons spanning from (a) the inevitable convergence of the same threat to any Internet enabled device (sometimes addressed as the Internet of Things), (b) the difficulty of implementing a performing security solution on a mobile device, (c) the absence of basic security settings on mobile devices (e.g. antivirus and firewall) and (d) the increased variability of attack vectors directed towards mobile devices.

4. IMPLEMENTATION

Introduction:

Implementation is the carrying out, execution, or practice of a plan, a method, or any design, idea, model, specifications, standard or policy for doing something. As such, implementation is the action that must follow any preliminary thinking in order for something that actually happens.For an implementation process to be successful, many tasks between different departments need to be accomplished in sequence. Companies strive to use proven methodologies and enlist professional help to guide them through the implementation of a system but the failure of many implementation processes obtain stems from the lack of accurate planning in the beginning stage of project.

Java Technology:

Initially the language was called as “oak” but it was renamed as “Java” in 1995. The primary motivation of this language was the need for a platform-independent (i.e., architecture neutral) language that could be used to create software to be embedded in various consumer electronic devices.

Java Architecture:

Java architecture provides a portable, robust, high performing environment for development. Java provides portability by compiling the byte codes for the Java Virtual Machine, which is then interpreted on each platform by the run-time environment. Java is a dynamic system, able to load code when needed from a machine in the same room or across the planet.

Java Database Connectivity :

JDBC is a Java API for executing SQL statements. (As a point of interest, JDBC is a trademarked name and is not an acronym; nevertheless, JDBC is often thought of as standing for Java Database Connectivity. It consists of a set of classes and interfaces written in the Java programming language. JDBC provides a standard API for tool/database developers and makes it possible to write database applications using a pure Using JDBC, it is easy to send SQL statements to virtually any relational

database. One can write a single program using the JDBC API, and the program will be able to send SQL statements to the appropriate database. The combinations of Java and JDBC lets a programmer write it once and run it anywhere.
Simply put, JDBC makes it possible to do three things:

- Establish a connection with a database

Two-tier and Three-tier Models :

The JDBC API supports both two-tier and three-tier models for database access. In the two-tier model, a Java applet or application talks directly to the database. This requires a JDBC driver that can communicate with the particular database management system being accessed. A user's SQL statements are delivered to the database, and the results of those statements are sent back to the user. The database may be located on another machine to which the user is connected via a network. This is referred to as a client/server configuration, with the user's machine as the client, and the machine housing the database as the server. The network can be an Intranet, which, for example, connects employees within a corporation, or it can be the Internet.

Python Technology:

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.[26] Van Rossum led the language community until stepping down as leader in July 2018.[27][28]

5. SYSTEM ANALYSIS

Existing System:

The phishing detection system is confined to Google Chrome Browser. It uses Google's Safe Browsing Database to warn user before entering a possible phishing site. It's more flexible for systems rather than mobile devices.

DISADVANTAGES OF EXISTING SYSTEM:

- Although, it protects user from entering into phishing sites on Chrome Browser, it fails to do so on other browsers.
- It cannot detect newly created phishing sites (Someone has to report it)
- Doesn't work in other mobile browsers
- Doesn't scan all the links opened by the device(Mobile)
- Cannot intercept all network traffic like GET requests.

PROPOSED SYSTEM:

To overcome the drawbacks of existing system, we propose a framework RPAD-ML (REALTIME PHISHING ATTACK DETECTION USING MACHINE LEARNING). It is based on Machine learning classification model and can detect phishing attacks in mobile. The machine learning model is kept on server and the client device is provided with an app which can communicate with this ML server and intercept the web pages and apps.

ADVANTAGES OF PROPOSED SYSTEM:

- Functional Capabilities
- Performance Level
- Data Structures
- Safety
- Reliability
- Quality

6. MODULES

Here we are using two modules. The name and description of each module are given below
In this module, the user can open an app or browser and then input an URL if the opened app is a browser. The user can load web pages and open web links in apps.

RPADML:

In this module, the web page/link to be loaded is sent as data to the server. Where this server having machine learning model at its backend checks whether the requested URL is phishing or not. And gives back the result to the user.

Android App:

It acts as an interface for the user and machine learning server. All the website URLs/Links to be opened are first scanned by the app and then sent to the machine

learning server. The server then classifies and gives a result. Based on the response received from the server the app intercepts the web page/app which is loading a possible phishing site/URL.

7. TESTING & VALIDATION

INTRODUCTION:

Application Testing is the process used to help identify the correctness, completeness, security, and quality of developed user application. Testing is a process of technical investigation, performed on behalf of stakeholders, that is intended to reveal quality-related information about the product with respect to the context in which it is intended to operate. This includes, but is not limited to, the process of executing a program or application with the intent of finding errors. Quality is not an absolute; it is value to some person. With that in mind, testing can never completely establish the correctness of arbitrary computer software; testing furnishes a criticism or comparison that compares the state and behavior of the product against a specification. An important point is that software testing should be distinguished from the separate discipline of Software Quality Assurance (SQA), which encompasses all business process areas, not just testing.

Test levels

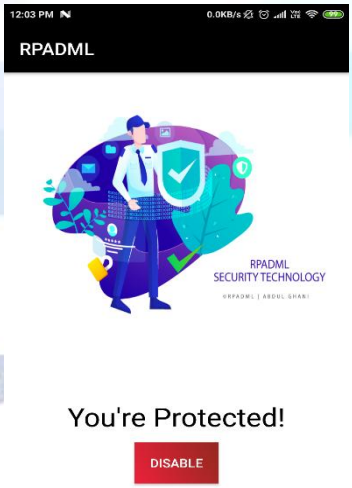
Unit testing tests the minimal software component and sub-component or modules by the programmers.

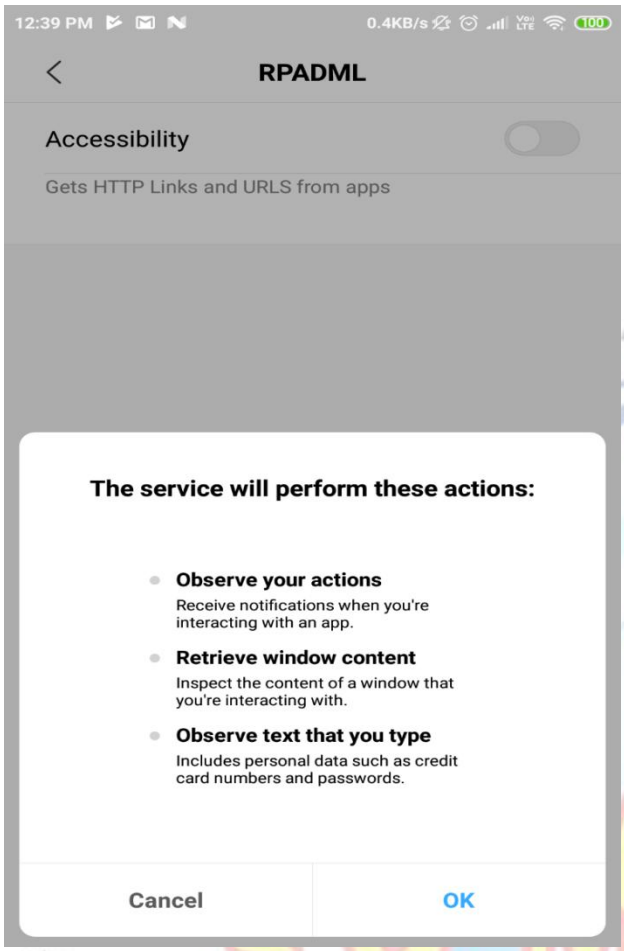
- Integration testing exposes defects in the interfaces and interaction between integrated components (modules).
- Functional testing tests the product according to programmable work. System testing tests an integrated system to verify/validate that it meets its requirements.
- Acceptance testing can be conducted by the client. It allows the end-user or customer or client to decide whether or not to accept the product. Acceptance testing may be performed after the testing and before the implementation phase. See also Development stage
- Beta testing comes after alpha testing. Versions of the software, known as beta versions, are released to a limited audience outside of the company. The software is released to groups of people so that further testing can ensure the product has few faults or bugs. Sometimes, beta versions are made available to the open public to increase the feedback field to a maximal number of future users.

8. TEST CASES:

A test case is a software testing document, which consists of event, action, input, output, expected result and actual result. Clinically defined (IEEE 829-1998) a test case is an input and an expected result. This can be as pragmatic as 'for condition x your derived result is y', whereas other test cases described in more detail the input scenario and what results might be expected. It can occasionally be a series of steps (but often steps are contained in a separate test procedure that can be exercised against multiple test cases, as a matter of economy) but with one expected result or expected outcome. The optional fields are a test case ID, test step or order of execution number, related requirement(s), depth, test category, author, and check boxes for whether the test is automatable and has been automated. A test case should also contain a place for the actual result. These steps can be stored in a word processor document, spreadsheet, database or other common repository. In a database system, you may also be able to see past test results and who generated the results and the system configuration used to generate those results. These past results would usually be stored in a separate table.

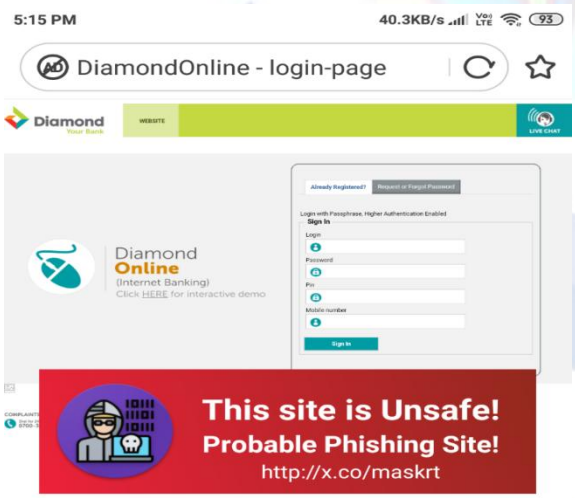
9.RESULT ANALYSIS:





Screen No.5.2.1.2 Accessibility Permission page

10. OUTPUT SCREENS:



Screen No.5.2.2.1 Phishing Website detected Page

11. CONCLUSION:

By using RPADML system, we solve the problem of detecting phishing sites on mobile devices in real-time. Now, the users are able to identify phishing sites/links without performing any activity.

RPADML system itself shows floating warning sign before entering such websites.

11. FEATURE ENHANCEMENTS:

- Adding compiled machine learning model in local devices
- Increasing the efficiency of API i.e., Response Time by leveraging Server Resources
- Ability to report false-positive results.
- Better garbage management in client device.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Gunter Ollmann, "The Phishing Guide Understanding & Preventing Phishing Attacks", IBMInternet Security Systems, 2007.

[2] https://resources.infosecinstitute.com/category/enterprise /phishing/the-phishing-landscape/phishing-data-attack-statistics/#gref

[3] Mahmoud Khonji, Youssef Iraqi, "Phishing Detection: A Literature Survey IEEE, and Andrew Jones, 2013

[4] Mohammad R., Thabtah F. McCluskey L., (2015) Phishing websites dataset. Available: https://archive.ics.uci.edu/ml/datasets/Phishing+Websites Accessed January 2016

[5] http://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/

[6] http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/

[7] https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html

[8] www.alexa.com

[9] www.phishtank.com

[10] J.Rajaram presented a Paper in 5th International Conference on Technical Advancements inComputer Science an Engineering (ICTACSE -2019) Paper on"Theoretical Research on NTBSProtocol:ItsImplementation and Possibilitiesin Nibble bits forVANET Security" in MRECW, Hyderabad, India,ISBN: 978-93-8880847-7

[11] https://www.researchgate.net/figure/Extraction-of-L7-PDUs-from-input-packets_fig2_300144295

[12] J.Rajaram presented a Paper in International Conference on "International Conference On Innovations In Signal Processing And Embedded Systems (Icispes-2021) 22nd Oct – 23rd Oct

2021)" Paper on "Network Intrusion Detection Using Machine Learning for Virtualized Data" in MLRIT, Hyderabad, India.

[13] Rajaram Jatothu, Dr. R .Vivekanandam, presented in the National Conference on Research Advamcements in Computational Informatics(RACI2018) paper on "The Time Efficient Privacy-Preserving Multi-KeywordRankedSearch moreEncrypted Cloud Data" at In Anurag Group ofInstitutions, Hyderabad, India Vol 5, Issue 4, April2018| ISSN: 2394-2320.

[14] Rajaram Jatothu, Published a paper on "A Review Paper On Data Mining with Big data analysis algorithms, Tools, Applications and Challenges' International Journal of Management, Technology And Engineering Volume 8, Issue IX, SEPTEMBER/2018, ISSN NO : 2249-7455.

[15] Rajaram Jatothu,Dr.R.P.Singh, Published a paper on"Distributed Protocol Using Quantum Cryptography For Secure Communication In Ad hoc Networks"International Journal of Pure and Applied Mathematics Volume 116 No. 23 2017, 253-260 SSN:1311-8080 (printed version); ISSN: 1314-3395.

[16] Rajaram Jatothu, R.P.Singh, Published a paper on "The Efficient Route Management Protocol for Misbehavior in Ad Hoc Networks" International Journal of Innovations & Advancement in Computer Science ISSN 2347 – 8616 Volume 6, Pg No: 346-359, Issue 10 October 2017.

[17] Rajaram Jatothu, Dr.R.P.Singh, Published a paper on "The Role of Quantum Cryptography under Distributed Protocols for Secured Communication in Ad Hoc Networks" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169, Pg No. 1089 –1096Volume: 5 Issue: 6.

[18] Rajaram Jatothu International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 NATIONAL CONFERENCE on

[19] Developments, Advances & Trends in Engineering Sciences (NCDATES- 09th & 10th January 2015)Paper Title: "Concerning Numerous-Hop, Numerous-Path Fault-Tolerant and Load Balancing Stratified Routing Protocol for Wireless Sensor Network".

[20] M.Dhasaratham Published a paper on"Distributed hybrid AODV Algorithm for path Concern in MANET Using Bio Inspired Techniques" International Journal of Adv Research in Dynamical & Control Systems, Vol 11 No. 1,2019, ISSN 1943-023X(Scopus Indexed)

[21] M.Dhasaratham Dr.RP singh Published a paper on"A Survey on data anonymization using map reduce on cloud with scalable two-phase top-down approach" International Journal of Engineering & Technology (IJET), 7(2.20) (2018) 254-259(Scopus Indexed)

[22] M.Dhasaratham Published a paper on"Criminal investigation analysis using ID3 a machine learning algorithms"The International Journal of Analytics and Experimental Model Analysis, Vol. XII, Issue V, May/2020, ISSN NO:0886-9367.

[23] M.Dhasaratham Dr.RP singh Published a paper on"Enhanced Security for data sharing in clouds through policy and access control management" International Journal of Research in Advanced Computer Science Engineering, Vol. 3,No. 6,November 2017, ISSN 2454-423X.

[24] M.Dhasaratham Dr.J Rajaram, Dusara Deksisa Geleto, Fayisa Dekebi Tusamo Published a research paper on"Low Resolution image Improvement of Aerial images using SCIKIT Tools"The International Journal of Analytical Aand Experimental Model Analysis (IJAEMA), ISSN no:08886-9367

[25] M.Dhasaratham, Mubeena Shaik, Naseema Shaik Published a paper on "Data mining concepts with customer relationship management" International Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol.4,Issue7(Version 6), July 2014,pp.98-100