MIT ADT University
MIT School of Engineering
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
S.Y. (CSE-3)
Mini Project - I

ARK Security

Group members:

Name	Roll no.	e-mail id	Contact no.
S. Utkarsh Rao	2203153	utkarshrao1177@gmail.com	6204644100
Manan Kukreja	2203180	manan241202@gmail.com	9511169915
Mohammad Musaib Akhter	2203280	musaabmallick163@gmail.com	7256013404

Project title: - Phishing to prepare a utility software
Project guide: - Dr. Rajani Sajjan

Copyright document

Introduction: -

Phishing is an act attempting to acquire information such as user name password and credit card details as a trustworthy entity in an electronic communication. Communications purporting to be from popular social websites, auction sites, online payment process or IT administrators are commonly used to lure the unsuspecting public. Phishing emails may contain links to website that are infected with malware.

Background (Literature survey): -

1. Extracting URL with trained model is lightweight operation as compared to downloading whole webpage and using its contents for extracting purpose. There exist number of URL features that allow for detection of phishing sites.
2. Server's identity is achieved through the use of an IP address. A trust website usually has domain name for its verification and phishing sites normally use some unauthenticated Zombie system to host that particular site. For avoiding from domain registration or user checking, the IP address is a way used to hide from all identification and verification.
3. When '@' or '_' these suspicious characters are present in URL that URL is Suspicious URL. This URL is used to detect phishing web. When phishing web try to access the victims, then URL is checked, if the URL is suspicious then that site is phishing site. Generally, the URL should not contain more number of slashes. If URL contains more than five slashes then that URL will be a phishing URL.
4. Rabab Alayham Abbas Helmi et al [4] have designed and developed a tool to detect the source code of a phishing website which is attached to email by using a decision tree algorithm. In order to improve the protection of user's information from the fake website. Anti-phishing detection is suggested to overcome the problems through the following features. The first feature of an anti-phishing detection login system is by using the user's email and password. Second feature is detecting phishing websites which are attached to a user's email by using a decision tree algorithm. Lastly, a phishing website will be detected and generate a report to the user.
5. Since phishing e-mails must resemble online Banking and retailers to gain the trust of the user in divulging their information, the phisher in the e-mails mimic the appearance of a reputational company. The companies spoofed most often are Citibank, eBay, and PayPal.

6. Frequently, phishers attempt to conceal the destination website by obscuring the URL. One method of concealing the destination is to use the IP address of the Web site, rather than the hostname. An example of an IP address used in a fraudulent e-mail message's URL is http://210.14.228.66/sr/. Also, the URL can be hidden through representation in DWORD, Octal, or Hexadecimal format.
7. Most of the phishing e-mail use the underlying context such as invoking a sense of false urgency, threat, wheedle, and concern to deceit the user in clicking on the visited hyperlink. Therefore, it is important to build such context graphic models for detection.

Proposed system: -

The proposed approach for phishing email classification employs the model of Knowledge Discovery (KD) and data mining for building an intelligent email classifier that is able to classify a new email message as a legitimate or spam; the proposed model is built by applying the iterative steps of KD to identify and extract useful features from a training email data set, the features are then fed to a group of data mining algorithms to identify the best classifier. The proposed model for email classification utilizes linguistic processing techniques and ontologies to enhance the similarity between emails with similar semantic term meaning, also the principle of term document frequency is applied in weighting the phishing terms in each email such that emails phishing terms weighting helps indiscriminating phishing from legitimate emails. The proposed model also reduced the number of features used in the classification process into 16 features only; which enhances the classification performance and efficiency and minimizes the noise of including many features and hence improves the classification accuracy. These enhancements and are discussed in detail in the following subsections. In the fig.1 we have proposed the block diagram and how your software will stop the attack while you receive a mail or a URL.
In the fig.2 we have shown the flowchart that how are software will go about with the process when it finds phishing mails or URLs.

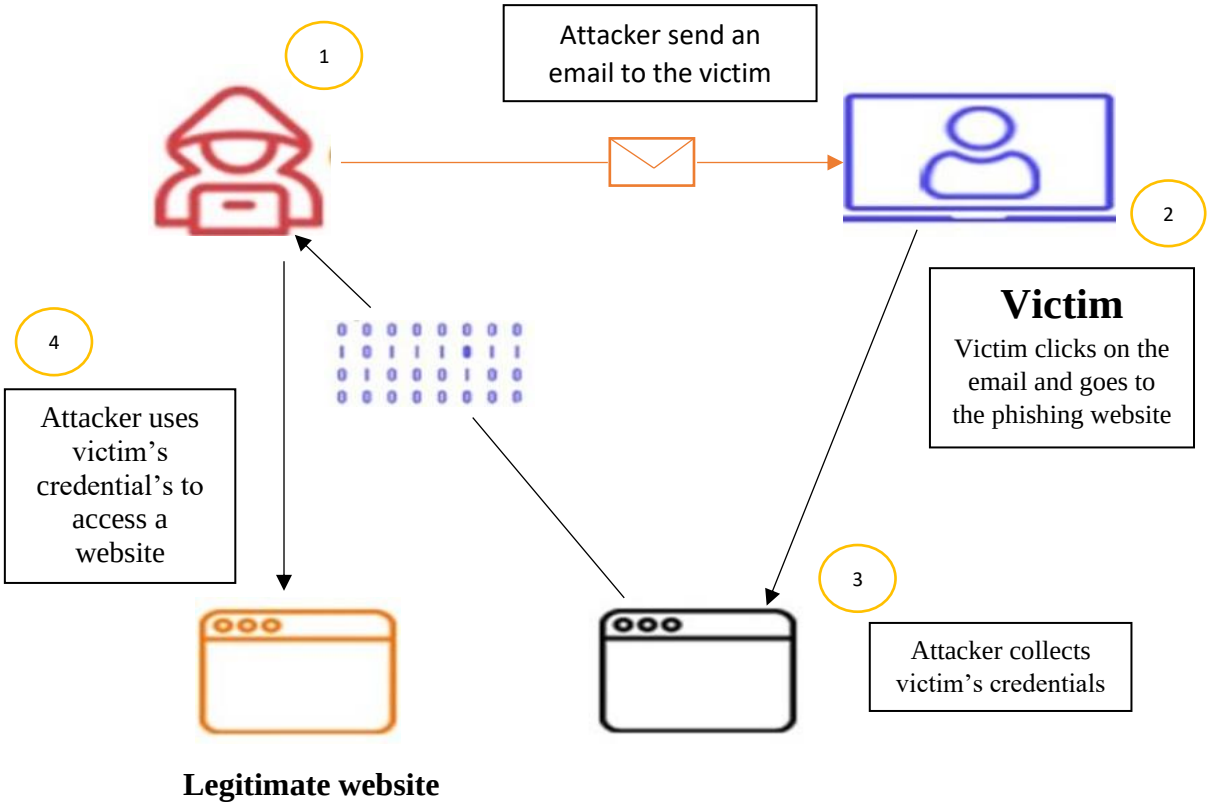

Fig.1- Block Diagram

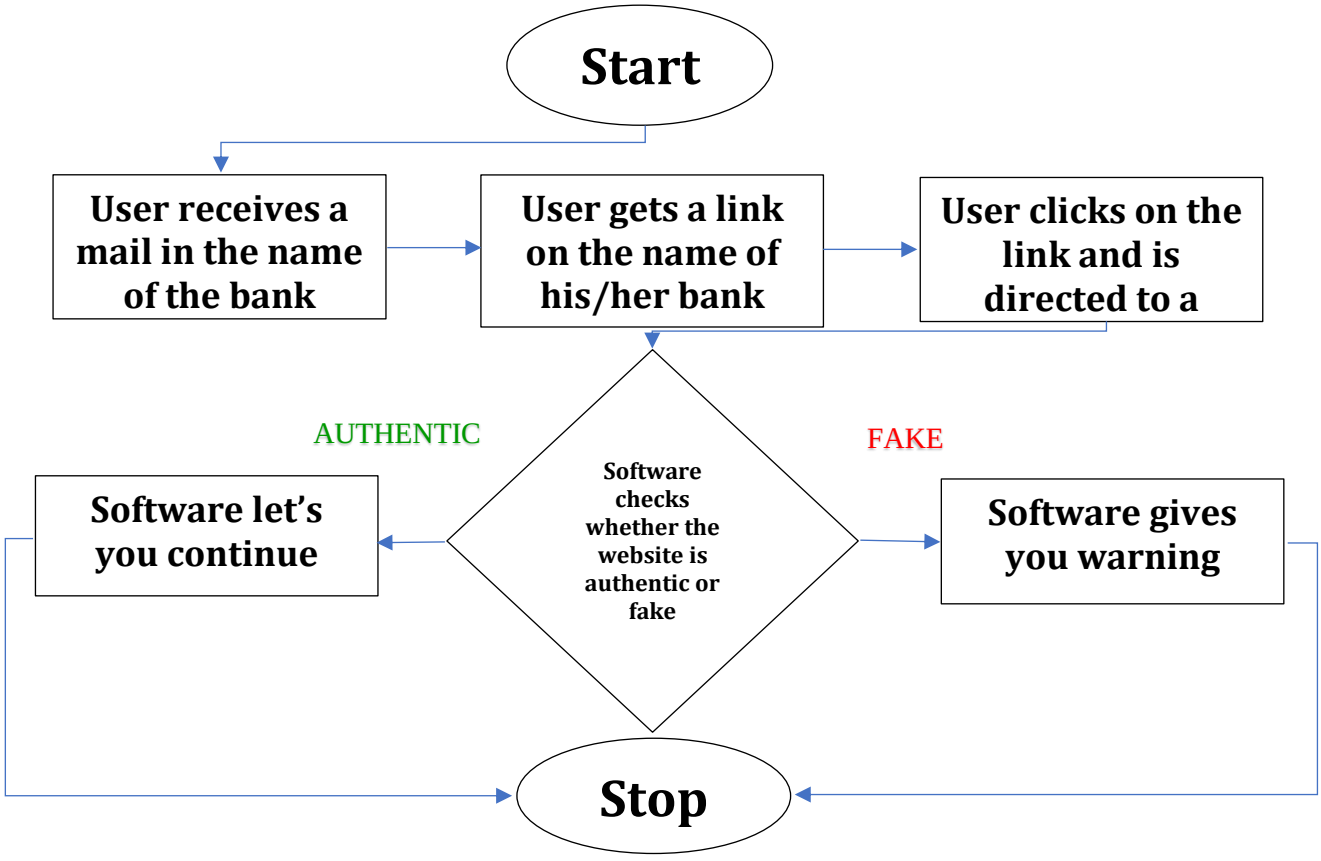

Fig.2- Flowchart

Working of system: -

The user receives a mail (say in the name of their bank) asking to submit details for regular updates or something as such through an e-mail then as the user will click on the URL it will direct them to a website. Here our software comes into play, it will detect that the website is really an authentic one or not. If it is authentic then the software will let you proceed and if it is fake then it will send a warning message that this website might be fake.

Application: -

Phishing protection software can help to prevent an attack by shielding employees from suspicious emails, blocking malicious links, stopping weaponized attachments and identifying signs of fraud and impersonation. It is often integrated with web browsers and email clients as a toolbar that displays the real domain name for the website the viewer is visiting, in an attempt to prevent fraudulent websites from masquerading as other legitimate websites.

References: -

1)"Phishing Knowledge based User Modelling in Software Design" - Linfeng Li, Timo Nummenmaa, Eleni Berki, Marko Helenius. Beijing Institute of Petrochemical Technology, Beijing, China, University of Tampere, Tampere, Finland, Tampere University of Technology, Tampere, Finland http://ceur-ws.org/Vol-1525/paper-16.pdf

2) "Types of anti-phishing solutions for phishing attack" - Siti Hawa Apandi, Jamaludin Sallim and Roslina Mohd Sidek Faculty of Computing, College of Computing and Applied Sciences, University Malaysia Pahang, Gambang, Kuantan, Pahang, Malaysia https://iopscience.iop.org/article/10.1088/1757-899X/769/1/012072/meta

3) "Machine learning based phishing detection from URLs"- Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, Banu Diri. Istanbul Kultur University, Computer Engineering Department, 34158 Istanbul, Turkey b Marmara University, Technology Faculty, Computer Engineering Department, Istanbul, Turkey c Yildiz Technical University, Computer Engineering Department, Istanbul, Turkey.
https://www.researchgate.net/profile/Ebubekir-Buber/publication/344952543_Machine_learning_based_phishing_detection_from_URLs/links/5f9acba0299bf1b53e4f22e1/Machine-learning-based-phishing-detection-from-URLs.pdf

4) “Phishing E-mail detection based on Structural Properties” - Madhusudhana Chandrasekaran, Krishnan Narayan and Shambhu Upadhyaya Department of computer science engineering, state university of New York at Buffalo
https://www.albany.edu/wwwres/conf/iasymposium/proceedings/2006/chandrasekaran.pdf

5) “An Efficient Approach to Detecting Phishing A Web Using K-Means and Naïve-Bayes Algorithms” - Ms. Kranti Wanawe, Ms. Supriya Awasare, Mrs. N. V. Puri. Computer Department Universal College Of Engineering and Research, Pune
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.428.4810&rep=rep1&type=pdf

6) “URL based Email Phishing Detection Application” - Roshan Ravi, Abhishek Arvind Shillare, Prathamesh Prakash Bhoir, K.S. Charumathi, B.E. Student, Department of Information Technology, Pillai College of Engineering, New Panvel, Navi Mumbai, Maharashtra – 410206, India 4Assistant Professor, Department of Computer Engineering, Pillai College of Engineering, New Panvel, Navi Mumbai, Maharashtra – 410206, India
https://irjet.com/archives/V8/i4/IRJET-V8I466.pdf

7) “A Transformer-based Model to Detect Phishing URLs” - Pingfan Xu School of Computer Science University of Guelph Guelph, Canada
https://arxiv.org/pdf/2109.02138.pdf

8) “URL Phishing Analysis using Random Forest” - S. Jagadeesan (Asst. Professor) (jagadeesan.s@ktr.srmuniv.ac.in), Anchit Chaturvedi (anchitmudit@gmail.com), Shashank Kumar(shashank.kumar14@gmail.com), Department of Computer Science and Engineering SRM Institute of Science and Technology Chennai.
http://www.acadpubl.eu/hub/2018-118-21/articles/21e/49.pdf