A Smart Approach to Avoid Phishing

Paras Gandhi
Department of Computer Engineering
Bharati Vidyapeeth University College of Engineering

Osho Tripathi
Department of Computer Engineering
Bharati Vidyapeeth University College of Engineering

Abstract

This research reports on the smarter way to tackle phishing and various web attacks. In addition, observations on the outcomes obtained from examination, group-project coursework, and informal feedback from victims have been analysed. This resulted in uncovering the same pattern and habits of the victims and how to utilize them to improve their security. This study led to devise specific pattern recognition using classification algorithm of machine learning to improve the user's security.
Keywords- Phishing, Classification Algorithm, Machine Learning

I. INTRODUCTION

Phishing is defined as the fraudulent acquisition of personal information by tricking an individual into believing the attacker is a trustworthy entity this is normally done through emails and an instant message. Phishing email will generally direct the user to visit a site where they are asked to update personal information, such as a password, credit card, social security, or bank account numbers, that the valid organization already has. The website, however, is unsafe and will record and steal any information the user enters on the page. It is the attempt to acquire sensitive information such as usernames, passwords, and credit card details (and sometimes, indirectly, money), mostly for malicious reasons, by masquerading as a trustworthy entity in an electronic communication.

The word is a neologism created as a homophone of fishing because of its similarity to using a bait in an attempt to catch a victim. Communications purporting to be from popular social web sites, auction sites, banks, online payment processors or IT administrators are commonly used to lure non-technical victims. Phishing emails may contain links to sites that are infected with malware. Phishing is typically carried out by email spoofing or messaging, and it often directs users to enter details at a fake site whose look and feel are almost identical to the legitimate one.

Phishing is an example of social engineering techniques used to cheat users, and exploits the poor usability of current web security techs. Attempts to deal with the growing number of reported phishing incidents include legislation, public training, awareness, and tech security measures. Many websites have now created secondary tools for applications, like maps for games, but they should be clearly marked as to who developed them, and users should not use the same passwords anywhere on the internet.

Phishing is a big threat, and the risk is even bigger in social media such as Facebook, Twitter, and Google+. Hackers could create a clone of a website and tell you to enter personal information, which is then mailed to them. Hackers commonly take advantage of these websites to attack public using them at their workplace, homes, in order to take personal and security information that can affect the user or company.

Phishing uses the trust that the user may have since the user may not be able to tell that the site being visited, or program being used, is not real; therefore, when this occurs, the hacker has the chance to gain the personal information of the targeted user, such as passwords, usernames, security codes, and card numbers, among other things.

II. CASE STUDY

2000 saw the proliferation of a phishing scam in which people received mails supposedly from eBay claiming that the user's account was about to be suspended unless he clicked on the provided email link and updated the credit card information that the genuine eBay already had. Because it is relatively simple to make a website look like a legitimate organization's site by copying the HTML code, the scam counted on people being tricked into thinking they were actually being contacted by eBay and were actually going to eBay's site to update their account information.
Here is an example of what a phishing scam in a mail may look like.

Fraudsters send fake mails or set up fake sites that mimic site's sign-in pages (or the sign-in pages of other trusted companies, such as Flipkart or SBI) to trick you into disclosing your username and password. This practice is often referred to as "phishing" — a play on the word "fishing" — because the hacker is fishing for your private account info. Typically, these hackers try to trick you into providing your user name and password so that they can gain access to your online account. Once they gain access, they can use your personal info to commit identity theft, charge your cards details, empty your bank accounts, read your mails, and lock you out of your online account by changing your password.

If you receive a mail or message from someone you don't know directing you to sign in to a site, be careful! You may have received a phishing mail with links to a phished site. A phishing site (sometimes called a "spoofed" site) steals your account passwords or other confidential information by tricking you into believing you're on a legitimate website. You could even land on a phishing site by mistyping a website name.

Is that site legitimate? Don't be fooled by a site that looks real. It's easy for phishers to create sites that look like the genuine article, complete with the logo and other graphics of a trusted sites.

Important: If you're unsure about a website, do not log in. The safest thing to do is to shut down and then restart your browser, and then retype the URL into your browser's URL bar. Typing the correct website name is the best way to be sure you're not redirected to a phished website.

III. HOW THEY ATTACK

Almost all attacks of phishing use some form of deception designed to make a link in an email appear to belong to the phished organization. Misspelled URLs or the use of subdomains are common tricks used by phishers, such as this example URL, http://www.sbionline.com/. One method of spoofing links used web addresses having the '@' symbol, which is used to include a username and password in a web URL.

Phishing is different from worms or virus attacks that primarily use sophisticated technology to get this kind of valuable information. Phishing instead tries to fool you into handing over this information yourself. Because phishing relies more on targeting people than technology, it's sometimes referred to as a type of "social engineering" attack. Since phishing generally doesn't try to install malware like Trojan horses or keyloggers, regular antivirus and anti-malware may not help shield against it. But more sophisticated security suites do include "phishing filters" and web reputation services that can help protect you from spoofing attempts. Phishing is often sent out to thousands or millions of people as part of a spam attack, very often sent from zombie computers that are part and parcel of large botnets.

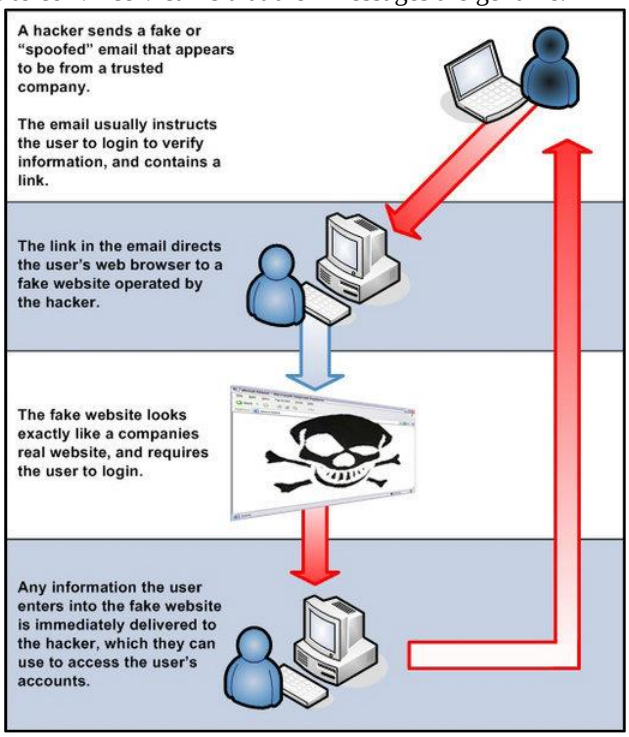
Most people associate spoofing with mail messages that spoof, or mimic, banks, credit card companies or other business like flipkart and snapdeal. These messages look genuine and attempt to get victims to reveal their personal info. But mails or messages are only one small piece of a phishing scam.
From beginning to end, the process consist of:

- Planning Phishers decide which business to target and determine how to get mail addresses for the customers of that business. They generally use the same mass-mailing and address collection techniques as spammers.
- Setup once they know their targets to spoof and who their victims are, phishers produce methods for delivering the message and collecting the data. Generally, this involves mail addresses and a Website.
- Attack this is the step people are most familiar with -- the phisher sends a phony message that seems to be from a famous sources.
- Collection Attackers record this info that victims enter into Websites or popup windows.

– Identity Theft and Fraud the attackers use the information they've gathered to make illegal purchases or commit fraud. Only three fourth of the victims fully recover.

If the phisher wants to coordinate another attack, he studies the successes and failures of the completed attack and begins the cycle again.

Spoofing scams take advantages of software and security weaknesses on both the client and server sides. But even the most high-tech spoofing scams work like con jobs, in which a hustler convinces his mark that he is reliable and trustworthy. Next, we'll look at the steps attackers take to convince victims that their messages are genuine.



They can attack in following ways:

A. By Email

The most common form of phishing is by e-mails. Pretending to be from your banks, or a legitimate retailer or government agency, the sender asks you to "confirm" your personal information for some phony reason. Typically, the email contains a link to a spoofed site that looks just like the real thing – with hi-tech visuals and images. In fact, the fake sites are close-replicas of the actual one, making it hard even for experts to distinguish between the real and fake websites. You enter your personal info onto the site – and into the hands of identity thieves.

B. By Phone

Attackers also use the phone to hunt for personal information. Some, posing as employers, call or send emails to people who have listed themselves on job search Web sites.

C. Something's Phishy If…

While spoofing scams can be sophisticated, the following features are often indicators that something is "phishy." Be aware of a potential scam if:

– Someone contacts you unexpectedly and asks for your personal information such as your account number, an account password or PIN, credit card number or Aadhar number. Legitimate companies and agencies don't operate that way.
– The attacker, who is a supposed representative of a company you do business with, asks you to confirm that you have a relationship with the company. This information is on record with the real company.
– You are warned that your account will be closed unless you "reconfirm" your financial information.
– Links in the mail you receive ask you to provide your personal info. To check whether a mail or call is really from the company or agency, call it directly or go to the company's Website.
– You're a job seeker who is contacted by someone claiming to be a prospective employer who wants your personal info.

D. Sample Phone Calls

1) Sample 1

"Is this Mr Raj? I'm calling from XYZ Bank. Do you have a Visa card? I need to verify your account number because it appears that someone may be fraudulently charging purchases to your account. Can you read me the account number and expiration date on the front? OK, now the last four digits on the back..."

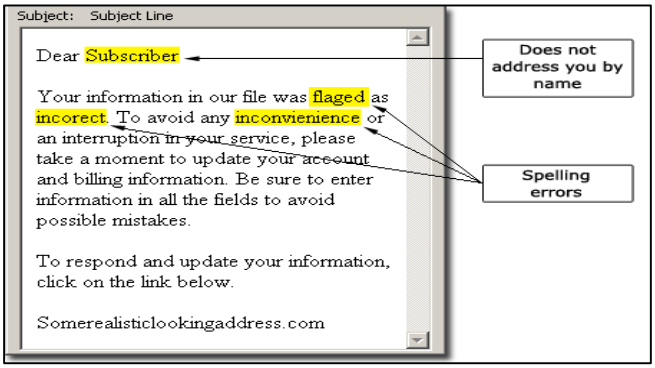
2) Sample 2

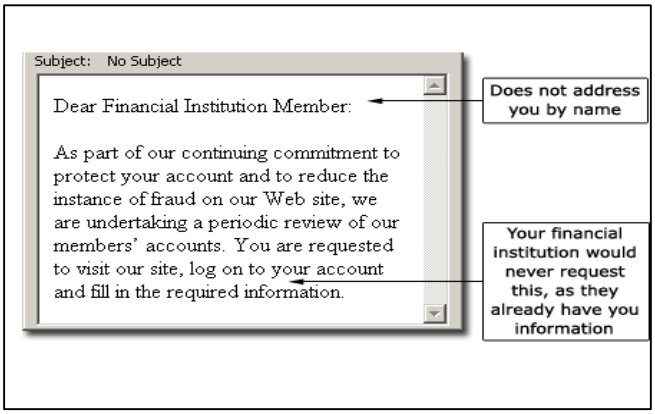
"Hello, Mrs Hudson? I represent the PQR Company and our records show that you have an overdue bill of $5000 plus interest and penalties. You don't know anything about this bill? Well, there could be a mix-up. Is your address 21B Baker Street? What is your Social Security number...?"

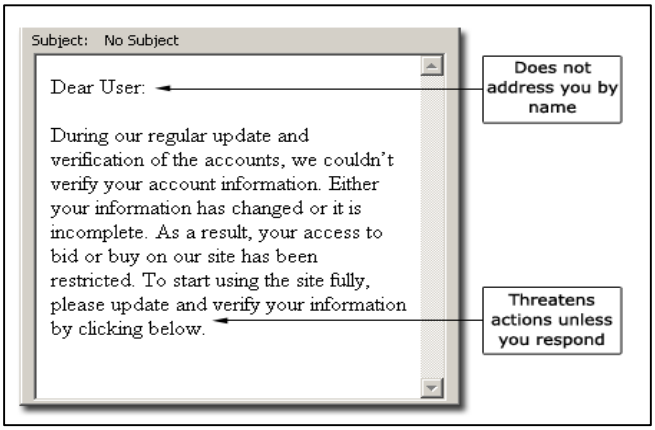
3) Sample 3

"This is Detective Conan calling from the Federal Consumer Agency. Are you Mr White? We have received several reports of telemarketing fraud involving attempted withdrawals from bank accounts in your area. In order to safeguard your account, we need to confirm your account number..."

E. Sample Phishing Emails







IV. OUR CONCEPT

One easy way is we check every URL that user enters from our database for Phishing. But with thousands of users at a time it will never be efficient instead we can match the URL against only certain selected ones.

A. History Pattern

Attackers have a group of people they like to phish that includes people with a little less technical knowledge etc.
There is a Pattern in these cases.

- A person with shopping hobby is more likely to be phished in a financial fraud on a fake Shopping website.
- A normal user might end up with identity theft by filling a fake credential form etc.
- Now there has to be certain set of websites he visited before and after he got phished, which provides us with a pattern.
- Let's say that x.com is a fraud website and a user gets phished. His history being u.com, v.com, w.com, x.com, y.com, z.com.
- We check for all the phishing case of x.com in our database finding one for example p.com, u.com, w.com, x.com, z.com.
- We can extrapolate that x.com has visitors visit u.com and w.com before reaching x.com.
- So if a user is visiting u.com we can check for all the possible cases like x.com for example and keep checking on REAL TIME basis.

B. Database

We will be using a database that will have all the past cases of phishing. In the first column we will have the user history (last three domains) before the fraud website and in the second column we will be having the name of the actual fraud website.

USER HISTORY	*FRAUD WEBSITES*
x.com , y.com	*z.com*
a.com , b.com , c.com	*fr.com*
w.com , s.com , r.com	*q.com*
er.com , e.com	*tar.com*
u.com , g.com	*y.com*
rt.com , x.com	*d.com*
ty.com	*po.com*

C. Reporting System

1) Manual System

In this system we can give the privileges to any user to report the phishing manually, so we can update our database accordingly.

2) Automatic System

In this System we assume that the user has anti phishing extension, now if the browser lands on a fraud website, the extension or the application will alert the user and report the history to our system that will update the database.

D. Alert System

A user will have a Real Time anti phishing enabled and whenever he is about to enter a fraud site, he will be alerted by the extension or the application.

E. Application Extension

User can embed an extension in his browser to report the phishing automatically or for alert against phishing.

F. Application Software

An Application Software could also be used for the same purpose meaning it will work like an anti-virus in the background and keep checking for phishing in real time.

G. Anti-Phishing (Working)

For anti-phishing we will have a database to have a track of our sites. In the database we will have different fields which will store information. In the first field we can have a pattern of sites which the user have visited. In the second field v can have a corresponding site name which is a phishing trap and on the third field we can have a remark field which give us the status about the site.

When a user will start browsing, the application or the extension in the browser will read the pattern & will try to match it with the patterns in the database. If the pattern is found it will then send the information about the phishing site.

H. Mathematical Model

1) Database at the server

This Database consist of the user history along with the reported phished web sites [p1, p2, p3, p4]. We take into consideration 8 hypothetical users for this scenario. The variable x, y, z etc. are references to the websites, user visited before jumping on the phished site.

User 1: [x, y, z, p1]
User 2: [p, q, r, p2]
User 3: [a, b, c, p3]
User 4: [l, m, n, p4]
User 5: [x, z, t, p1]
User 6: [x, r, c, p3]
User 7: [c, y, z, p1]
User 8: [i, h, n, p4]

After analysing the database, we'll extrapolate a phishing pattern 'P' pattern for all the phished sites [p1, p2, p3, p4].

P(Phished Site) = [site(frequency)]
P(p1)= [x(2),y(2),z(3),t(1),c(1)]
P(p2)= [p(1),q(1),r(1)]
P(p3)= [a(1),b(1),c(2),x(1),r(1)]
P(p4)= [l(1),m(1),n(2),i(1),h(1)]

In a general sense the above collection of data shows us the behavioural pattern of the victims of phishing attacks. Now we can create a mathematical relation to use this pattern to avoid future victims.

The following scenario explains the use of above equations.
'U' is an online user trying internet banking for the first time.
U has currently visited(C[]) x.com and z.com.
C(U)=[x, z]
So now we will create a list of possible attacks he might face.
H(U)=[p1,p3]
To calculate the probability percentage, we'll add the frequency and proportionate it.
Total frequency = x(2+1) + z(3)
Total frequency=6
Frequency for p1 , f(p1) = 5/6
Frequency for p3, f(p3) = 1/6

With this we can conclude that, there is 83.3% chances that user will go to p1, and a 16.7 % chance that he'll go to p2. Knowing this fact, we can improve user's safety by many folds.

V. CONCLUSION

This approach of classification based on history pattern can be a new arsenal to fight against the phishing spam. Furthermore, with just a few minor adjustments, the same tactic can be used against malwares or Trojans or any kind of malicious file. Similar to the history pattern for phishing, there will be a history pattern that will record user's browser history before he reached the particular site and use the similar working mechanism to alert the user.

VI. FUTURE SCOPE

Even though I've explained this technique with reference to Phishing, this can be implemented in malicious files, virus, and worms by using the similar database and creating a similar Pattern.

REFERENCES

[1] Jayakumar, N., Iyer, M.S., Joshi, S.D. and Patil, S.H., A Mathematical Model in Support of Efficient offloading for Active Storage Architectures.
[2] Naveenkumar, J. and Joshi, S.D., 2015. Evaluation of Active Storage System Realized through MobilityRPC.
[3] Naveenkumar, J., Bhor, M.P. and Joshi, S., 2011. A self-process improvement for achieving high software quality. International Journal of Engineering Science and Technology (IJEST), 3(5), pp.3850-3053.
[4] Salunkhe, R. and Jaykumar, N., 2016, June. Query Bound Application Offloading: Approach towards Increase Performance of Big Data Computing. In Journal of Emerging Technologies and Innovative Research (Vol. 3, No. 6 (June-2016)). JETIR.
[5] BVDUCOE, B., 2011. Iris Image Pre-Processing and Minutiae Points Extraction. International Journal of Computer Science & Information Security.
[6] Archana, R.C., Naveenkumar, J. and Patil, S.H., 2011. Iris Image Pre-Processing and Minutiae Points Extraction. International Journal of Computer Science and Information Security, 9(6), p.171.
[7] Kumar, N., Angral, S. and Sharma, R., 2014. Integrating Intrusion Detection System with Network Monitoring. International Journal of Scientific and Research Publications, 4, pp.1-4.

[8] Jayakumar, M.N., Zaeimfar, M.F., Joshi, M.M. and Joshi, S.D., 2014. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY (IJCET). Journal Impact Factor, 5(1), pp.46-51.
[9] Kakamanshadi, G., Naveenkumar, J. and Patil, S.H., 2011. A Method to Find Shortest Reliable Path by Hardware Testing and Software Implementation. International Journal of Engineering Science and Technology (IJEST), ISSN, pp.0975-5462.
[10] Namdeo, J. and Jayakumar, N., 2014. Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts. International Journal, 2(2).
[11] Jayakumar, N., Singh, S., Patil, S.H. and Joshi, S.D., 2015. Evaluation Parameters of Infrastructure Resources Required for Integrating Parallel Computing Algorithm and Distributed File System. IJSTE-Int. J. Sci. Technol. Eng, 1(12), pp.251-254.
[12] Salunkhe, R., Kadam, A.D., Jayakumar, N. and Thakore, D., 2016, March. In search of a scalable file system state-of-the-art file systems review and map view of new Scalable File system. In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on (pp. 364-371). IEEE.
[13] Naveenkumar, J., Makwana, R., Joshi, S.D. and Thakore, D.M., 2015. Offloading Compression and Decompression Logic Closer to Video Files Using Remote Procedure Call. Journal Impact Factor, 6(3), pp.37-45.
[14] Jayakumar, N., 2014. Reducts and Discretization Concepts, tools for Predicting Student's Performance. Int. J. Eng. Sci. Innov. Technol, 3(2), pp.7-15.