WEB PHISHING DETECTION
USING APPLIED DATA SCIENCE

TEAM ID : PNT2022TMID44694
TEAM MEMBERS : DHIVYA S (732519104004)
ABITHA S (732519104002)
VASUNTHARA R (732519104030)
PRIYATHARSHINI G (732519104703)
DEPARTMENT : CSE
COLLEGE NAME : SHREE VENKATESHWARA HI-TECH ENGINEERING COLLEGE, GOBI.

PROPOSED SOLUTION

Fields such as domain, sub domain, Top Level Domain (TLD), protocol, directory, file name, path and query allow creating different URL addresses. These related fields in the phishing URLs are generally different from the legitimate ones on websites. Therefore, URLs have an important place in detecting phishing attacks especially for classifying the web page quickly.

It was observed in the literature review that effective features obtained from the URL increase the accuracy of the classification. Additionally, third-party service usage, site layout, CSS, content, meta information, etc. features can also improve accuracy. However, these features will cause an increase in the classification time of the new websites which needed to be classified. The proposed model, trained only with the features obtained from the URL, is expected to classify in a shorter time than other models. Considering this information, only URL analysis is planned in the study. Thus, the classification results of the obtained features in different algorithms in machine learning are compared. In addition, the results from another study with the same dataset are compared with those of the current study.

NOVELTY

Attackers take advantage from published phishing counter measure to modify their websites and bypass detection systems. We propose 27 novel features with relevant information to achieve high performance for today's phishing detection tasks. We include a novel type of feature, web technology analysis.

SOCIAL IMPACTS

Stolen data can have many uses. Master card information will be accustomed purchase goods and services, ATM card information may well be accustomed duplicate ATM card sand use them for withdrawal of money. Account be accustomed steal information or to be ready to act as another user online.

FEASIBILITY

This is the first attempt to systematically understand the threat posed by the ease of cor-relating user information across caller ID lookup application (True caller), and social networking application (Facebook). This was executed using phone num-as unique identifiers. We show the attack is feasible with easily available computational resources, and poses a significant security and privacy threat. An attacker can use these cross-application features to launch highly targeted attacks on multiple channels like OTT applications, voice, e-mail, or SMS

SCALABILITY

Indian phone numbers that we enumerated, it is possible to launch social and spear phishing attacks against 51,409 and 180,000 users respectively. Phishing attacks can be launched against 722,696 users. We also found 91,487 highly influential victims who can be attacked by crafting whaling attacks against them.

BUSINESS MODEL

The main reason is the lack of awareness of users. But security defenders must take precautions to prevent users from confronting these harmful sites. A business is aware that a spoof site has launched, the next step is to alert customers to ensure they don't visit the fake website and enter credentials. But organizations still need to provide a countermeasure in case customers aren't notified in time. As deception technology matures, defenders have new ways to foil phishing, even if hackers have managed to gather victim credentials. One approach involves injecting the spoof site with decoy credentials. Decoys are highly convincing fake credentials that lessen the value of stolen credentials to the point where the attacker is unsure if they've

taken anything they can use. An email-focused strategy fails to extend protection to an organization’s customers. All one needs to do is look at phishing attacks against British Airways or American Express to see that consumers are an extremely vulnerable group. An email-centric strategy can't protect customers because they’re all using different email providers that a company cannot monitor, much less control. It’s also not possible for a large enterprise to train every one of its customers to recognize a phishing attack..