PROBLEM STATEMENT FOR WEB PHISHING DETECTION

PROBLEM

Phishing attacks are becoming more and more sophisticated, and our algorithms are suffering to keep up with this level of sophistication. They have low detection rate and high false alarm especially when novel phishing approaches are use. The blacklist-based method is unable to keep up with the current phishing attacks as registering new domains has become easier. Moreover, comprehensive blacklist can ensure a perfect up-to-date database. Various other techniques such as page content inspection algorithms have been used to combat the false negatives but as each algorithm uses a different approach, their accuracy varies. Therefore, a combination of the two can increase the accuracy while implementing different error detection methods.

BACKGROUND

The amount of trade conducted electronically has grown extraordinarily with widespread Internet usage. Electronic marketing or e-commerce consists of the buying and selling of products or services over electronic systems such as the Internet and other computer networks. This type of facilities is growing rapidly over internet to comfort human being living. In this electronic world, internet is an easy and fastest way of payment for a service.

People want reliability in every place of life. And when it's a matter of money they
want more security for transaction. Computer security has thus become a major concern for merchants and e-commerce service providers, who deploy countermeasures such as firewalls and anti-virus software to protect their networks. Phishing is another danger, where consumers are fooled into thinking they are dealing with a reputable retailer, when they have actually been manipulated into feeding private information to a system operated by a malicious party. There are several different techniques to combat phishing, including legislation and technology created specifically to protect against phishing. One strategy for combating phishing is to train people to recognize phishing attempts, and to deal with them. Education can be effective, especially where training provides direct feedback. The Anti-Phishing Working Group, an industry and law enforcement association has suggested that conventional phishing techniques could become obsolete in the future as people are increasingly aware of the social engineering techniques used by phishers. Another effective strategy is the technical contribution to recognize phishing attempts. Anti-phishing measures must be provided to the users for e-commerce

or online transaction security. Our attempt is to provide an automated technical support to the users by detecting phishing web-pages. We found the procedure of phishing and worked to detect them and then publish for user verification.

OBJECTIVE

Nowadays Phishing becomes a main area of concern for security researchers because it is not difficult to create the fake website which looks so close to legitimate website. Experts can identify fake websites but not all the users can identify the fake website and such users become the victim of phishing attack. Main aim of the attacker is to steal banks account credentials. Phishing attacks are becoming successful because lack of user awareness. Since phishing attack exploits the weaknesses found in users, it is very difficult to mitigate them but it is very important to enhance phishing detection techniques. Phishing may be a style of broad extortion that happens once a pernicious web site act sort of a real one memory that the last word objective to accumulate unstable info, as an example, passwords, account focal points, or MasterCard numbers. all the same, the means that there square measure some of contrary to phishing programming and techniques for recognizing potential phishing tries in messages and characteristic phishing substance on locales, phishes think about new and crossbreed procedures to bypass the open programming and frameworks. Phishing may be a fraud framework that uses a mixture of social designing what is additional, advancement to sensitive and personal data, as an example, passwords associate degree open-end credit unpretentious elements by presumptuous the highlights of a reliable individual or business in electronic correspondence. Phishing makes use of parody messages that square measure created to seem substantial and instructed to start out from true blue sources like money connected institutions, online business goals, etc, to draw in customers to go to phoney destinations through joins gave within the phishing websites.

CONCLUSION

we have seen how phishing is a huge threat to the security and safety of the web and how phishing detection is an important problem domain. We have reviewed some of the traditional approaches to phishing detection; namely blacklist and heuristic evaluation methods, and their drawbacks. We have tested two machine learning algorithms on the Phishing Websites Dataset and reviewed their results. We then selected the best algorithm based on its performance and built a Chrome extension for detecting phishing web pages. The extension allows easy deployment of our phishing detection model to end users. We have detected phishing websites using Random Forest algorithm with and accuracy of 97.31%.

For future enhancements, we intend to build the phishing detection system as a scalable web service which will incorporate online learning so that new phishing attack patterns can easily be learned and although the use of URL lexical features alone has been shown to result in high accuracy (97%), phishers have learned how to make predicting a URL destination difficult by carefully manipulating the URL to evade detection. Therefore, combining these features with others, such as host, is the most effective approach.