

Prediction and Real-Time Detection of Phishing URLs/Domains

Project Description:

Phishing is a social engineering attack where attackers deceive users, often directing them to malicious websites with a look and feel similar to popular sites (e.g. Paypal.com), to steal their sensitive information such as usernames, passwords and credit card information. Phishing costs billions of dollars worldwide including Qatar and is on the rise. This project aims to help reduce phishing attacks.

The project consists of two components, but both are related as they use similar techniques identify phishing domains/URLs:

- a. Prediction of phishing domains given a set of legitimate domains (e.g. qcri.com.qa) :
With the ability to predict the phishing domains which are likely to be misused by Internet miscreants, organizations can take proactive measures to protect their users from such domains.
- b. Real-Time detection of phishing URLs:
Using the real-time streaming data feed on Google Transparency Logs we have, we want to detect URLs most likely to be Phishing URLs. We would also like to understand how many of them are related to Qatar.

Duties/Activities:

- Design and implement phishing techniques (e.g. combosquatting, typosquatting) based on the latest research publications (student will be guided to identify and understand the relevant research papers)
- Design and implement a web-based system to demonstrate the usage of the system and visualize phishing domains

Required Skills:

- Design and implement algorithms (Python is the preferred language of implementation) – Software Engineering skills in general
- General understanding of computer networks
- Working knowledge of Linux operating system (e.g. Ubuntu)
- Familiarity with web application development
- Familiarity with database systems

Preferred Intern Academic Level:

Senior level under graduate or master's student

Learning Opportunities:

- Learn Phishing techniques used by Internet miscreants (malicious activities in general)
- Design and implement a practical system
- Application of data structures and algorithms
- Application of machine learning algorithms
- Design and implement a web based system (client-server architecture)
- Dealing with large amount of real-time data (scalable algorithm design)

Expected Team Size: *it is preferable to have team projects*

It is a single person project.

Mentors

Name: Mohamed Nabeel

email: mnabeel@hbku.edu.qa