DTonomy Internship Project Proposal

06.14.2020

R. Kevin Oberlag

Summer 2020

DSA 5900

Credit Hours: 4

Company: DTonomy

Company Sponsor: Peter Luo

Introduction

The project being proposed will be used to meet the Masters of Data Science and Analytics *Professional Practice* curriculum requirements at the University of Oklahoma. Specifically the project will be performed for and under the supervision of the cybersecurity company, DTonomy. DTonomy aims to solve critical problems that security teams must deal with on a daily basis, utilizing automation and artificial intelligence to enable smarter organization of security alerts and incidents. The company describes their technology as follows:

> DTonomy's technology includes a unique adaptive learning engine which continuously learns, provides contextual insights that are not easily discoverable, and makes relevant recommendations and automated workflows to guide IT teams through steps and procedures, resulting in up to 10 times quicker resolution of incidents, decreased downtime, and reduced alert fatigue for staff.

In short, the business objectives of DTonomy are to deliver meaningful insights of security related threats and alerts to customers, reducing the level of convolution while also providing contextual recommendations that would otherwise be difficult to obtain by an individual or team. This project will attempt to expand these insights by integrating with email providers, such as G-Mail and Outlook, which have the potential to reach a larger user base, therefore leading to a potentially larger customer base of DTonomy's main platform.

Objectives

The insights and recommendations discussed above are currently delivered through DTonomy's online platform, however, it would be desirable to also deliver meaningful insights closer to the source of some security incidents, such as through email platforms. Phishing emails are one of the many issues that security teams must face, so the goal in working with DTonomy will be to build an add-in for applications such as G-Mail and Outlook, which will deliver direct access to users within each respective platform to meaningful security insights, such as IP Address, URL extraction, and WHOIS information.

The objectives include software development of the add-in application, as well as feature extraction of important indicators in suspicious emails, visual representation of relevant information, and implementation of a recommendation engine to recommend relevant insights.

I. Email Add-in software development

Specifically the add-in will initially have the features of extracting URLs from an email, and allowing the user to view a screenshot of each web page in question, without actually visiting the web page. In addition, the user will be able to view the WHOIS website data for the domain of each url, in order to determine if the hyperlink in their email seems legitimate. The email add-in will involve developing a new application, which utilizes the frameworks provided by the email providers. Along with learning the specific framework, additional skills of working with APIs will also be required. The application itself will utilize third-party APIs for capturing screenshots of web pages, and for providing WHOIS information. Therefore, there are several aspects of software engineering to be learned and improved on via this portion of the project.

II. Feature extraction in suspicious emails

This objective will be a more analytical approach, requiring data extraction and manipulation skills, along with exploratory data analysis, in order to find relevant features for insights into suspicious emails. I will be conducting analysis on data sets that contain data collected from “Suspicious” flagged emails. Once important features are determined, this knowledge will be applied to the add-in for more user insights. This will provide an excellent learning opportunity to apply analytic skills learned in the DSA program.

III. Visual representation of information

Add visualizations to add-in, such as a map, for visualizing geo-data related to the ip address of the site. Creating visualizations to better understand a set of data is an important skill of a Data Scientist. This step will be an opportunity to learn and develop more skills for visualizing data.

IV. Implement a recommendation engine

Provide recommendations on what actions users should take with explanations. This will utilize insights from objective II, and be programmatically implemented into the add-in.

Plan

I. Email Add-in software development

A. Read documentation and learn API of email add-in framework, while also learning recommended application development practices of the platform.

B. Develop functionality for extracting URLs from selected email messages, and displaying them as selectable options in the add-in user interface.

C. Setup API integration with urlscan.io, in order to retrieve a screenshot of the chosen web page.

D. Setup UI for displaying the screenshot.

E. Setup API integration with whois.com, in order to retrieve the WHOIS information of the chosen web page's domain.

F. Setup UI for navigating to WHOIS information.

G. Test add-in

H. Research deployment to marketplace and sharing of add-in.

II. Feature extraction in suspicious emails

A. Research publicly available datasets and kernels related to phishing emails.

B. Perform exploratory data analysis on data sets, in order to determine important features that help to define a phishing email.

III. Visual representation of information

A. Use third-party APIs to pull geo-data from emails.

B. Apply geo-data to mapping visualizations and implement them into the add-in.

C. Test add-in.

IV. Implement a recommendation engine

A. Conduct research on possible remediation steps for phishing emails.

B. Implement a rule engine for the add-in, to suggest relevant remediation for a given email, utilizing the feature insights from objective II.

C. Test add-in.

Deliverables

I. Email add-in application

An initial version of the email add-in will be developed, tested, and possibly deployed for users of the respective email add-in platforms. Development will include research of documentation, including email add-in frameworks and third-party APIs, writing JavaScript code to build the application, code reviews and discussion, and application testing. Research will also be performed and shared, on how to deploy the add-in to the respective email platform's marketplace.

II. An analysis report on important features of phishing emails

Analysis will be performed on various phishing email datasets found online in order to determine relevant features of phishing emails, and delivered in an organized and documented report.

Schedule

The four objectives are roughly planned to be completed every 1.5 weeks, leading to a project completion of July 15th.

I. Email Add-in software development

To be completed on or before June 12th

II. Feature extraction in suspicious emails

To be completed on or before June 24th

III. Visual representation of information

To be completed on or before July 3rd

IV. Implement a recommendation engine

To be completed on or before July 15th