Project Initiation Document

Fall 2016

Daniel Smith

A. Project Scope Statement

For this project, the sponsor, ASRC Federal Mission Solutions, will be provided with code, research, and analysis of a tool which performs scanning functions on employee emails. The goal of this tool is to minimize human error in sending confidential/sensitive information which is a mistake that can ultimately lead to an employee's termination. Code for the project will focus on working algorithms which can process emails and determine if they contain predefined confidential information. The end goal for the code of this tool is to package it for use as a plugin with existing email clients. Research involved in this project will be documented and submitted as reference for the sponsor. This includes any and all open source code utilized as well as developed algorithms. Performance analysis will be conducted and reported. The necessary CPU, Memory, Disk Space, and Network Utilization will be recorded and graphed.

Not included in the scope of this project is code maintenance following the final increment of the project. Any and all code maintenance will be performed by the sponsor or outsourced by the sponsor at their discretion. Additional infrastructure, such as servers, necessary for the project is the responsibility of the sponsor after receiving requirements from the software developers. The product will be developed incrementally, and it is up to the discretion of the sponsor to determine when the product will be put into product in their environment. Each increment will be a functioning tool.

B. Project Background

This project was initially proposed at ASRC Federal Mission Solutions by James Warner in the winter of 2016. The project was not activated until September 2016 in conjunction with the Rowan Collaboration Program which links ASRC with Rowan Software Engineering students to complete a product. James Warner as the author of the idea for the tool is the lead contact point

for the project at ASRC and acts on the behalf of their best interests in order to allow Rowan Collaboration students to complete the project within the given time frame of a school semester.

The project was proposed because ASRC Federal Mission Solutions has employees who work on contracts from the Department of Defense (DoD). These contracts frequently deal with proprietary intellectual information which should not be exposed to insecure networks or people without proper clearance. Because email is an inherently insecure protocol, no data or mention of sensitive information should be passed using this conduite. In cases of human error, such information can inadvertently be sent in an email and ultimately lead to an employee's termination. To curb this the email scanning tool will act as a safeguard to prevent these mistakes from happening as frequently. At the end of the day, the tool cannot be perfect, but the intention is for it to act as a backup to human vigilance.

C. Feasibility Analysis

The tool that is to be created in this project is an e-mail scanner. E-mail scanners have been been in existence for at least a decade with their primary use being to analyze e-mails for indicators of spam. That makes the theories and algorithms for this project based in the well studied practice of detecting spam. With e-mail spam, the delivery and implementation continuously evolve, this makes it inherently more difficult to detect, but with the scanning of e-mails for confidential information, there is an established pattern in the sensitive terms as well as the natural flow of language from the sender. This makes it easier to write a more accurate algorithm to pinpoint potentially errant entries of confidential information in insecure emails. Overall, the project is one-hundred percent possible based on prior research and modern implementations of spam email filters.

D. Assumptions, Dependencies, and Constraints

The initial assumption for this project is that the language of the email is English. Without this assumption the project becomes impossible to complete without a stronger handle on foreign languages. Additionally, the assumption is that all emails are sent following the Multipurpose Internet Mail Extensions (MIME) protocol. This ensures the tool focus on a well established standard.

The largest dependency for the project is the proper input of starting data by ASRC Federal. Without proper insertion of data on their end the tool will have no data to work with. The development versions of the tool will include dummy data and tests, but the final implementation will depend on proper input of data by an ASRC Federal employee.

For the end product of the tool, the implementation is constrained by Microsoft Outlook's plug-in framework. Because employees at ASRC Federal use Outlook as their e-mail client, the tool will ultimately need to be integrated with this client in order to perform the pre-send scan.

E. Initial Project Plan

Sprint 1 - 10/14/2016

- Practically impossible to reverse engineer encryption
- Open source software research

Sprint 2 - 10/29/2016

- Word Stemming
- Rough detection algorithm

Sprint 3 - 11/11/2016

- E-mail scoring
- Algorithm improvement (speed, consistency)

Sprint 4 - 12/2/2016

- Artificial intelligence algorithm
- User interface

Sprint 5 - 12/16/2016

- Outlook Plug-In

F. Organization and Governance

The project team for this project is structured in the style of Agile Scrum. The team consists of six members, a product owner, scrum master, and four development team members. In typical Agile Scrum layouts the product owner and scrum master are independent of the development team. Due to the nature of this project in that it consists entirely of undergraduate computer science students from Rowan University, there is foreseeable overlap in the product owner and scrum master assisting the development team with code due to their particular knowledge and experience.

In Agile Sprint, each two week sprint ends with a sprint review which acts as a check between the sponsor at ASRC Federal and the project team. During the sprint review, the product owner presents an increment of the product to the business and establishes that the product is moving in the right direction. This also acts as a check to ensure that the development team is interpreting requirements correctly.

G. Communication Plan

The project team will conduct what is considered daily scrum meetings on Mondays, Wednesdays, and Fridays during each sprint. These meetings will last for fifteen minutes and establish what has been accomplished since the last daily scrum. Minutes for these meetings

will be taken by the scrum master for purposes of ensuring proper scrum protocol is upheld. Additional communication will occur between the project team via digital messaging technologies. For sprint reviews, GoToMeeting will be used to share computer screens to demonstrate increments and highlight information and an ancillary phone conference system will enable audio for the conference.

H. Quality Plan

To be provided to James Warner:

- Research summary of open source software used
- Junit testing for code
- Performance Whitepaper
- Accuracy Whitepaper

I. Risk Assessment

The largest risk in the creation of this tool is the creation of an unforeseen attack vector on the encrypted database of sensitive terms. This could allow a malicious entity to decipher the encrypted terms. To mitigate this, one of the largest requirements for the tool is security of the database and preventing feedback to the user of which words were perceived as sensitive.

J. Estimated Cost ROI

The cost to ASRC Federal in this project is minimal. The largest cost to them is the time that James Warner spends assisting the Rowan Collaboration students. Because the students are completing the project for a class, there is no cost to be paid to the developers or development team. This means that while ASRC Federal's investment is small the return could be a fully working product.