Carnegie Mellon University
Software Engineering Institute

Technical Approaches to Bystander Engagement

SEPTEMBER 21, 2023

Dan Costa, CISSP, PSEM, GCITP

Document Markings

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM23-0935

Agenda

Where Does Technology Fit in Bystander Engagement?

- Supervisor / Manager Appraisals
- Suspicious Contact Report Filtering

Background / Context

The workforce is one of the most, if not the most, powerful sensors for insider risk management.

However – we're asking a lot of these sensors:

- Numerous reportables
- Each with their own reporting mechanisms
  - Usability concerns abound
- "Other duties as assigned"

Where Does Technology Fit?

Some Options:

- Remove the human from the reporting loop
  - Limited applicability, maximum impact
- Technology-driven prompts to potential reporters
  - Some probabilistic model predicts the presence of a reportable, the reporter is asked to confirm
  - We're already using the "check with the user" method in other security functions (e.g. removable media)
- Leverage technology to minimize reporting burden on the reporter
  - Maximize user experience to increase reporting fidelity
- Nowhere!
  - There are some things better left handled low-tech

Suspicious Contact Reporting Overview

32 CFR 117.8(c)(2) *Suspicious contacts.* Contractors will report information pertaining to suspicious contacts with employees determined to be eligible for access to classified information, and pertaining to efforts to obtain illegal or unauthorized access to the contractor's cleared facility by any means, including:

(i) Efforts by any individual, regardless of nationality, to obtain illegal or unauthorized access to classified information.

(ii) Efforts by any individual, regardless of nationality, to elicit information from an employee determined eligible for access to classified information, and any contact which suggests the employee may be the target of an attempted exploitation by an intelligence service of another country.

Typical Suspicious Contact Reporting Workflow

Threat Actors

Benign Email Senders

Employee

Employee's Security Office

Issues With the SCR Workflow

- Some people don't **report** enough of the suspicious contacts they receive
- Some people don't ***see*** enough of the suspicious contacts they receive
  - E.g., spam filters deployed at the enterprise level, email filtering systems deployed at the individual user level
- Human review time is costly and error-prone

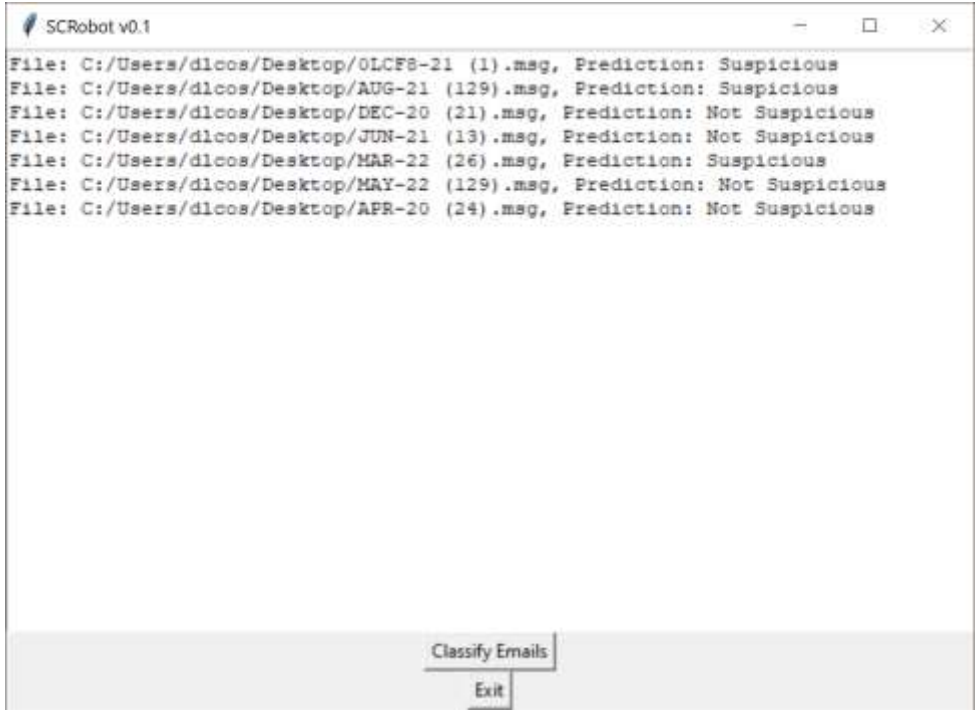
Enter SCRF: The Suspicious Contact Report Filter

- Machine learning model that predicts whether an email message is suspicious
- Model built on 2,566 potential suspicious contact report email messages double-coded by security staff
  - 1123 (43%) of these emails were deemed suspicious
  - Inter-rater reliability of 84%
    - This figure was used as the proxy measure for the target accuracy for the model – if it's as good as having two security people review it, we're happy with it

So, What Actually Makes An Email Suspicious?

Features of Concern	Mitigators ("Good" Features)
Sender identifies as self-funded	Sender known by security
Font changes implying copy-pasted script	Organization known by security as reputable
Form letters	Exchange initiated by employee, or employee responds positively
Sender email address not matching message signature	Long email threads with conversational, familiar tone
Stilted, overly formal message tone	Newsletters, group emails
Mismatched fields of study in message / attachments	
Self-identifying by countries of concern	
Unsolicited requests	
Resume / CV attached without prompting	

How's It Working?



- ~300 Emails reviewed by SCRF once deployed
  - 87% Accuracy – 3% better than two humans reviewing SCR emails
- 93% improvement in human time needed to review emails

Future Improvements

- Support for non-English language in message contents and attachments
- Augment the model with additional features of concern
- Explore adapting this approach to similar problems
- Place the model in front of the email spam filter

Suspicious Contact Reporting - Future Workflow

Threat Actors

Benign Email Senders

SCRF

Security Office

Questions / Discussion

Presenter Contact Information

Dan Costa
Technical Manager, Enterprise Threat and Vulnerability Management
CERT Division, Software Engineering Institute
Carnegie Mellon University

Telephone: +1 412.268.8006
Email: dlcosta@sei.cmu.edu