PhishED:
Automated contextual feedback for reported Phishing

Adam Jenkins, Nadin Kokciyan, Kami Vaniea; University of Edinburgh
{adam.jenkins, nadin.kokciyan, kami.vaniea}@ed.ac.uk

Project Overview

People often receive **suspicious emails** that claim to be from trusted persons or organisations (e.g., a bank), but are actually from attackers aiming to deceive recipients into giving away valuable information, known as **phishing**.

In this project, we are developing the **PhishED** system that will support people by providing **automatically generated advice** to suspicious emails they report. Using contextual cues the advice is meant to both help make an **informed decision** and **provide education** in a teachable moment.

Proposed Solution

PhishEd leverages Artificial Intelligence's ability to extract and reason about contextual features of phishing in support of user decision making. Example of such features are::

Contextual keywords - If a reported email uses terms like "shutdown", "email", and "account", the user may think that such a message was sent through an organization.

Uniform Resource Locators (URLs) - If an email contains the "PayPal" keyword but the URLs do not lead to PayPal's official website, that information can be provided to the user.

Email headers - E.g. the From address can be checked against DKIM signatures or whitelists of organisation domains.

Template Initial Designs

MSc Thesis: Design of Auto Responses
Created initial designs for PhishEd, using a **user-centered approach** creating final mock up, shown in Figure 2.

NEAT & SPRUCE
We will improve our template designs by integrating **Microsoft security warning** design guidance, designed for use within their own product teams, see Figure 1.

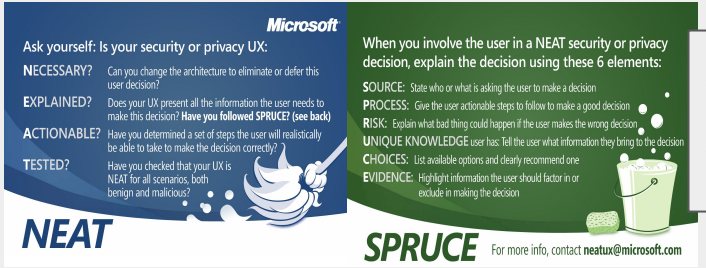

Figure 1: Microsoft's NEAT and SPRUCE guidelines for Security Warning Design

Figure 2: Initial design of automated responses by Zeyu Zhang (MSc)

Executive Summary, provides the range of contextual features in reported email

Traffic Light Colour Scheme to highlight risk levels or facts

Sender email address, highlighting domain and ***authentication information*** from headers

URL breakdown for example age of the domain

Language cues explained to the user

Provide reassurance, encourage using **unique *knowledge*** to make an informed decision

Future Work

Iterative Design
Improve on our initial designs using a **User centered design** process.
Focus groups with potential users to inform designs.

Evaluation
Lab studies will identify suitability with potential users.
Develop an **Outlook add-in**, and then deploy and monitor the usage of tool in a **longitudinal study** with partner organisation.