CASE STUDY

SMART SHIELD: Artificial Intelligence Anti-Phishing System

CYBER SECURITY COOPERATIVE RESEARCH CENTRE

PROJECT SNAPSHOT

Smart Shield is an anti-phishing solution that uses advanced machine learning algorithms to examine and learn from phishing emails and malicious websites, enabling greater detection of advanced phishing attempts.

WHO'S INVOLVED

This project was funded by the Cyber Security CRC and the Government of Western Australia. R & D has been undertaken by CSIRO's Data61.

WHAT'S THE ISSUE?

Malicious emails are one of the most commonly reported cyber incidents. While there are many products on the market to detect and stop phishing, attackers continuously and rapidly evolve to evade detection. Hence, even a small number of phishing attempts may pose a risk for organisations.

WHAT WERE THE OBJECTIVES?

Examine and learn from phishing emails and malicious websites;

Better understand prevalent phishing threats;

Find a solution that offers better protections against phishing emails

TECHNOLOGY DESCRIPTION

Unlike many existing solutions, Smart Shield system learns not only from email and web pages themselves, but also via feature correlation with alternative trusted entities. These evaluate email sender reputation, goal and consistency, using advanced techniques like graph neural networks, and natural language processing to extract and classify multi-dimensional features. These can detect more sophisticated unseen phishing attacks. Smart Shield scans incoming emails and provides traffic light banners on top of each email indicating the warning level- red, yellow, or green. The banner also provides information on what aspects of the email triggered the warning.

The Smart Shield system is cloud-ready. It can be deployed in an organisation-controlled public or private cloud infrastructure.

APPLICATION

Smart Shield system deployed in AWS is mature, fail-safe and scalable. It is being tested in a small pilot with participants, primarily within the distributed systems security lab of CSIRO Data61. In terms of usability, the feedback so far has been 100 per cent successful.

Smart Shield uses both contemporary and novel algorithms to detect phishing. The contemporary algorithms are trained with approximately 38,000 employee reported emails, and 20 million phishing and benign URLs that are relevant to Australia. The employee reported emails were primarily received from CSIRO, representing over 5 years of phishing and spam emails received by their employees. Additional email samples were received from the WA Government and used for independent testing. This approach ensures that our algorithms learn from data that is relevant to us.