Vulnerability of ID-Based Email Systems to Social Engineering Attacks

Abhi Siri, Maksym Yarosh, Rached Arda, Mannan Syed, Dr. Ihab Wattar, PhD

Cleveland State University

Choose Ohio First

ABSTRACT

When identifying targets for phishing campaigns, the use of scraping tools, such as theHarvester, is necessary, as most organizations use name-based email systems. When using ID based systems, however, it becomes possible to calculate target addresses en masse with relatively low input – which this project aims to demonstrate.

OBJECTIVES

- Compare the effectiveness of reconnaissance via web scraping tools vs. algorithmic target generation against an organization using an ID-based email system
- Gain insight into how modern email systems prevent mass email campaigns and attempt to bypass the relevant safeguards.

PROCEDURE

- Strategy
  - Copy a legitimate campus communication on a Gmail account, getting users to visit a cloned version of a CSU page (myCSU, see Fig. 2)
- Infrastructure
  - Leveraged HTTrack to clone the myCSU site
  - Cleaned all external links, redirected all site links to an “exit form” for data gathering
  - Generated bit.ly links to track link clicks
- Reconnaissance
  - Via Web Scraping:
    - Leveraged theHarvester on “vikes.csuohio.edu”
    - Validated output emails
  - Algorithmically:
    - Wrote a simple Java program to determine all possible addresses given the first 4 digits of a student ID
    - Tried 5 separate sets of starting digits
    - Validated output emails in samples of 50 emails
- Execution
  - Used a generic email builder (Unlayer) to create a convincing university communication (see Fig. 1)
  - Introduced sender to the CSU environment by simulating correspondence with group members
  - Take 1: Single mass e-mail
    - 500 recipients, all at once
  - Take 2: Dispersed e-mails
    - Sent individually to sets of 10 samples

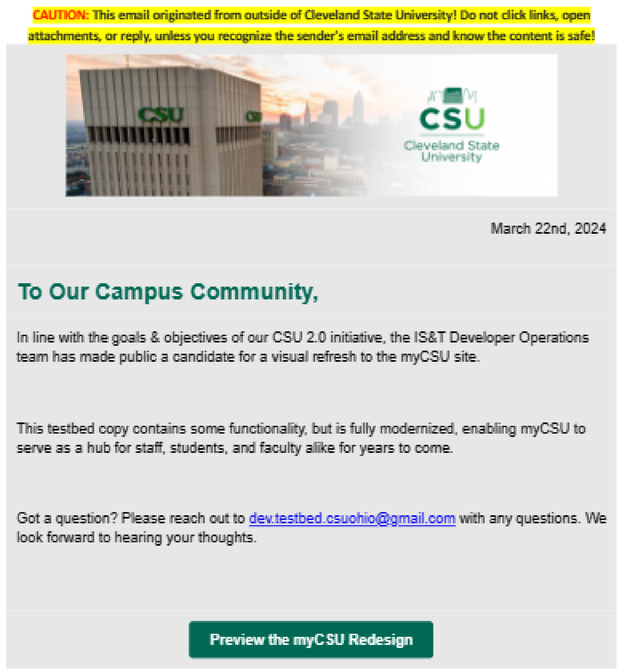

Figure 1. The phishing email sent out to all targets.

RESULTS

- Reconnaissance Phase (see Fig. 3)
  - theHarvester yielded 35 addresses, 18 of which were valid (51%), with a 10 minute runtime
  - Java program yielded 5,000 potential addresses, with 537 valid ones (11%). Validation took ~3 hours manually, but could have been automated
- Execution Phase
  - Both attempts took several hours to send, and only went to accounts that had interacted with the account prior to the campaign
  - Likely blocked by Gmail, not CSU – CSU email and external email were both CC’d on all campaigns

Figure 2. A screenshot of the myCSU clone.

DISCUSSION/FUTURE WORK

Area for future work include:

- Exploring strategies to deliver content straight to inboxes (better message crafting, spoofing, etc)
- Executing bait attacks to better simulate real threat actors and increase sender-organization interaction
- Automating target validation to reduce process time

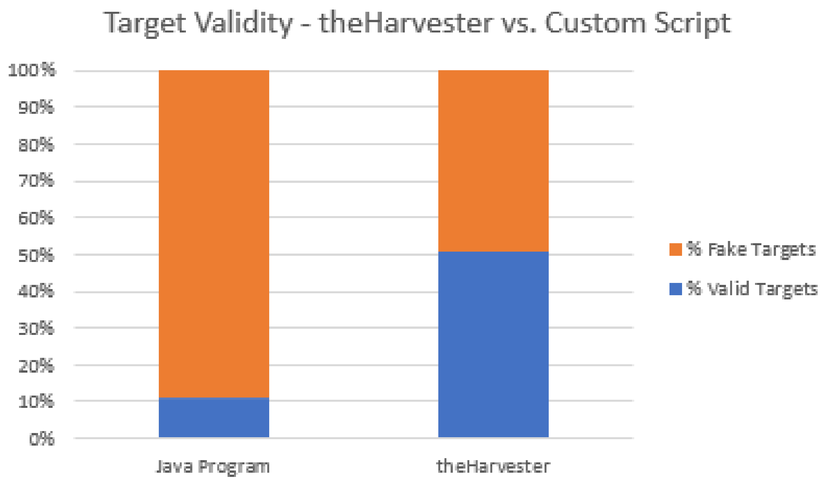

Figure 3. A chart depicting the percent validity of each reconnaissance method

CONCLUSIONS

Despite the longer completion time (which can be automated), algorithmic target generation is a powerful tool that enables mass email campaigns in organizations with numerical or simple alphanumeric email address conventions.

However, it may still be advantageous to use web scraping over an ID generation script, as web scraping tools had a significantly higher percentage of valid results.

To conclude, it is wise for all organizations to minimize potential avenues for attack as much as possible. ID based email systems are but another potential threat vector – which is why care must be taken to institute compensating email security tools to prevent any phishing campaign – no matter how its targets were generated.