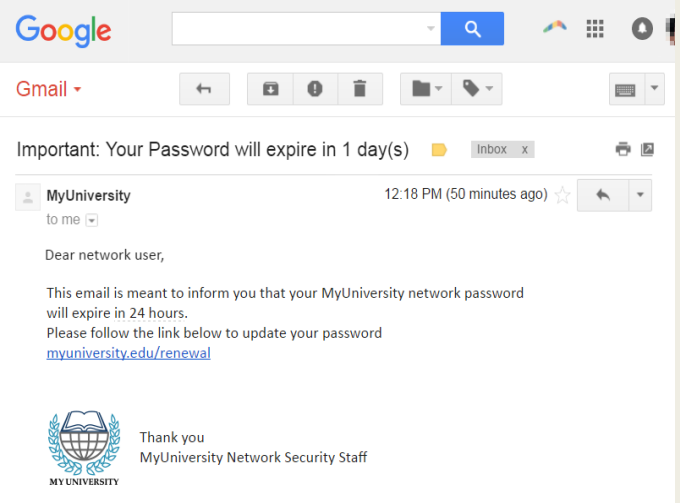
Ethical Implications of AI & Phishing

Emily Herron, Kevin Chen, Will Berger

What is Phishing?

Phishing is a form of hacking that tricks victims into giving up sensitive information (passwords, IDs, documents, etc.) by posing as a legitimate source.



Phishing is all about imitation. Phishing emails generally use two techniques: 1) creating a sense of urgency to trick you into clicking on a malicious link, or 2) attempting to 'blend in' as an innocuous email that preys on your curiosity.

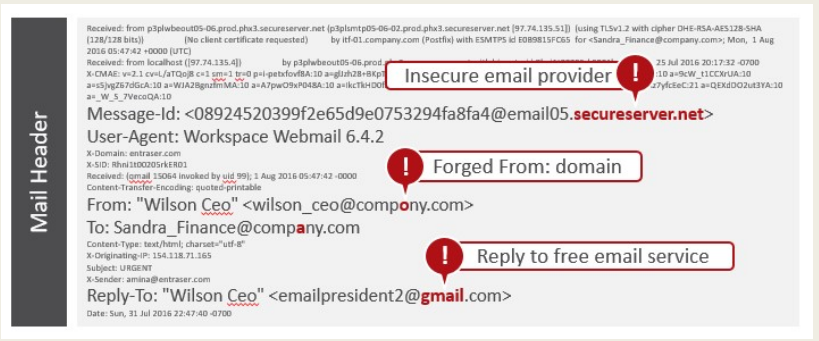
Vanderbilt's Email Quarantine

After interviewing Vanderbilt's IT department, they stated that Vanderbilt uses both Gmail's and Microsoft Office 365's third-party proprietary spam filters to eliminate spam. These services inspect the subject line and the email's contents to determine if the email is to be released into the user's inbox or to be quarantined for 15 days and then deleted. The quarantine is not accessible via students' Gmail accounts and requires a separate log-in from a portal listed on the Vanderbilt website.

Many job interview invitations were marked as spam and sent to the quarantine, effectively locking them away without alerting students. When Vanderbilt added Microsoft's quarantine spam filter in the fall of 2018, Vanderbilt may have actually done more harm to students and faculty than good by inadvertently increasing the false positive rate (missing important emails) at the expense of trying to decrease spam entering your inbox.

How to Protect Yourself:

- Always remember to question emails, and if something seems off, there s probably a reason.
- Check for spelling and grammatical errors in the email address of the sender, the subject line, and the actual content of the email.
- Hover over (but don't click on) hyperlinks that look suspicious to see where they lead.
- Install a phishing filter on your email application and also on your web browser. These filters don't keep out all phishing messages, but they do reduce the number of phishing attempts [7].



Ethical Implications

AI on both sides of the phishing battle has effectively created an AI arms race. The tools and techniques we've created to make our lives better have also been turned against us. Now, both sides have to stay on the cutting edge of the technology to achieve their desired goals.

Designers of these algorithms need to make trade-offs between flagging false positives as spam or potentially missing true phishing attacks.

Acknowledgments

1] “What is phishing | Attack techniques & scam examples | Imperva,” Learning Center. [Online]. Available: https://www.imperva.com/learn/application-security/phishing-attack-scam/. [Accessed: 14-Apr-2019].
[2] N. Ghazim.jameel and L. E. George, “Detection of Phishing Emails using Feed Forward Neural Network,” International Journal of Computer Applications, vol. 77, no. 7, pp. 10 15, 2013.
[3] J. Vincent, “Gmail is now blocking 100 million extra spam messages every day with AI,” The Verge, 06-Feb-2019. [Online]. Available: https://www.theverge.com/2019/2/6/18213453/gmail-tensorflow-machine-learning-spam-100-million. [Accessed: 14-Apr-2019].
[4] E. Benishti, “Artificial Intelligence is Revolutionizing Phishing and It s Not All Good,” Ironscales, 21-Mar-2018. [Online]. Available: https://ironscales.com/blog/Artificial-Intelligence-Revolutionizing-Phishing/. [Accessed: 10-Apr-2019]. [
5] J. Seymour and P. Tully, “Automated E2E Spear Phishing on Twitter,” Black Hat. [Online]. Available: https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Soc ial-Engineering-Automated-E2E-Spear-Phishing-On-Twitter.pdf. [Accessed: 10-Apr-2019].
[6] KCCross, “Anti-spam protection FAQ,” Microsoft Docs. [Online]. Available: https://docs.microsoft.com/en-us/office365/SecurityCompliance/anti-spam-protection-faq. [Accessed: 14-Apr-2019].
[7] J. Huang, “Beyond Catching Sender Spoofing using AI to stop email fraud and Business Email Compromise,” Simply Security News, Views and Opinions from Trend Micro, Inc, 06-Dec-2018. [Online].

Available: https://blog.trendmicro.com/beyond-catching-sender-spoofing-using-ai-stop-email-fraud-business-email-compromise/.