To gather our dataset, we collected approximately 15,000 known phishing emails from multiple public sources (see sources below) Individual web-crawlers were written to gather these phishing emails from each of the websites and store them as HTML text documents for analysis. The web-parsed emails were then cleaned and formatted to remove errors, noise (e.g., images, HTML elements), and any extraneous formatting so that only the raw email text remained. Only text indentations and paragraph formatting were retained. We then developed a coding manual that allowed us to map principles of persuasion and framing to the email text.

For example, an email was to be labeled as applying gain if it lured the user into a perception that acting on the email would potentially lead to some gain or reward, for example, a discount online. Two lead coders trained a team of six coders over a period of three weeks using the coding manual. The coders' training comprehension was tested using 20 emails that were chosen by the trainers as representative of the principles of persuasion and framing. Coders were considered ready once labeling agreement was achieved on these representative emails. Because multiple principles of persuasion can be present within an individual email, Cohen's kappa was not computed to measure reliability. Instead, for each given email, two randomly assigned coders needed to reach an agreement on the labeling of a given email.

## References for Public Sources of Phishing Emails


[1] "Phishing scam reports archive," Miller Smiles. [Online]. Available: http://www.millersmiles.co.uk/archives/current. [Accessed: 28-Dec-2019].

[2] "Alerts & Notifications | Information Technology," University of Pittsburgh. [Online]. Available: https://www.technology.pitt.edu/news-and-alerts. [Accessed: 28-Dec-2019].

[3] "Phishing Scams Targeting the UMN," University of Minnesota. [Online]. Available: http://phishing.it.umn.edu/. [Accessed: 28-Dec-2019].

[4] "Phish Bowl/Phishing Scams," UCLA IT Services, 19-Apr-2016. [Online]. Available: https://www.it.ucla.edu/security/alerts/phishing-scams. [Accessed: 28-Dec-2019].

[5] "Phishing," Penn State | Office of Information Security, 10-Jul-2018. [Online]. Available: https://security.psu.edu/phishing/. [Accessed: 28-Dec-2019].

[6] "Recent Phishing Examples | Library & Technology Services," Lehigh University. [Online]. Available: https://lts.lehigh.edu/phishing/examples. [Accessed: 28-Dec-2019].

[7] "Phishing Alerts | UA Security," University of Arizona. [Online]. Available: https://security.arizona.edu/phishing_alerts. [Accessed: 28-Dec-2019].

[8] "Phishes & Scams," University of Michigan. [Online]. Available: https://safecomputing.umich.edu/phishing-alerts. [Accessed: 28-Dec-2019].

[9] R. B. Cialdini, "The Science of Persuasion," Sci. Am., vol. 284, no. 2, pp. 76–81, 2001.