Email Phishing Detector Report

We have the overall email score which will range from 1-10. 1 likely, 10 not very likely. We then run different tests on the email address and body which will create sub scores which are added together for the main score. We examine the email in the following categories, we will discuss each one by one:

- Whitelist/Blacklist
- Spelling Errors
- Phishing Language
- Suspicious sender account

Whitelist/Blacklist

This is mentioned first because it trumps everything. Known and trusted sources are in the whitelist, such as gmail.com, yahoo.com, hotmail.com, etc. If the user is sending from one of these sources, then the sender score will automatically be a ten regardless of the body of the email or their username. If the sender is on a blacklist their email will be immediately scored as a 1. If the sender is on neither, but their email gets scored a 1, they will be added to the blacklist for future use. This list will change with time as users get added to the blacklist and as we can add (or remove) reputable sources to the whitelist.

Spelling Errors

For spelling errors, we use the library spellchecker from Python. We compare the total amount of words in the body, compared to how many spelling errors it contains and find the percentage. A certain percentage (which could be tweak more) is tolerable, currently set to 15% (to account for actual human error), the higher the percentage the more this will contribute to the overall score. This could be tuned and tweaked with more data as to how common are spelling errors for humans.

Phishing Language

For these we created a list of common phishing terms. We examine the body of the email and similar to the spelling check we examine the percentage of terms that are in that list to the total number of words in the body. If it's over a certain threshold, currently 5% (can be adjusted), the percentage starts to contribute to the sub score and gets greater and greater as it gets larger. This can be improved as we add to the common phishing terms list and also with more data so we can narrow down thresholds.

Suspicious Sender

Determining a suspicious sender is more complicated. First, we separate the sender's email address into 2 parts, before the '@' and after. We do this because there are some principles that are applied differently. For example, we search for numbers in the email. However, if there are numbers contained in the second part it does not affect the score that heavily, because the business name could contain a number, but it will greatly affect the score having a number in the first part of the email as that is inexcusable.

We then start looking for symbols in the email, and count symbols like '.' and '-'. It is common for emails to have those symbols but when more than one each exists the score of the email will start to decrease.

We check for spelling errors; however we do not count for words separated by dots or hyphens. For example, if the start of the email is 'confirm' that's fine, I spelt wrong like 'confrm' that will trigger the score and greatly negatively affect that. But something like 'confrm-to-own' or 'confrmtoown' won't register as word for spell check. This can be greatly improved on and advanced.

This one is one of the most complicated. The numbers being generated are arbitrary and we tweak them based on our knowledge of social engineering. So when running tests if we know a certain type of email should be flagged we adjust the numbers so that will be calculated.