Big Data and Phishing Detection

What is phishing and why I choose to investigate it

Phishing is a type of online fraud where the one who is committing the fraud attempts to gain access to your personal information such as passwords and bank details (TechTarget 2016). Phishing is accomplished through fraudulent emails, instant messages and other method of online communication which are disguised as reputable companies (TechTarget 2016). I chose to research phishing detection and how big data is used to prevent this because since massive increase in the use of online communication phishing has become a real issue. Phishing is a big problem that can effect anyone, I myself have received phishing emails and that is why I think it is important for phishing to be detected to protect people who may be a victim to this crime.

How Big data is used in phishing detection

When a phishing email is sent a link is provided and on this webpage is where your data is request-ed. Phishing websites are discovered by analysing “end user confidential data submission statistics” (GlobalSign 2013). A central process (a google bot for example) is used as a receiver to receive data which indicates that personal and confidential information is submitted to a webpage, this data is taken from several computers (GlobalSign 2013).

Once the data has been gathered it is then analysed and aggregated and it is this stage that deter-mines whether there are anomalies in the behaviour of the submission of confidential data for ex-ample If there is an unexpected increase in confidential information submissions which indicated that this webpage is used for phishing (GlobalSign 2013). Once the big data analysis has determined whether the webpage is used for phishing many measures can be taken to prevent further phishing such as black listing the webpage, shutting it down completely or an alert can be sent to the appropriate party or automated system (GlobalSign 2013)

Another technique used to prevent phishing is the use of a large database of previous phishing sites that have been detected and using machine learning algorithms (Kitten, T. 2013). The machine learning algo-rithms analyse the data a use it to determine characteristics of the phishing sites so that when a new one arises it can compare the characteristics and inform businesses of the indicators (Kitten, T. 2013).

Has big data science successful in meeting business objectives

The object of a business when it comes to phishing is total prevention as these attacks can cost the company a lot of money. In fact a Cloudmark survey found that a successful spear phishing attack on average costs a company $1.6 million (Cloudmark 2016). Allow though big data science is helping to detect phishing websites the overall success has been very low as new phishing schemes are constantly being created in fact 85% of organisations have reported being victims to phishing attacks in 2015 which is a 13% increase from the previous year (Crowe, J 2016) .

My thoughts on how big data can be further used in phishing detection

In my opinion I believe awareness is key when fighting phishing attacks. Companies are a big target for phishers as they want access to their payments systems to gain the most amount of money so I believe it is key to inform the people in charge of payment systems through training. If you look at figure 1 you will see that the majority of organised don’t training to all of the staff or simple just handle the problems as they arise. Furthermore phishing attack most commonly occur through email so emails should be limited on payment systems. Big data can also be used to determine who is most commonly falling victim to phishing attacks as these are the people who should be made the most aware of what phishing is and how to avoid these attacks.

Conclusion

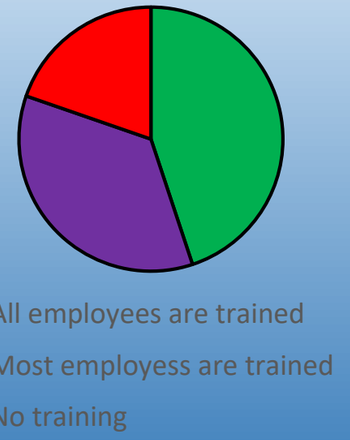
In conclusion it is clear to see that big data science is constantly being used to combat phishing attacks and although it has aided in the detection of phishing it has not done much in the way of preventing businesses and individuals falling victim to it. In my opinion I believe two factors are the reason we haven’t seen a greater improvement the fight against phishing. The two factors are a lack of awareness in individuals to know when they are under attack and the fact that phishing scams are developing with technology and constantly adapting. The techniques used to combat phishing can also be improved upon and expanded into other areas of security domains to further their progress in fighting off cybercrimes.

By Alex Stacey

Analysis of how these big data solutions can be transferred to different security domains

As we have already discussed phishing is a form of online fraud that mainly involves getting a victims bank details by sending fraudulent emails or through fraudulent websites. Given the fact phishing is a type of fraud the techniques could be transferred to other security domains, for example the use of machine learning in order to predict and detect online fraud sites. Furthermore, the use of a central process to receive data (detecting a large increase in the number of confidential data to a certain site) can be used for the detection and prevention of identity theft because the goal of these types of attacks are to gain personal data such as someone’s social security number.

Figure 1 : Do organizations train staff in phishing detection? (Malik, J 2016)



References

- Crowe, J. (2016) Phishing by the Numbers: Must-Know Phishing Statistics 2016 [online] available from < https://blog.barkly.com/phishing-statistics-2016> [21 November 2016]
- Cloudmark (2016) Survey Reveals Spear Phishing as a Top Security Concern to Enterprises [online] available from <https://blog.cloudmark.com/2016/01/13/survey-spear-phishing-a-top-security-concern-to-enterprises/> [23 November 2016]
- GlobalSign (2013) The Detection And Prevention Of Phishing Attacks [online] (1), 2 . available from <https://www.globalsign.com/en/resources/white-paper-phishing-attacks.pdf> [16 November 2016]
- Kitten, T. (2013) Using Big Data to Fight Phishing [online] available from <http://www.bankinfosecurity.com/interviews/malcovery-i-1909> [16 November 2016]
- TechTarget (2016) phishing [online] available from <http://searchsecurity.techtarget.com/definition/phishing> [16 November 2016]
- Malik, J. (2016) Clicking With The Enemy [online] available from <https://www.alienvault.com/blogs/security-essentials/clicking-with-the-enemy> [23 November 2016]