KF7004 – MComp Computing Research Project Reflection

16018262

Sept 22

Word count:

Contents

1 Background

1.1 Scope

This research attempted to look at whether there was enough data in website log files for a formula to be applied, in order to detect attacks. Part of the scope was to collect a representative sample of real-world data to measure the effectiveness of the proposed methodology. By using real-world data I could prove it is more accurate when validating security models; therefore, challenging conventional thinking that studies have to generate their own data.

The scope of this work was correct due to the novel nature of the work and the difficulty I had getting ethical approval, which will be explored in a later section. Even given these challenges, I was still able to gather enough data to test my hypothesis. The methodology was tested on a local website and could be applied to any across the world. The risks associated with this region were known and acted as a control. However due to countries' different internet patterns, it would be difficult to understand different traffic patterns.

1.2 Objectives

The objectives in my proposal were to; identify how attacks are evading current techniques, understand attack characteristics and refine a formula that can detect and determine risk. The high accuracy of the formula indicates I have understood current attacks and their evasion techniques proving a successful detection methodology. Although, I had to remove an objective, which would have been to develop a better way to assign risk to a country. I intended to look at the number of attacks coming from IP addresses within a country compared to the size of the population. But, I was unable to achieve this. I had issues with my support throughout most of the year, which forced time-constraints upon me; also, this data is difficult to determine. Another reason for its removal was that I was unable to contact the lecturer who suggested this idea, because they now work at another university. Therefore, I kept the original methodology used in the previous research, which still was able to prove that looking at risk for a country is a valid factor to consider.

1.3 Risk

In my proposal I identified the risk that website owners may be unwilling to give me their data for analysis. However, after engaging with them, they were happy for their data to be used. The research could potentially inform them of attacks on their website; if any happened to be found, I could advise them on how they could be mitigated. The biggest risk associated with this project was the lack of previous work. There was a question in my mind that as this methodology seemed relatively straight-forward, why had nobody tried this before? Was it because it was unsuccessful? This risk was mitigated due to the fact that, even if the conclusion did not support my hypothesis I would still be adding knowledge to this field. Therefore, this was a worthwhile risk and one that I would gladly take again.

2 Group vs Individual

Primarily I worked alone, although I did receive a lot of assistance from Nick (a Computer Science lecturer). Thurow et al. 1999 argues that multidisciplinary settings allow individuals from different disciplines to contribute their disciplinary perspectives in an attempt to solve complex problems. Myself and Nick have complimentary skill-sets that aided in a multidisciplinary approach. We focussed on different areas of the research, with Nick able to use his knowledge of algorithms and programming to help improve the formula. Whilst I focussed on which aspects of the data that pertained to cyber security. Hence I was able to identify risk factors and Nick would then advise on how to efficiently calculate the risk attached to the data. For example, I determined that user agents could be an identifier of risk, I had originally tried a Levenshtein distance in order to find the nearest user agent (this did not come to fruition). Upon reflection of this approach, it may have resulted in misidentification of user agents due to their similarities. However, Nick suggested doing a bag of words approach, which was ultimately a stronger and more accurate methodology. This stopped me overcomplicating the detection of user agents. In future projects I will now begin by looking at simpler approaches first. Furthermore, Nick had prior knowledge of the project, due to supervision at undergraduate level and therefore understood the project and its complexities.

Throughout the year myself and Nick critically evaluated our techniques; which meant we developed a habit of meeting regularly. This prevented delays in the research, therefore keeping momentum. Despite the high moment in the project overall, there were still weeks were progress was slow. These made me feel like I hadn't done anything, which negatively impacted my motivation. The project was not initially planned out, which naturally generated a more agile approach. Furthermore, due to the novel nature of the work there was a danger that I would set unrealistic deadlines in order to complete the research; which is deemed by **GanttPRO** to be one of ten major pitfalls of projects success(**GanttPRO**). This approach yielded results

for me because I had to deal with additional complication. The level of support I require was not met for a elongated period of time, which made it difficult to plan what activities could be undertaken in advance. An example of this was, when I received a phone call at 6pm one night, informing me I could come into University the next day. In addition to this, I also had times were I thought I would be in and would find out that I could not attend. This meant some of our meetings had to take place on Teams and sometimes involved brainstorming on whiteboards. This was sometimes difficult to read, if internet connection was poor. On reflection the whiteboard feature on Teams may have been better. The circumstances dictated this work required a different approach to manage the project. This is inline with AJ Shenhar's theory on project management (Shenhar 1998), who suggests that different projects require different styles of management. So due to external factors an agile project management style was needed.

This occurred organically in the present research as ideas could be effectively implemented without disagreement, regarding the direction of the project.

I largely worked alone on this project as a result of the work being heavily linked with that of my undergraduate dissertation. Most of the underlying software was already written, therefore I worked to identify small improvements that could be made. Then developing a methodology to prove the accuracy of the program. However, new ideas may have led to a better outcome, this would be because of the ability to collaborate with other researchers. However, I did get a high degree of collaboration with Nick; this is in line with Robert E. Levasseur (2010) who states "to initiate and... the high level of two-way communication" (Levasseur 2010). The initiation process occurred very naturally between myself and Nick. While on placement, we discussed a problem caused by an IP address attacking my website and I proposed that I programme something to detect the attack. This sparked a reciprocal flow of ideas that continued throughout the research, even if some elements were absent from the final project. It can be shown that the more collaboration that exists in research projects, the higher the quality of the research (Figg et al. 2006). Therefore the collaboration within this project led to a better output but other collaborators might have increased this further. For example, when classifying the accuracy of the program; having someone to classify the data as well, helping to remove any bias that I might have had.

3 Background, Motivation, Sources of Knowledge

Throughout this project, there have been many stages where reflection has been needed in order to understand the progress that has been made. By regularly putting my research on hold and taking a step back, I was able to develop a regular reflective habit (Dyment and O'connell 2010). For example, after making changes to the formula, I would run the data again and reflect on the output. However, at other times throughout the year, there have been ongoing issues with a lack of available support in order to complete work. It added more stress and I often felt like I was rushing work that I should have taken more time over. There have been periods where reflection in action have been required, often when doing this we are unaware that we have derived meaning from an experience. Through this process we find ourselves doing the same action again, as it has a positive outcome (Schon 1983). When I reflected on where I was in the project, and what I needed to do to progress it resulted in more effective decision making and problem solving. This was mainly evident by last minutes changes made to my proposal. This meant having to do an online zoom session, so that someone was able to assist me in completing my final draft. The positive outcome here was that I got the work done and I was able to learn that doing work on zoom, although not ideal, still contributed to the success of the project. One thing I'd change is, not asking other people to proofread my work near the deadline at risk of being too last minute to get any meaningful changes in.

The motivation behind this research was to solve a challenge that I faced, running my own website. The lack

of literature in this area to do with cyber security was surprising, however this led to freedom in creating a new methodology. I had to combine some existing techniques and test some new ones. Due to the lack of relevant literature, the literature review was challenging to construct. Therefore it became more about identifying gaps in previous work. Hertz (1997) points out that 'the reflective researcher does not merely report the findings of the research but at the same time questions and explains how those findings are constructed'. (CITE) This shows that I was able to reflect on the available literature and ask questions that led to my hypothesis. A lot of these questions were not complicated in nature, but came from a working knowledge of websites and servers, the key lesson that I learned in this was that just by asking simple questions about the research, it was easy to identify flaws. Moving forward I will read research with a more critical point of view.

4 Relevance

This work was relevant due to the fact that all previous cyber security work in regards to low rate DDoS attacks could be seen as deeply flawed. For example, Tripathi generated their own data to validate their own hypothesis and methods. One of my goals was to show that cyber security work needs to be validated by real data. The work is also relevant due to an increasing reliance on websites and the internet in general. If there aren't effective ways to check attacks, then they have the potential to disrupt daily life. I saw this actualise itself when, after helping a Newcastle University Researcher I was able to reduce the running cost of their web severs by fifty percent. This shows the real world effect that the software has, along with the harm that attacks can cause if they are able to go undetected. The work was also relevant to me on a personal level; as it allowed me to create some research that has never been done before and go on to develop it as a product to sell to other websites.

5 Ethics, Legal, Social, Security and Professional Issues

5.1 Ethics

there was a delay in receiving ethical approval due to the ambiguity of user consent in the collection of website log data, and whether users had given informed consent about the collection. Websites say that data can be analysed however due to the university policy of expressed consent, there were questions about whether the data could be used. However, due to the study wanting real-world data, this might have changed user behaviour, knowing that they would've been tracked. After my ethics form asked for clarification the second time, I met with Jamie Mahoney who advised that we should look at GDPR. In particular article 14 paragraph 5, which states if it is too much effort to get a form of consent then you do not need it (look at the actual statement). Upon reflection I should have contacted Jamie earlier. This is also a legal issue and it is the same as the ethical problem.

Another ethical implication of the work may be the fact that entire country is given a risk. This is mainly done to give the software an idea of context, however a user from a high risk country can still get a low overall risk score.

6 Legal Issues

There was a potential legal issue as people could not opt out of the data analysis. This is due to the fact that as soon as they went on to the website their IP address was logged. Most websites privacy policy state that IP addresses will be logged and used for analysis, therefore most users should be aware of how their data will be used.

7 bibliography

References

Dyment, Janet E and Timothy S O'connell (2010). "The quality of reflection in student journals: A review of limiting and enabling factors". In: *Innovative Higher Education* 35, pp. 233–244.

Figg, William D et al. (2006). "Scientific collaboration results in higher citation rates of published articles". In: *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 26.6, pp. 759–767.

Levasseur, Robert E (2010). "People skills: Ensuring project success—A change management perspective". In: *Interfaces* 40.2, pp. 159–162.

Schon, D (1983). *The reflective practitioner. How professionals think in action*. London: Temple Smith.

Shenhar, A. J. (1998). "From theory to practice: toward a typology of project-management styles". In: *IEEE Transactions on Engineering Management* 45.1, pp. 33–48. ISSN: 0018-9391. DOI: 10.1109/17.658659.

Thurow, Amy Purvis et al. (1999). "The dynamics of multidisciplinary research teams in academia". In: *The review of higher education* 22.4, pp. 425–440.