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 Project Management

Primarily I worked alone, although I did receive some guidance from Nick (a Computer Science lecturer). Thurow et al. 1999 argues multidisciplinary settings allow individuals from different disciplines to contribute their perspectives attempting to solve complex problems. Myself and Nick have complimentary skill-sets that enabled a multidisciplinary approach; focussing on different areas of the research. Nick used his knowledge of programming algorithms 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 any risk attached to the data. For example, I determined user agents could be an identifier of risk, originally trying a Levenshtein distance to find the nearest user agent. This proved unsuccessful. This approach, 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 I will begin by looking at simpler approaches.

Looking back on this project, working with Nick was enjoyable, due to our different approaches. Nick always had a positive outlook on the project and its challenges; therefore keeping my motivation high and I came away from our meetings believing the project could be completed with ease. He also helped by, not only generating ideas for the project but helping to implement them. This further helped reduce any stress on me and he enabled me to have faith in our methodology. However, a slight problem was, Nick was difficult to reach at times. The primary way to ask a question to him was looking for a space in his calendar that was mutually convenient. I often felt bad for booking in meetings for issues that could have been solved via email but, this was unavoidable. Which added to my stress, particularly at key times; for example, when submitting my ethics for the third time. I was going to wait for Nick to double-check, but I was getting more stressed about the ethics not coming back on time; rather than the potential of it being rejected for the fourth time for problems Nick may have spotted. Overall, working with Nick proved beneficial and if I were to do a similar project I would ensure we had better communication outside of meetings if possible.

Throughout the year our techniques were critically evaluated. We developed a habit of meeting, therefore maintaining momentum in the research overall. Despite this, there were still weeks where progress was slow. These made me feel like I hadn't done anything, negatively impacting 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 because I had to deal with additional complication. The level of support I require was not met for an elongated period of time, which made it difficult to plan what activities could be undertaken in advance. Once, I received a phone call at 6pm, informing me I could come into University the next day. This is an example showing I had varying amounts of time to work on the project each week. 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; also, I should have tried to get more support online, if not in person. 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.

The work was 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 from different researchers may have led to a better outcome, because of the opportunity for collaboration. 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. Myself and Nick would often challenge one another's ideas in order to ensure our methodology was robust. It may have been useful to include another cyber security researcher, in order to further challenge our thinking. Additionally, 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. 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 program 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.

If I were to do this again with the knowledge of my lack of support and doing a new project I would have worked in a larger group because the level of stress I put on myself would have been reduced. However, there may have been deadlines that I might have been unable to meet; potentially causing a greater amount of stress

I also found that when working in novel areas some of the concepts maybe difficult to explain, therefore continuity of the researchers no matter the size is important. If new people had joined at the start of this year it may have been difficult to explain the principles behind the methodology.

I had already worked with Nick previously in undergrad, so my confidence communicating was already high. Throughout the year I gained more confidence communicating my ideas to various other people. For example, with my different support workers being able to communicate where I was up to in the project quickly and explain what they could do to help me. Furthermore, I had to communicate concisely with Jamie in order to get my ethics.

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.

I felt highly motivated to do this work as it solves a real world issue that I had. Also, some people told me that doing this analysis as a mathematical model was impossible. It worked at undergrad level so I was motivated tp actually prove that, but I was unable to previously.

At the start of year motivation was high, quickly fell away when i found out that I had no support.

The fight with the uni to cone back next year increased my motivation, because I wanted to show hem what I could do.

After xmas break my motivation increased further as I actually got support more than one day a week.

Ethics 2nd rejection was a low point. "why bother if I can't get ethical approval?" and then Jamie ensured I got it.

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.

I'm also pleased that I have shown that my work is relevant in the cyber-security field and provides valid conclusions. It may help other researchers think about the way that cyber-security research is conducted.

5 Ethics, Legal, Social, Security and Professional Issues

5.1 Ethics and Legal Issues

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 or completed further research alone of GDPR. This is also a legal issue and it is the same as the ethical problem.

Ideally I would have liked to have had the ethics in sooner in the year, but it could only come after the proposal was marked. I couldn't rally do anything with that. Should have possibly asked someone on the module to peer review my ethics form. Could I have generated my own data?

COuld have used a website's old data, providing it was over 6 months old; after this point it falls out of the purview of data protection

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.

Should have anonymised the IP addresses as much as possible by using a hash function, however that may have made any analysis more difficult. I had known how complex the ethixs wpld be for this project and given the time-constraints. I maybe should have done a simpilar project.

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.

6 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.