Daily Thoughts & Experience Diary

*Thursday 16th January 2020*

* Received an email from Mr Bustard confirming my first choice project.
* Just elated.
* But nervous because I’m unclear on the full extent of the project. I’ve never used a raspberry pi before but I have heard of them.
* The project specification states:

“Create **a raspberry pi, npcap and wireshark** based **packet sniffing network monitor** that can be attached to **wired and wireless networks** and **passively monitor network communications**. Time permitting explore how anomalous transmissions can be detected using machine learning based modelling of normal network behaviour. This project also requires an analysis of the data protection and privacy implications of the technology.”

* Out of the three technologies stated at the beginning of the spec, I’ve only used wireshark before, and even then it was only briefly for my networking module last term.
* I doubt I will be able to get onto the additional functionality regarding machine learning.
* Suddenly the Web Technologies module looks a lot more inviting…

*Friday 17th January 2020*

* Met with John today in person, felt reassured as he offered me all of the equipment I needed and told me how to get started.
* These steps included dual booting my laptop and installing a Raspberry Pi emulator (Qemu)

**Rough Story**

I never thought that I would have to dual boot my laptop for this project. It seems I’ve met my match on the very first hurdle. Dual-booting a computer can be quite risky if not done correctly, as it could lead to a deletion of the Windows OS and all files on the hard drive. Once I have carried out this I will have to install the emulator, which I hope is just a simple application with its own installer. Once this is done, I can then begin looking into programming the network monitor aspect of the project. John recommended C++ as that is the most common language for programming using Raspberry Pies. I’ve never used C++ before, only Python and Java, however I don’t think that this will be too tricky to handle, as C++ has the same structures as most programming languages and so I will only have to vary the syntax. The actual packet sniffing code will be very difficult and will most likely constitute the most work in this project as I have never programmed anything that accessed packet data across a network. I don’t think I will get around to the machine learning aspect of the project.

John’s email asks us to consider what part of the project will be of most use to others – to be honest I can’t work out at this stage what that might be. Will mull this over and hopefully inspiration will hit.

**Task List for Weekend Work:**

* Start Diary
* Dual Boot Laptop
* Install Qemu

*Saturday 18th January 2020*

* Worried about dual booting my laptop due to the risks involved (previously mentioned). As a complete beginner at messing around with OSes I was unsure if it was wise for me to proceed.
* I backed up all of my uni work/project files to my OneDrive account, in case things went wrong.
* I looked up a tutorial on how to install Ubuntu and perform a dual boot.
* I followed the tutorial step by step cautiously, scared to deviate in case I made a mistake.
* Everything was going well until I made the final configurations to the BIOS of my laptop and went to restart. What I expected was the GNU GRUB bootloader to appear, what I got was a windows blue screen asking for a Bitlocker key.
* I panicked, thinking that my Windows operating system had been deleted. Luckily, I had saved my files to my OneDrive and I had also saved a copy of the Bitlocker key. Creating a copy of the Bitlocker key was **not mentioned** in the tutorial video. I also had another laptop with me so I was able to retrieve the key. I scrolled down to the comments of the video and there were a lot of comments complaining about having to ring Microsoft support for their Bitlocker key. I was one of the lucky ones.
* After a bit of troubleshooting I was able to successfully load Ubuntu!
* I felt very pleased with myself, after conquering the fear of losing all of my files.
* When I switched back to Windows however I had to reinput my Bitlocker key again. This is troublesome, so I made a mental note to attempt to fix this issue.
* I hoped that installing QEMU would be a simple download and run deal, however this turned out not to be the case.
* To install QEMU you must first download three things; the application package, the raspberry pi signature and the kernel for the raspberry pi. Then to run QEMU you must carry out complex console command.
* Overall installing QEMU was much more complicated than I expected however I persevered. This was my first time using the command line to install an application and it was interesting to learn about how all of the commands worked, e.g “sudo” standing for “superuser”.
* I felt really impressed with myself having completed the technical tasks set for this weekend.
* On running QEMU I noticed that the emulator itself was rather slow. I made a note to fix this.

**Side Quests Gained**

* Fix Bitlocker key re-entry on swapping to Windows.
* Attempt to make Raspberry Pi emulator to run faster.

**Task List for Week ahead.**

1. Collect physical equipment from John Monday/Tuesday
2. Ask John if we need to create a website for our tutorials regarding the project, or simply go through a pre-existing website (Week 4).
3. Explore npcap on internet as I know nothing about it.
4. Setup Github page.
5. Complete Side Quests.

*Sunday 19th January 2020*

* Fixed the issue regarding Windows requiring the Bitlocker key every time upon switching.
* It appeared I had not disabled the bitlocking at drive level.
* Had some work from other modules, so did that rather than continue with the project, as most tasks required to be at the university
* Felt pleased that the dual booting was a success and there were no more issues regarding it.

*Monday 20th January 2020*

* Decided to figure out what Npcap was.
* Npcap is the packet sniffing component of a larger piece of software called Nmap.
* Nmap is a utility software which is free and open source. Its main purpose is to scan large networks and retrieve characteristics of each computer (e.g the OS type,types of firewalls in use etc).
* Npcap seems to be a superior version of Windows own packet sniffing library called WinPCap.
* WinPCap is depreciated.
* I downloaded and installed Npcap and the Npcap SDK.
* The Npcap SDK contained a variety of .html files. These html files constituted the documentation of Npcap.
* I had never seen documentation displayed like this before (typically documentation is hosted on the library’s website)
* This was incredibly frustrating to work with as most of the links in the documentation were broken.
* I did eventually find the .lib files for Npcap.
* I downloaded Visual Studio in order to prepare to code in C++.
* Unfortunately, I could not figure out how to add these .lib files to my project.
* I spent hours shifting through documentation and old forum posts, however none tackled the issue head on, or were using WinPCap or something else.
* Alas to no avail
* I had spent a lot of time on this and had made no real progress on installing the library.
* This made me a bit upset but I promised myself I would stop if I ever got too frustrated.
* I installed Wireshark in preparation for the next step.
* Wireshark uses NPCap to capture packets.
* Today was not a great day on the project front.

*Tuesday 21th January 2020*

* Continued searching for a way to import the NPCap library into Visual Studio, but yet again I could not find an answer.
* This was vexing as I have now spent a day and a half on something which should have been relatively simple.
* Instead of ramming my head into a brick wall, I decided to try to make the raspberry pi emulator run faster.
* I downloaded the “Virtual Machine Manager” which offers a more streamlined approach for instantiating Virtual Machines and also allows for the use of KVM (Kernel Based Virtualization).
* Unfortunately, Raspberry pi emulation is not supported as the Raspberry Pi uses an ARM CPU core, and the slowness is caused by the conversion of x86 processor instructions to ARM CPU instructions.
* Though this search did end up being fruitless, it did teach me a lot about different types of emulation and virtualisation.
* I launched the emulator when no other programs were being run. This improved the performance of the emulator. I suppose I can live with this limitation.
* I’m glad that I got QEMU to run a bit faster as I needed a win.
* I also obtained the hardware required from John today, so I will assemble it tomorrow.

*Wednesday 22th January 2020*

* Downloaded the OS for the raspberry pi (Raspbian).
* I am amazed that I have had to download 3 Operating Systems in order to set up this Networking project.
* I flashed the Micro SD card with the software used to flash the pen drive when I was installing Ubuntu.
* I set up the raspberry pi and loaded the OS.
* I went to connect the mouse and keyboard to the raspberry pi however they did not connect.
* I checked the peripherals on my computer, and they worked fine, so it must have been the USB ports on the raspberry pi which were causing the issue. I gave the ports a clean with a damp cloth, but still no luck.
* I went back to John and he gave me a different raspberry pi. This one worked perfectly, and I was able to set it up with no problems and even connect it to my Wi-Fi Network.
* This is a big step forward.
* Reviewed my Main Quests for this week and have yet to complete the GitHub/tutorial setup and have yet to fully install the NPCap library.
* John has contacted a student in an above year to help with the NPCap library and has confirmed that it is quite difficult to get it up and running. This made me feel content with myself and the progress I have made.
* I have been collating the websites and resources I have used up to this point in preparation for the tutorial/Github part of the project.

*Thursday 23rd January 2020*

* What a joyous day!
* Nathan Nash (the student John contacted) as rescued me from my woes.
* Nathan and I sat and went through the settings for the Virtual Studio Project.
* Sat there for about an hour and changed a variety of settings.
* Turns out I as not too far off the correct settings, and it was simply a matter of incorrect file paths.
* I’m over the moon.
* …yet
* I’m annoyed that it took two and a half days to obtain the solution.
* Then it hits me.
* I get to write a tutorial.
* Which means I can save poor students from having to endure the hardship I have had to.
* Excellent, I shall make it my main priority for the weekend to start my tutorial on github.
* Talking to Nate made me appreciate how much I have already learnt regarding .h and .lib files aswell as Visual Studio.
* I guess if I hadn’t endured the frustration, I wouldn’t have fully grasped the solution.

*Friday 24th January 2020*

* Had my first meeting with John today.
* He gave me a lot to think about regarding the project.
* He said that success in his eyes was to make something useful for others, and not just completing the final project.
* He also said to do the tutorial on something which has poor tutorials and/or something that you found hard.
* Something which fits this bill is the NPCap library (or “architecture” as the developers call it, as it interacts with the network adapter. I don’t see the difference.)
* NPCap official docs are woeful at explaining things, for example; The second example used in the beginners’ guide breaks due to the file referencing a user defined class which is not given.
* I could make a guide for beginners looking to get started with the library which I did mention in yesterday’s entry.
* However, a beginner teaching beginners might not be the best course of action – not enough experience…
* A tutorial that would be a lot more straightforward would be the dual booting tutorial. I could get the message out about backing up the Windows Bitlocker key.
* Though even with the Bitlocker setback I still found the dual boot, for the most part, straightforward and the tutorials were of good quality.
* The same goes for installing QEMU.
* Doing a brief google search for NPCAP tutorials yields (mostly) nothing but tutorials relating to Wireshark or WinPCap, so there is a desperate need for some tutorials.
* I think that settles it.
* I will do my tutorial on NPCap as it will make the most impact.
* Now to pick the HTML layout…joy

*Saturday 25th January 2020*

* Today is the day I make a start on the tutorial web page.
* The first step is to look at what NPCap tutorials have been created.
* Of course, there is the NPCap docs itself.
* I’m **very** intimate with it already.
* The official documentation for NPCap states it assumes “a good knowledge of C, networks and network protocols”.
* I can tell you it assumes much more.
* For example, the docs reference to specific, complicated validation algorithms for packets (the CRC algorithm explicitly).
* Another example of the docs assuming too much is that the docs do not provide a “Getting Started” guide regarding the setting up of the library/architecture with Visual Studio or any IDE for that matter.
* Overall it is quite good at explaining how the library/architecture works which is the main objective of documentation at the end of the day.
* Another tutorial I found is from a site called Poftut. This tutorial just tells you how to install the NPCap service and not how to use the SDK. Further, the tutorial does a poor job explaining what the NPCap service actually is.
* I suppose some very basic beginners may need help double clicking the installer wizard but other than that this tutorial is of little use.
* The rest of the tutorials on the internet are either about:
  + The depreciated WinPCap library
  + Debating whether to use NPCap or WinPCap.
  + Using Wireshark.
* So, I don’t think I’m entering a saturated market.
* In terms of my target audience for my tutorial web page, I believe that targeting a student is the best course of action.
* Why?
* Because a student will (hopefully) know the basics of procedural and OOP programming paradigms and so can follow the logic of any program presented to them.
* They **may not have** an understanding of C/C++ (like me) but since they will know the constructs of a programming language, (which don’t change between language to language) they will only have to get use to the syntax.
* I’m a bit concerned though about this tactic however as C requires manual memory management. Hopefully this aspect of C is does not come up.
* The code examples in the SDK for NPCap perform the desired task, however the code itself is sparse on comments. I will add more comments to the code so that it is more readable.
* I gave a thought to other types of audience, for example; experts in the field and complete beginners to programming.
* Experts in the field of networking would be hard to teach, as by definition, they understand the intricacies of networking and how sniffers work etc. I wish they could teach me! I am but a lowly student armed with “Schuam’s Outlines Computer Networking” by Ed Tittel. Experts could happily read the official documentation for NPCap, further limiting purpose to pursue this audience.
* Complete beginners should not be attempting this project. There are plenty of tutorials regarding the basics of programming so they can read/watch those.
* I went for the “Black White” HTML template from html5webtemplates.co.uk. I don’t claim to be a web designer, but this will nicer to look at then the official documentation, which is default HTML, with no styling.
* I’m using Komodo Edit to edit the HTML. This editor has a very clean interface and does not require a great deal of effort to learn. The editor also includes an inbuilt browser tab, allowing the user to view their HTML webpage while remaining in the editor.

**Creation Process for How-To Webpage**

Firstly, I download an HTML editor, I’m using Komodo Edit to edit the HTML. This editor has a very clean interface and does not require a great deal of effort to learn. The editor also includes an inbuilt browser tab, allowing the user to view their HTML webpage while remaining in the editor. Here is the download link

https://www.activestate.com/products/komodo-ide/downloads/edit/

Secondly, I decide on who my target audience is for the how-to guide.

Then I break down what I have done with NPCap down into steps, starting with installing the required software.

Each step is will constitute a page on my website.

I write what I want to say in each step in a word document, as it is difficult to edit sentences in the HTML editor. I even place the images I want to use in the word document itself. I call this my how-to guide “script”.

Once the “script” is finished I then place the word documents contents into a word to HTML converter:

<https://word2cleanhtml.com/>

This produces the HTML content for my pages and provides the precise height and width for images.

Then I downloaded a website template from the internet. I used this site: <https://www.html5webtemplates.co.uk/templates.html>

Next, I create a new Project in Komodo Edit and then, using the downloaded template, begin creating my html pages. I then copy and paste my HTML content into these pages.

In Komodo Edit I then create an images folder in which I store all my images for the website. It is important to have many screen caps of what you are doing in the tutorial as it helps the student follow the guide.

Finally, I link up all of my HTML pages together, so that it is easy to navigate between them. Typically, the template will have a navbar to facilitate this.

*Sunday 26th January 2020*

* Feel exhausted after spending around 5 hours writing the tutorial in a word document, complete with images etc.
* It was very tedious but I’m glad to say I’ve made huge progress in it.
* Translating it into HTML should be easy enough to do.
* A little upset that I didn’t progress with the coding aspect of the project, however at the end of the day, the tutorial is slightly more important as it has a deadline.
* Writing the tutorial was a good test of memory and I feel proud that I could write it with minimal googling.
* As the week draws to a close it is time to assess the tasks for the week.

**Task List for Week of 19th – 26th ahead.**

1. ~~Collect physical equipment from John Monday/Tuesday~~
2. ~~Ask John if we need to create a website for our tutorials regarding the project, or simply go through a pre-existing website (Week 4).~~
3. ~~Explore npcap on internet as I know nothing about it.~~
4. Setup Github page.
5. ~~Complete Side Quests.~~

I’m a bit miffed that I did not complete all of this weeks tasks, however their seems to be little agency for setting up a Github page, as no suitable code for the page has been written yet. So really I don’t think I should be too hard on myself.

Now that mostly everything is set up I can now concentrate on the tutorial and the project itself.

**Task List for Week ahead.**

1. Complete Tutorial. (including HTML etc)
2. Capture my first packets using NPCap.

Two tasks for the week ahead, but they are quite big ones. Firstly, completing the tutorial will be more of a time sink then anything. I might investigate an automated process for that…

**Side Quests Gained**

* Figure out a way to convert a Word Document into HTML (within a reasonable time frame)

Secondly, I’m still far off from sending capturing packets using NPCap. It will be a challenge, but I guess that’s why the modules called Computer Science Challenges.

*Monday 27th January 2020*

* Carried on with the website tutorial for the majority of the day.
* Had a great time commenting the NPCap Code as I learnt a lot about C in the process.
* Have a slight issue with my laptops Wi-Fi chip not being supported by NPCap.
* John has assured me that if I run my code in linux and use the USB Wifi Dongle, it should work.
* We shall see how that prediction turns out tomorrow.
* I’ve gotten to the point in the web tutorial where I **need** a screen cap of packets being captured.
* Hopefully tomorrow I can obtain them.

*Tuesday 28th January 2020*

* Frustration
* I’ve hit a wall.
* Firstly, I needed to install an IDE for the C code.
* Visual Studio itself cannot be run on Linux.
* I downloaded Visual Studio Code instead.
* However, Visual Studio Code is just a glorified text editor with no compiler/debugger.
* I followed a tutorial on how to get a complier for Visual Studio Code however this fell through as It only worked for simple scripts (I could not add the header/library files.
* Annoyed, I set up the raspberry pi to see if I could run the code using it.
* I used the Genie IDE however again I couldn’t figure out how to install the header/library files. Also, how am I meant to download the Windows only NPCap Service on raspberry pi?
* Compounding this is that NPCap takes into account the processor, so I’m not sure how it will work with an ARM processor. We shall hopefully see.
* I am irritated with myself….
* It feels like all the pieces are at my disposal, but assembling them is too difficult or too obscured.
* I decided to try the Wi-Fi dongle on my laptop using the Windows OS.
* Thankfully it worked and I was able to capture the much-needed packets for my web tutorial.
* A minor relief
* However I can’t shake the overarching problem I’m facing in terms of NPCaps flexibility over the three Operating Systems I’m using.
* Of course, there are alternatives I could use. Although that seems like a cop out more than anything else and I have already dedicated a lot of time to understanding NPCap.
* I hope I don’t fall victim to the Sunk Cost Fallacy.
* I will continue the web tutorial for the rest of the day.

*Wednesday 29th January*

* Continued working on the tutorial today.
* I used an online tool called word2cleanhtml.com to convert the word file to HTML.
* The tool advertises clean HTML, but I was correcting some formatting mistakes, such as putting an image of some settings after the text that explains it.
* Anyway, I had a look online and found the submission form for the webpage.
* It asks for a static HTML page.
* Sigh,
* I have broken the guide into multiple pages, so it is easier to read, and have also linked them together using hyperlinks.
* Oh well, I guess I will just have to reformat the pages into one page.
* That shouldn’t take too long.

*Thursday 30th January 2020*

* It didn’t take too long.
* I can happily ay the first draft of my web page is finished.
* Just need to go to John for some feedback.
* Still vexed on the raspberry pi situation.
* I’ve done some research into the issue and it does not seem at all possible.
* I am feeling worried as this issue has put the whole project in a state of stasis and I feel as though I’m wasting time.
* On a brighter note, I believe that my web tutorial has gone very well.
* I could have added a little more detail into the packet sniffing section however overall…
* This guide will defiantly help people use NPCap.
* I do feel proud about that.
* **It will have made a lasting impact and will be valuable for people**.

*Friday 30th January 2020*

* Had my meeting with John today.
* He mentioned that he thinks that the diary needs improved and was worried that myself and Ben (partner) were too far ahead of the pack essentially.
* He and I cannot help it if others are still clambering around wondering what to do.
* Anyway another topic brought up was about the diary and that the deliverable (webpage) was not in itself worth anything academically and it was the diary that was of most importance.
* I disagree with his philosophy.

As previously mentioned, time and again the point of the module is not to achieve the goal but to produce something that will be valuable to other people.

It then follows that the webpage produced should be worth at least something as that is the actual thing that is produced.

Again I understand that the diary is the most important thing.

But it is simply not the thing that will be invaluable to other people.

* After that tangent the diary itself needs to be improved.
* I need to produce essentially a learning section and a planning “daydream?” section.
* In order to stop this document from being too crowded and I will separate these sections into two other word docs.
* I shall spend the weekend working on this.

*Saturday 30st January 2020*

* Helped at the Offer Holders event today.
* This was a great experience as I got to talk to a lot of aspiring students about Queens.
* I felt very tired after the event.
* I took the evening to myself, however I did send off my research into the NPCap compatibility to John.
* Hopefully he replies soon.

*Sunday 1st February 2020*

* Worked on retroactively creating these two new diary documents.
* Tiresome as it is hard to express your theoretical plans.
* Also looked at alternatives to NPCap.
* TShark and TCPDump seem to be the most favoured alternatives.
* It seems that both have their own benefits and drawbacks.
* For live capturing of packets, tcpdump seems to be favoured, due to it being “lighter” than TShark.
* However, for analysing large packet files, TShark is better, due to its wider variety of filters and settings.
* This has got me thinking.
* I could use both TCPDump and TShark together in my project.
* I will have to see what John says tomorrow.

*Monday 2rd February*

* John has allowed me to use TCPDump and TShark.
* This is great news as I can now resume work on the project.
* I can also now review my tasks for the previous week and make new ones.

**Task List for Week 26th – 3rd.**

1. ~~Complete Tutorial. (including HTML etc)~~
2. ~~Capture my first packets using NPCap.~~

**Side Quests**

* ~~Figure out a way to convert a Word Document into HTML (within a reasonable time frame)~~

In all honesty I forgot about the tasks set for the last week as I was so stressed with making the How-to guide and the project brick wall, however now that these issues are resolved I feel much happier. I also elated that I completed all the tasks.

**Task List for Week ahead**

* Capture packets using raspberry pi.
* Using my laptop, obtain and display these packets.

This week I want to move the project forward. I want to make use of all the parts. I want to assemble the machine.

* Installed a lot of things today.
* I Installed tcpdump on the raspberry pi
* I also installed TShark on the raspberry pi.
* In order to test configurations for TShark I also installed it on my laptop as it is much faster and easier to set up than on the pi.
* A screen shot of a computer

  Description automatically generatedCaptured my first packets:
* Spent some time just thinking about the project in my head.
* I’m going to have to extract the packet file from my raspberry pi somehow…
* Will think about that tomorrow.

*Tuesday 3th February*

* I’m going to use WinSCP in order to wirelessly exchange files to and from the raspberry pi.
* This will remove the annoyance of having to take the microSD out and then putting it back in again etc.
* Spent a bit of time looking into Wireshark and I have formulated a logic that I could to pursue.
  + Raspberry Pi captures packets (on loop)
  + Raspberry Pi sends packet files to laptop
  + Laptop analyses packets.
  + If a foreign device attempts to connect, remove the device off the network wirelessly.
* Granted this idea does not use machine learning, however this could be the first step towards getting to that point.
* I do feel unsure about all this as I’m not sure if everything listed is possible.
* I have set up WinSCP and so tomorrow I will have a look at trying to automate the sending of files.

*Wednesday 4th February*

* Struggling to get WinSCP to work.
* Apparently, for what I want to do with the software, I need an SFTP server.
* After losing my mind halfway through a tutorial to host my own on my own local machine, I decided to look at AWS.
* Taking a step back, I realise that I may not need this type of file transport. I could send the files via email.
* However, this process could be too taxing on the raspberry pi.
* I’ll see what John has to say about this.
* There could be a better solution then the ones I am proposing.

*Thursday 5th February*

* Instead of looking at the WinSCP file transfer stuff I opted to instead look at some forms of network attacks.
* This decision made for the sake of my sanity of course.
* I dusted off my networking textbook and had a look at the section titled “Security”.
* I read this chapter.
* It focused primarily on encryption, which did not help me.
* However, it did mention briefly about Denial of Service (DoS) attacks.
* Essentially a computer floods the network with lots of packets, making the network too slow for users to use.
* This would be one of the threats my device could hopefully detect.
* There are many different versions of this attack.
* I chose to emulate a SYN DoS attack.
* This attack involves sending lots of SYN packets.
* SYN packets are used as part of the three-way hand shake. This handshake being the basis of all connection orientated communication between devices.
* I created a script using Scapy which would flood my personal network with SYN packets. Scapy is a packet creation library which works with Python.
* This did not turn out as planned.
* The packets were sent but my router seemed unhindered. I was still able to use the network with no slowdown.
* This troubles me
* I sent an email to a networking lecturer with the attached .pcap (packet capture) file to see if he could help explain what had happened.
* I have set up an appointment with him tomorrow…

*Friday 6th February*

* Uploaded the website and diary submission that was due for today.
* Hopefully everything is in order.
* Also had my meeting with John today.
* He recommended that I take step back from the technical aspect of the project and have a look at products on in the market that are in the same area as my project.
* I confessed my issues with WinSCP and automatic file transport from my raspberry pi.
* John recommended Flask, a lightweight python-based web server I could run on my raspberry pi.
* I think I can set that up over the weekend.

**Side Quests Gained**

* Send a pcap file from the raspberry pi to a laptop using Flask.
* In order to explore the anomaly detection aspect of the project I have also been tasked to watch some Machine Learning lectures on fast.ai. I’m looking forward to this challenge, however this may have to wait till next week.
* In order to have a bright start tomorrow, I am writing my notes for my other two modules tonight. This allows for maximum project time.

*Saturday 8th February*

* Flask has taken a while to set up.
* First I had to find a good tutorial, this took a bit of searching, as most applications of flask do not involve a raspberry pi
* Secondly, IDLE (a Python IDE) was not installed on my raspberry pi! I thought this was rather odd, however I did note that their were multiple other Python IDEs installed on the pi. These include Mu and Thonny IDE. I decided to try one of these.
* Bad idea.
* Both Mu and Thonny are both IDEs meant to introduce people to programming who have never programmed before.
* The result of this is that the UI is filled with large, obnoxious buttons and you are restricted to just the inbuilt libraries of python.
* This would no suit as I needed to install Flask which was not an inbuilt library of python.
* After a rather embarrassing amount of time I realised I could just type sudo apt install python3.7 into a terminal window in order to install the desired IDE.
* The next and most strenuous part was trying to remotely access the webserver from my laptop.
* I must have read every stack overflow thread on the matter.
* In no particular order, here are the things I did in order to try and connect the two devices
  + Turn my router on/off
  + Swap port numbers
  + Realise that when you reboot your router the ip addresses are reassigned.
  + “Open particular ports to allow inbound connections by using UFW (Uncomplicated Firewall) and then rebooting the raspberry pi.
  + Toggling Flask’s debug mode
* I was composing my last paragraph of my stack overflow question when suddenly my other tab glowed with the familiar words of “Hello World” in Times New Roman.
* I was stunned that it had worked. It is a shame that I had to try so many different troubleshoots as now I do not know which got it working.

A screenshot of a cell phone

Description automatically generatedFlask Webserver code:

* Pushing my luck, I moved onto the last section, which involved the sending of the file.
* Flask is a very robust library, so all I had to do was call one method and put the file I wanted in the desired directory.
* When I downloaded the file for the very first time, the file appeared as a default file with no extension.

PyCURL receiving code:

A screenshot of a cell phone

Description automatically generated

* However, upon opening this file with WireShark, I was able to see the packet data!
* Truly this has been a triumphant day. Feeling swell

**Side Quests**

* ~~Send a pcap file from the raspberry pi to a laptop using Flask.~~

*Sunday 8th February*

* Today I watched the first lecture of the fast.ai Machine Learning for coders course.
* I attempted to participate with it however when I rented a computer to use via the recommended site Paperspace.
* However the files which where used in the tutorial were not present.
* People in the comments of the video where complaining of this issue aswell.
* After creating multiple “notebook” instances and a couple of hours looking for the filepath using commands I gave up and simply watched the first video to the end.
* I will ask John for the best course of action.
* The content in the video was highly applicable to my situation as in the video the lecturer was analysing a csv file.
* A pcap file is very similar to a csv file so I’m looking forward to watching the next video but I’m unhappy that I could not follow along with the video.
* I also scheduled an appointment with Dr.Gilla for Monday as our Friday arrangement was scuppered by a staff meeting.

*Monday 9th February*

* Had my meeting with Dr.Gilla
* He proposed that I focus on two types of attacks.
* DHCP attacks and HTTP attacks.
* Both are DoS attacks.
* He gave me further detail regarding these types of attacks.
* I wonder however if I fully managed to convey the challenge to him, as I’m not sure he grasped that I was aiming to make something valuable which may not have been created before
* Either way I found the meeting helpful.

*Tuesday 11th February*

* Worked some more on the project, deciding on how I was going to go about writing algorithms to analyse the packet files.
* I’ve decided to opt for python as it is the language I have the most experience in and their seems to be ample library support.
* I chose a library called pyshark which would parse the pcap files.
* However due to the return type of the parse not being iterable, I couldn’t access the elements that I wanted.
* So, I circumvented this by writing the console output to a file and then reading each line of the file as a string.
* This allowed me to create my first piece of diagnostic data.
* A barchart showing the occurrence of each protocol.
* This is valuable, as I had to circumvent the pyshark library in order to obtain the data.
* Feeling pleased.
* I will take a break from project work tomorrow in order to research devices that already exist.
* This will aid in the decision to proceed with the suggestion posed by Dr.Gilla.

*Thursday 13th February*

* Spent some time researching into pre-existing products in my field.
* I have created a new document which contains my findings called “Research Into Existing Products”.
* I only chose two examples, as most of the software available seems the same.
* It is all about gathering data from the network, analysing it and then alerting the appropriate person if something is wrong.
* It was interesting to note that since this is such a new field, there is no commercially available product that directly states that it is using machine learning algorithms to detect network anomalies.
* So I suppose any sort of machine learning solution might be valuable.
* I also took this time to look at a couple of research papers, however I couldn’t understand what any of them where saying.
* I may have to have another go at reading them when I’m a bit more experienced.
* Overall, I think that proceeding with the HTTP/DHCP DoS machine learning detection algorithm seems like the best idea now.
* When built, it could provide an entry point for other students looking to get into the field.
* It occurs to me I may have lost count of the days in this document, so I’m going to create a new task list.

**Task List for Week ahead**

* Determine necessary fields for the machine learning algorithm
* Write the machine learning algorithm code.
* Obtain a meaningful response.

**Side Quests**

* Automatically download the .pcap files from the webserver, so that continuous monitoring can be achieved.

*Friday 14th February*

* Had my meeting with John today.
* I got all the attention because Ben was sick.
* Talked a lot about what I needed to do in the coming week(s).
* I will update my task list:

**Task List for Week ahead**

* Determine necessary fields for the machine learning algorithm
* Write the machine learning algorithm code.
* Obtain a meaningful response.
* Obtain (useful) academic papers in the field.
* Create a working prototype of the system.

**Side Quests**

* Automatically download the .pcap files from the webserver, so that continuous monitoring can be achieved.
* This seems like quite a lot of tasks for one week, however I can always strive for it.
* During the meeting I asked for some help on making my barchart algorithm more efficient.
* John removed the unnecessary file output and instead used a string variable.
* I didn’t think it would make the program much faster.
* However
* My original version took 0:00:01.798357 seconds
* The new version took 0:00:01.568598 seconds
* A difference of 0.229759 seconds
* So an overall efficiency improvement of 12.78%(2.dp)
* Using this new code I created my tutorial video and uploaded it to YouTube.
* <https://www.youtube.com/watch?v=dL3X4IAZVlA&t=97s>
* That should help budding network investigators.

*Saturday 15th February*

* Worked on “cleaning” the data for the machine learning algorithm.
* I’m using a library called pandas which is used for high level data manipulation.
* This is the first time I am using it.
* And as a result of this it is quite hard to work with.
* This is compounded by the fact that PyShark has very little documentation supporting it.
* I’ve learnt a lot about different protocols from this.
* For example, the ARP protocol uses MAC addresses instead of IP addresses and ICMP addresses don’t have source or destination ports and does not use TCP or UDP.
* I have handled these exceptional circumstances.

*Sunday 16th February*

* Continued to work on cleaning the data.
* It is important to get this stage correct, as once cleaned, I simply have to call a prebuilt algorithm from sklearn.
* I also set up a git repository and posted all of the code related to the project there.
* I also added John as a contributor.
* I feel bad that I am not making as much progress as I would have liked over the weekend, however this stage is the most tedious, yet important part of the machine learning process.

*Monday 17th February*

* I took the time to learn more about the fundamentals of pandas.
* It is quite well documented so there were no issues there.
* I don’t like feeling like I’m stuck on something as it niggles away at me every day.
* Even troubles me when sleeping.
* It also doesn’t help that I’m juggling a Java project for another module.
* That project is fine, but since it is a group project, I must coordinate with other people.
* Which would be fine, if people pulled their weight.
* Oh well I suppose that’s life.
* Anyway, pandas is quite like SQL where theirare a lot of different commands and functions that will each do a vastly different thing. Then the complexity comes in by having to “chain” together these commands into one statement (again similar to SQL).

*Tuesday 18th February*

* I am so bored of pandas and data cleaning.
* Within the span of 4 days it as destroyed my will to work.
* But I will work harder as John is always right.
* I decided to have a look into PyCurl (Python wrapper for cURL).
* I managed to download the .pcap file form the raspberry pi, however only in the form of a bytes string, not a real file.
* A lot of threads on the topic simply say to “look at the docs” which is very helpful (not).
* I feel happy with the progress I made with cURL though. It at least feels doable.

*Wednesday 19th February*

* YES!!!
* My first model has been created!

Image showing score of model object (1, as no testing data is present)



* I’m so happy, I feel like I should celebrate.
* Coincidentally I was invited out to a bar to hear John speak.
* I probably shouldn’t put that in the diary though for fear of getting him into trouble.
* Anyway, yes I’m very pleased with myself.
* It is now time to think of the project as a research project.
* The next thing to do is to obtain data regarding attacks.
* I will detail this more in the planning section.

*Thursday 20th February*

* I finished off the cURL code for transmitting files from my raspberry pi.
* This feels good as that was hanging over me for a while.
* It is now time to review my weekly tasks

**Task List for Week ahead**

* ~~Determine necessary fields for the machine learning algorithm~~
* ~~Write the machine learning algorithm code.~~
* Obtain a meaningful response.
* Obtain (useful) academic papers in the field.
* Create a working prototype of the system.

**Side Quests**

* ~~Automatically download the .pcap files from the webserver, so that continuous monitoring can be achieved.~~

Arguably, I have achieved a meaningful response from my model since I have no validation data (so the score should be 1). I’m now turning my focus to dataset retrieval in order to perform an experiment to prove that the model works. I am overall very pleased with the achievements I have made over this week.

*Friday 21st February*

* Had my meeting with John today.
* The top priority now is obtaining suitable training and testing data for the model.
* I feel a bit unhinged, because I’am no longer simply creating things, I am now searching for things.
* Since the issue of finding suitable data for this niche field of cyber security is major, anything I produce from this stage will be of lasting value.
* The only trouble is I’m not sure If I will be able to make anything.
* I also want to create a tutorial for tshark.
* As In retrospect, it is quite poorly documented in a tutorial sense. The documentation however is pretty good.

*Saturday 22nd February*

* Emailed Kiernan McLaughlin, a lecturer at Queens, who may be able to help me on my search.
* Apparently, he has connections with many of the top network IT firms.
* If these IT firms have the data, they may let me use it.
* I also had a look at the application form for Invent NI 2020.
* Will need the help of John to fully complete it.
* I feel a bit powerless in these situations.
* I am totally at the mercy of other people in order to progress. It is not a great feeling. Stuck in perpetual anticipation.

*Sunday 23rd February*

* Made a post on the networking stack overflow site, appealing for data.
* Whilst I cannot progress further, I may as well document what I have done, so I will create the tshark tutorial.
* I have created the tshark tutorial.
* <https://www.youtube.com/watch?v=NvKiyHCRXrU&feature=youtu.be>
* I wish I had this sort of video when I began this challenge as it goes over the basics of tshark.

*Monday 24th February*

* Tried to have a look at some data sets I had downloaded.
* I couldn’t open a 6GB file on Wireshark. It said there was an internal unknown error.
* I could read the file using TShark though.
* My home wifi was getting upgraded today, so I didn’t have access to the internet, which limited my progress.

*Tuesday 25th February*

* I stumbled across a very highly rated paper relating to my field titled “An empirical comparison of botnet detection methods”
* The paper covered a variety of different attack methods and vectors, from Malware to DDOS attacks.
* They had created a very public dataset which I could use.
* I will have to speak with John about this however, as I’m unsure as to what datasets I should be using.
* I also watched the second lecture in the fast.ai course.
* In the lecture, the random tree was illustrated in a graphical form.
* I made a graphical form of my model however I discovered that my model never splits into separate trees.
* I must investigate this.

*Wednesday 26th February*

* More time spent looking for datasets.
* I must say this is not my favourite part of the project.
* Still the word document I have produced is of value so I’m happy about that.
* I just must remember the goal is not to reach the project goal but instead make things of value.

*Thursday 27th February*

* Information overload from John today regarding project.
* Going to change the goal of the project so that the device will be able to remove infected computers from a network.
* To detect infected computers, we monitor the packet data from those computers.
* This is all well and good however I really want to get a model working using packet data. I suppose that will be put on hold.

*Friday 28th February*

* Had my meeting with John today regarding project.
* The next hand in is due at the end of the this coming week.
* The task is to produce a “blog post” discussing the project.
* It is assumed that the “blog post” will be in a rough form and that the following 4 weeks will be used to tidy it up.
* It will take the form of an HTML webpage.
* I mentioned that I may not be able to get the model finished.
* As a contingency plan, John mentioned that I should have to look into open source firewalls, as the idea now is to use one of these fire walls to flag up that there is an attack (instead of the machine learning algorithm) this will then cause the device to break the affected computers connection to the rest of the network, in order to contain the problem.
* The substitution of the AI component for the firewall will allow the device to work, at a later stage we would add the ai component.

*Friday 29th February*

* I attended SISTEM 2020 today from 8am -10pm so needless to say I got no project work done, however I will go over what I learnt at this student conference.
* I attended a block chain workshop ran by MasterCard. I found it fascinating as I had always considered crypto currency and “mining” as fanatical. They showed how secure the whole process was and how the currencies themselves are backed by their encryption algorithms. I got to play around with a small scale version of it. It is interesting as in many respects a block chain is like a database which is more permanent and ever expanding.
* I attended a VR and AR conference which was okay. The speakers where students doing game development and those who had just graduated. They didn’t really have much to say, except plugs to follow them on twitter (sigh).
* Finally, I attended an employability workshop ran by Microsoft and it was one of the most interesting talks I have ever been in. The speaker , Stephen Howell , coloured is talk with anecdotes of his two children who suffer with Asperger’s and how he developed software to aid their learning. Must make a linked in account.
* Overall a great learning experience and I a very glad I went.

*Sunday 1st March*

* Today I did the research outlined on Friday.
* I feel under qualified to talk about business as I have never done a subject/module in business studies or economics.
* Still I gave it my best shot.
* Can’t progress further with the AI as John is working on an error, I simply can’t get my head round.
* Feeling tired. Doing lots of reading makes you tired. Also coming back from Dublin on the previous day makes you tried.

Now that the direction of the project has been re-addressed, I can now look at establishing some tasks for the week ahead.

**Task List for Week ahead**

* Write the draft blog post
* Finish the AI modelling
* Submit Invent NI application.

*Monday 2nd March*

* Spent the whole day trying to make my model work.
* This was very time consuming and frustrating.
* Working with pandas is very difficult. Even when you know your solution should work, issues still arise.
* It stems from the fact that Pandas Dataframes are not arrays, yet behave in arrays in some instances.
* For example you could type df[0] to return the first row but you would have to use df.value[0] to obtain the value of that first row.
* Simply frustrating.
* I didn’t get my model to work unfortunately.
* I am *extremely* close however the error I am encountering is due to sklearns .predict() method. Some serious debugging will need to occur in order to fix it.
* Still once that error is fixed, nothing else could go wrong
* As it is the last line of the code.
* Still, feeling annoyed.

*Tuesday 3rd March*

* Today is the day I start work on my blog post.
* My target audience will again, much like the how to guide, will be students.
* Why have I chosen this target audience again? There are multiple reasons:

1. As a student, I personally find conversational pieces more interesting than a more structured text ( i.e textbook)
2. Experts on the subjects may not care/have the time for a blog post which doesn’t get straight to the point. As this blog is an overview of the entire project, it may not contain anything of relevance to an expert until later.
3. The blog may not go into as much detail as experts would like.
4. As a student, it is refreshing to have a text assume some knowledge of the subject, but not assuming the more niche and tricky areas.

* I believe this reasoning is sound.
* The tone of the blog post will be conversational yet still retain an air of professionalism.
* The blog post will go from the beginning of the project up to this point. I will use my daily diary as a guide for chronology.
* I have opted for a similar style of webpage as that of my How-To Guide, as I find it appealing. I was never the best at web design but the text is clear and readable, which is the most important thing.

*Wednesday 4th March*

* Continued writing the blog post.
* It is quite difficult to blur the line between conversational and instructional.
* But I feel that with this blog, supported by the github repository, someone could easily pickup the project from where I left off.
* Must finish this off today as tomorrow I’m out all day at the EPS outreach event for the QCS and then at the formal.
* GitHub repository is up to date.

*Thursday 5th March*

* Bringing the joy of computing at Queens to children in Newry today (7-4).
* In the evening, QCS formal (7-11).
* So I’m writing this entry in advance.

*Friday 5th March*