PLAN OF ACTION

Honeyjar and Malware studies

ABSTRACT

This document describes the background of this project, as well as being a clear summary of expected actions and responsibilities for all project members involved.

1 VERSION MANAGEMENT

Version	Author	Date	Comments
0.1	Morcel	20-02-2018	Added chapters
0.2	Alexander	21-02-2018	Added front page, as well as further defining of the problem definition
0.3	Morcel & Alexander	26-02-2018	Improved Feasibility and Risks and applied Moscow-method to Project boundaries
0.4	Alexander & Morcel	27-2-2018	Added main + subquestions, small corrections
0.5	Alexander & Morcel	28-2-2018	Expanded on questions, added context to BCM, fixed spelling
0.6	Alexander & Morcel	3-2-2018	Expanded and added appendices
0.7	Alexander & Morcel	3-3-2018	Expanded on appendices, managed spelling in document

2 Table of Contents

1 Version management 1

3 Background 3

4 Problem Definition 5

5 Research Question and subquestions 8

6 Project Boundaries 9

7 Tools of Research 10

8 Organization 11

9 Costs 13

10 Feasibility and risks 14

11 Appendices 0

3 BACKGROUND

3.1 ORIGINS

The project originated from the University Of Aalborg (AAU). This project dates back to 2013 and was a master thesis project. In this first project a first architecture was developed and a first implementation was made. The motivation to initiate the project was:

"The need for good ground truth data to be used for research and training machines through machine learning algorithms."

While analyzing the network traffic the group realized that there were surprisingly few usable traces available. Because of this researchers still use old datasets such as the KDD'99 dataset from 1999. From this project onward a basic architecture has been created.

This architecture is very complex and requires improvement. So since 2013 this has been done on the initial setup. This setup consists of thirty old computers.
There have been several projects that were fairly successful. One project even got the Danish Tele Award in 2016. The project group had analysed 300.000 pieces of malware. They managed to find patterns which differentiated malware from cleanware.

Talaia is a Spanish company and has grown interest in the subject. They are in close collaboration with the project and provide feedback to the project where necessary.

3.2 ARCHITECTURE

Since then the term Honeyjar has been used as a host for many different projects. The term Honeyjar consists of many different projects. It is then split between different sections. Imagine this as a tree with a lot of different branches. They all go in a different direction but have the same roots. The Honeyjar in this projects consists of three parts;

3.2.1 The test environment

This environment is meant to look like a real network, not a virtual one. Furthermore it is possible to create virtual machines which can be

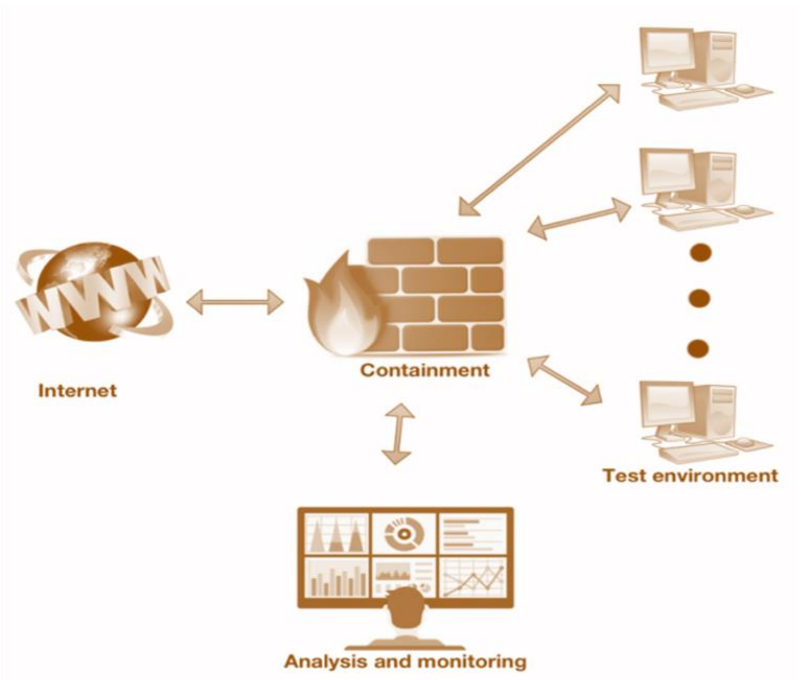


Figure 1: A Honeypot system

created and taken down again, automatic installation of software and operating systems as well as an emulated internet environment.

3.2.2 Containment

This part is created so that it is possible to connect to the outside but preventing harmful traffic to go outside.

3.2.3 Analysis

Here data is collected for analysis. With this data it is possible.

3.3 Proceedings

To continue the research there had to be a better setup, thirty old pc's wasn't sufficient. It also frequently happened that a computer broke down and caused problems with experiments. In 2017 a Danish foundation granted two new and very powerful servers to continue this research on. This is also the reason that new Honeyjar projects have been initiated.

These servers will be used for the Honeyjar project as well as "white hat hacker"-training. Both the project as well as the training will require virtual machines and networks. Another grant has been provided for the training as well, to create a Danish training platform for cyber security. Another perk of the new hardware is that there is a faster connection to the Danish Research Network.

3.4 Project group composition

3.4.1 Aalborg University

All project members except for the members from AAU are new to this project. The project members from AAU have worked on this project half a year prior to the rest and have made a basic simple Android-based honeypot system.

Their assignment was to improve cybersecurity. They took their research to Android phones since according to their research, people store a lot of sensitive data on their smartphones and cybersecurity on smartphones isn't deemed urgent by the masses at this point in time.

3.4.2 Saxion University of applied sciences

Saxion has brought two business IT students to the project. Their job is to create a system that is appealing for the market. In what form can this be a product that will be appealing to a potential customer? Other tasks that these students are specialized in are:

- Using different kinds of modeling such as BPMN, Database models, Data Flow Diagrams
- Translating customer needs into IT solutions
- Experience in working within a multidisciplinary group
- Facilitating the team where needed

3.4.3 Abdullah Gül University/University of Technology and Life Sciences

The students from AGU and UTP will both improve the current system as well as facilitating a network analysis to determine what is and isn't malware. With this analysis they will be able to create a basic machine learning algorithm.

4 Problem Definition

In this paragraph is described what the reason for the project is, as well as the desired end result when the project is finished. This will be accomplished using research questions. These questions will divide the project into smaller chunks. When all the sub questions have been answered the main issue will almost automatically be answered.

4.1 Statistics

Companies as well as regular consumers have a lot of sensitive data stored on their smartphones, for example (but not limited to):

- Credit card data
- Compromising photos
- Passwords
- E-mails
- Customer data

This in itself is worrisome, however the popularity of smartphones is rising as well. At this point in time smartphones outsell PC's. Computers have been susceptible to malware since they were first introduced to the market, however it seems that Android phones, while vastly gaining popularity over PC's, are still relatively under protected to Malware attacks. This is represented in the graph below, which is derived from a survey done by the project members from AAU.

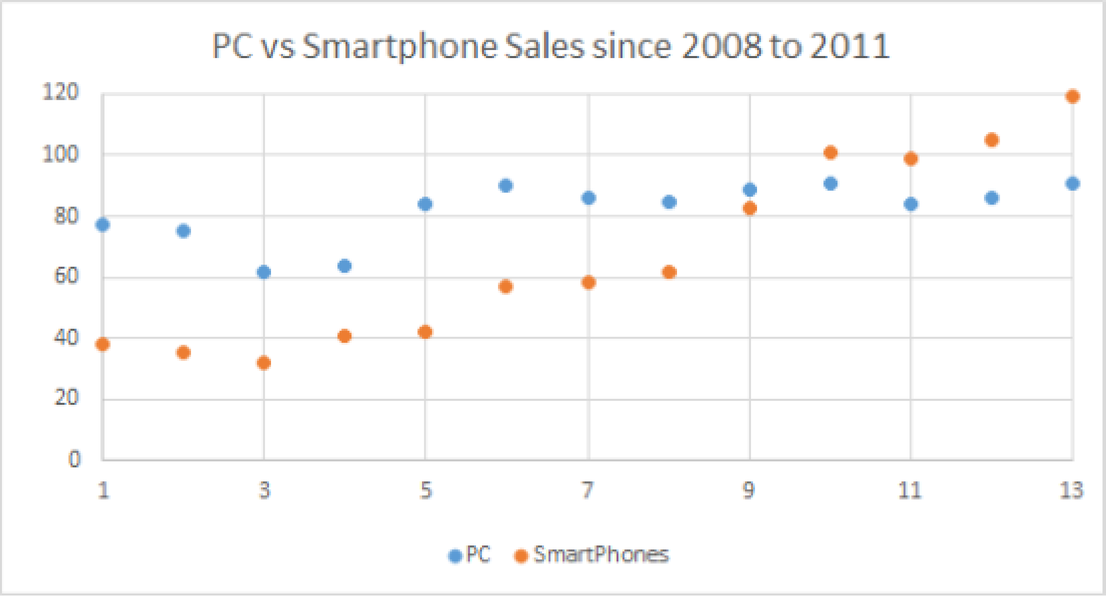


Figure 2: Smartphone sales vs PC sales

To exploit the fact that the market hasn't really focused on security on Android phones yet, the project members from AAU created a basic Honeypot system.
A Honeypot, as to be derived from its name, is a virtual machine that emulates a system, in our case an Android Phone, presenting it in such a way that it becomes interesting for malware to attack. This Honeypot system in its current state is still very basic and needs improving.

The three computers in the example will be running an emulated version of Android, making the emulation seem as realistic as possible, i.e. emulating human behavior. Once these systems are infected through the internet, the malware will be contained, and analyzed to determine what the category of the Malware is.

4.2 GOAL

The goal of this Honeyjar is to provide intelligence on the nature of the Malware that is captured and contained in the quarantine. Furthermore, based on the intelligence that is gained in containment, the Honeyjar algorithm has to show improvement in recognizing Malware.

There are a couple of ideas to create something similar to an antivirus system as well as an algorithm that can be sold to a company. To see how this might look the author would like to forward the reader to Appendix II.

4.3 END RESULTS

At the end of this project the following will be finished and accomplished:

1. A functional Honeyjar system that shows improvement in recognizing Malware through Machine Learning

2. A Business Plan to describe the Business aspects of the Honeyjar and how to make the system profitable
3. An interactive GUI-system with the functionality to control the virtualization process and generate relevant data sets in the form of PCAP files

5 RESEARCH QUESTION AND SUBQUESTIONS

5.1 RESEARCH QUESTION

The main research question of this project reads as follows:

"How can a Honeyjar-system for Android be made relevant for our target audience and turned into a profitable product?"

This is the question that needs to be answered for this project. This will happen through the answering of the following subquestions.

5.2 SUBQUESTION I: WHAT IS A HONEYJAR-SYSTEM AND WHAT IS ITS RELEVANCE IN CURRENT TIMES?

In the first subquestion the group will research the current state of the Honeyjar, what it should do, as well as defining the relevant information to study for this project.

5.3 SUBQUESTION II: WHY AND HOW SHOULD OUR TARGETED AUDIENCE USE A HONEYJAR-SYSTEM?

In the second subquestion, the aforementioned gathered research information will be put inside a business perspective to discover what businesses deem relevant in a Honeyjar product such as ours.

5.4 SUBQUESTION III: WHAT CAN A HONEYJAR MEAN FOR OUR TARGETED AUDIENCE?

In the final subquestion the Honeyjar has been activated and the project group will analyse the output data and will present this in a business context.

6 Project Boundaries

The following paragraph defines what the project will and will not deliver. The purpose of this paragraph is to define the scope of the project, and making clear what should not be worked on during the project, maximizing efficiency. For this, the Moscow-method is used.

6.1 Method

The Moscow-method is a way of prioritizing requirements. Moscow stands for:

1. M: Must haves
2. S: Should haves
3. C: Could haves
4. W: Won’t haves

Must haves describe the essential requirements of the project. These are the core requirements, and must be attained at all costs. The next step is *Should haves*, which describes requirements that are deemed very useful but aren’t integral to the success of the project. *Could haves* is another step below that, and can be seen as requirements that can be added if there’s time left. *Won’t haves* describe the requirements that will not be implemented.

Using this method sets a clear scope.

6.2 Applying the method

Must haves	Should haves
Functional Machine Learning algorithm	Dashboard
Virtual simulation of Android phones	Virtual Android phones that are as ‘‘human’’ as possible
An improved version of the already known ML: algorithm	Antivirus designs for future projects
Self-categorizing functionality of Malware	
Clear Business plan	

Could haves	Won’t haves
Web-application	Antivirus functionality
Real-time data collection	Apple iPhone-functionality
Demo for potential customer	

7 TOOLS OF RESEARCH

In this chapter will be described which tools and methods the project group will use to gain the information necessary to conduct the project.

7.1 DESK RESEARCH

As all project members are located in different countries, some necessary research has to be done by collecting information through the internet, books, and other sources. This includes research such as (but not limited to):

- Tutorials on Virtual Machines, servers, Android, etc.
- Manuals for the server
- Documents describing Honeyjar systems
- Articles
- Forums
- Troubleshooting

7.2 FIELD RESEARCH

For information that needs to be actual and from a unique perspective, field research will be performed. Information will be collected through interviews with companies focused on security, as well as other relevant people which are to be selected during the course of this project.

Through the conducting of interviews information will be collected that will be summarized through axial coding, making the information usable and measurable. This information can then be put into a program like Excel to showcase which results of the interviews are the most recurring, thus giving the project group insight in what is considered relevant.

7.3 SHARING OF KNOWLEDGE

To make sure everyone in the group is up-to-date with the progress that’s being made, during every virtual meeting all project members share what they have been researching the week prior. This keeps every project member informed, and allows for room for discussion.

8 ORGANIZATION

In this paragraph the project group is defined and elaborated upon.

The current project group consists of nine people coming from four different universities.
Everyone in the group are equal and have the same obligations. The group works together and makes sure everyone is able to do their part in the project. Since everything is intertwined it is important the group keep communicating and gives each other feedback where possible.

As said in background the project group consists of nine different people spread across four different universities. There will also be a supervisor which is closely involved with the project. On top of that every university has a teacher that will guide the student throughout the process.

Function/role	Name	Tasks
Stakeholder	Valentín Carela	Giving feedback throughout the project.
Supervisor/AAU Teacher	Jens Myrup Pedersen	Supervising/guiding the whole group during the project
Supervisor	Etto Salomons	Supervising and guiding the business students during the project
Supervisor	Mehmet Şükrü Kuran	Guiding Project member
Supervisor/UTP Teacher	*Unknown*	*Unknown*
Programmer	Jacob Vejlin Jensen	Improving the Honeypot system
Programmer	Peter Bolstad Møller	Improving the Honeypot system
Programmer	Daniel Britze	Improving the Honeypot system
Programmer	Robert Nielsen	Improving the Honeypot system
Programmer	Magnus Stensli	Improving the Honeypot system
Programmer/Networking	Ahmet Türkmen	Improving the Honeypot system and doing network analyses
Networking	Anna Switala	Improving the containment zone of the Honeypot
Business IT	Alexander Pluimers	Creating a vision as well as a business plan for the product
Business IT	Morcel el Ouahbi	Creating a vision as well as a business plan for the product

The following table contains contact information from everyone who is involved during the project.

Name	Phone number	E-mail address
Valentín Carela	+34 937 379 379	vcarela@talai.io
Jens Myrup Pedersen	+45 99 40 87 71	jens@es.aau.dk
Etto Salomons	+31 6 22 49 06 77	e.l.salomons@saxion.nl
AGU Teacher	05325210283	sukru.kuran@agu.edu.tr
UTP Teacher	*Unknown*	*Unknown*
Jacob Vejlin Jensen	+45 25 14 09 90	jvje17@student.aau.dk
Peter Bolstad Møller	+45 26 95 30 33	pmolle17@student.aau.dk
Daniel Britze	+45 28 64 31 17	dbritz17@student.aau.dk
Robert Nielsen	+45 28 76 82 75	rnni17@student.aau.dk
Magnus Stensli	+45 51 96 45 60	mstns17@student.aau.dk
Ahmet Türkmen	+90 541 204 37 48	f.ahmet.turkmen@icloud.com
Anna Switala	+48 782 33 55 97	annswi004@utp.edu.pl
Alexander Pluimers	+31 6 44 16 33 38	314831@student.saxion.nl
Morcel el Ouahbi	+31 6 21 35 29 90	423819@student.saxion.nl

9 COSTS

This paragraph will highlight the costs of the project and what the project group is supposed to deliver when the project concludes.

There is a budget of €40.000 for the whole project. This will include two trips of approximately five working days. Since the project group involved in this project is very large (nine people) two meetings is the bare minimum to keep working effectively. These meetings will make sure that ideas can be exchanged without the barrier of the internet, this will make the collaboration a lot easier.

At the end of the project it is expected that something has been delivered. The following will be delivered at least:

- A base for the machine learning which shows improvement through data provided
- A base infrastructure in which can be expanded further in future projects
- A business plan for a potential end product

When the above has been delivered the project is deemed successful and thus will the investment be worth it.

10 FEASIBILITY AND RISKS

This paragraph will describe whether the project is feasible enough to continue. The question that needs to be asked is: When this project is done, will it deliver results that are desirable?

10.1 FEASIBILITY

The paragraph costs has already made a clear understanding of what the project will deliver and if, at the end, it will be feasible. While the project group is aware that this is just an estimate it still gives a good idea of the project ahead. There are a lot of variables that come into play when the project is ongoing, so it is impossible to predict what will happen during the project. However with the current prognosis, the project has a solid base to continue on.

10.1.1 Money

Some money has been reserved for this project. The money is meant for all of the group members to get together and discuss progress on the project. Since the project group is divided into four countries this is mandatory. For more information about money check the paragraph "Costs".

10.1.2 Infrastructure

Since our project group is intercultural it is important to keep in touch through means like the internet and phones. The group is scattered around four different countries so a language barrier can be a problem as well so everyone in the group has to take that into account when communicating with each other. In the paragraph "Organization" is an overview of the contact credentials of each team member. This paragraph also includes the different teachers from the universities.

10.1.3 Materials

To successfully execute this project, proper materials are needed. One of these materials is the server which will host the HoneyPot system. With this server in place it will be possible to access and do research on the system. Other materials that are essential for everyone is a proper internet connection, hardware to work on and the platforms the group is going to work on:

1. Git
   For sharing documents, and relevant data. Git's advantage is that it uses Version Management automatically.
2. Discord
   For communicating between the group. Some advantages of Discord are the usage of channels, separating groups within the project so only the relevant information is shared with the relevant project members
3. Dropbox
   For the Business members only. Dropbox is a quicker way to share documents, which is the bulk of information for the Business students.

10.1.4 Expertise

As previously discussed the project group consists of nine people scattered across four countries. Everyone in the group has an expertise which will be beneficial to the project.

All these expertises will come into play when the project is running and everyone is dependent on each other in some way. For instance this plan of action is important for every project member.

10.1.5 Time

All project members are expected to work on the project every day, ranging from a few hours every day to full-time. Even though the project group is the largest among the EPIC groups, the project is big and ambitious, so maximal effort is required. The business students from Saxion are working on this part-time. They will have the most time available from everyone

10.2 RISKS

10.2.1 Different deadlines

As all project members work on this project in a different year and/or in a different format (i.e. the Business project members are doing this project for their Minor studies, while the Danish group is doing this for their first year of school), the deadlines for each project member differ. This opens up the possibility that project members are going to follow their own agenda as opposed to working towards a common deadline as a team. To prevent this, the group created a planning with each deadline carefully noted.

10.2.2 Miscommunication

Due to the intercultural nature of this project, the project members expect there to be miscommunication to a certain degree and cultural differences. The project members try to avoid this by setting up the following ground rules:

During meetings:

1. Give each other the space to talk, no interrupting
2. If something isn't clear, ask
3. One person is assigned to take notes during the meeting

In general:

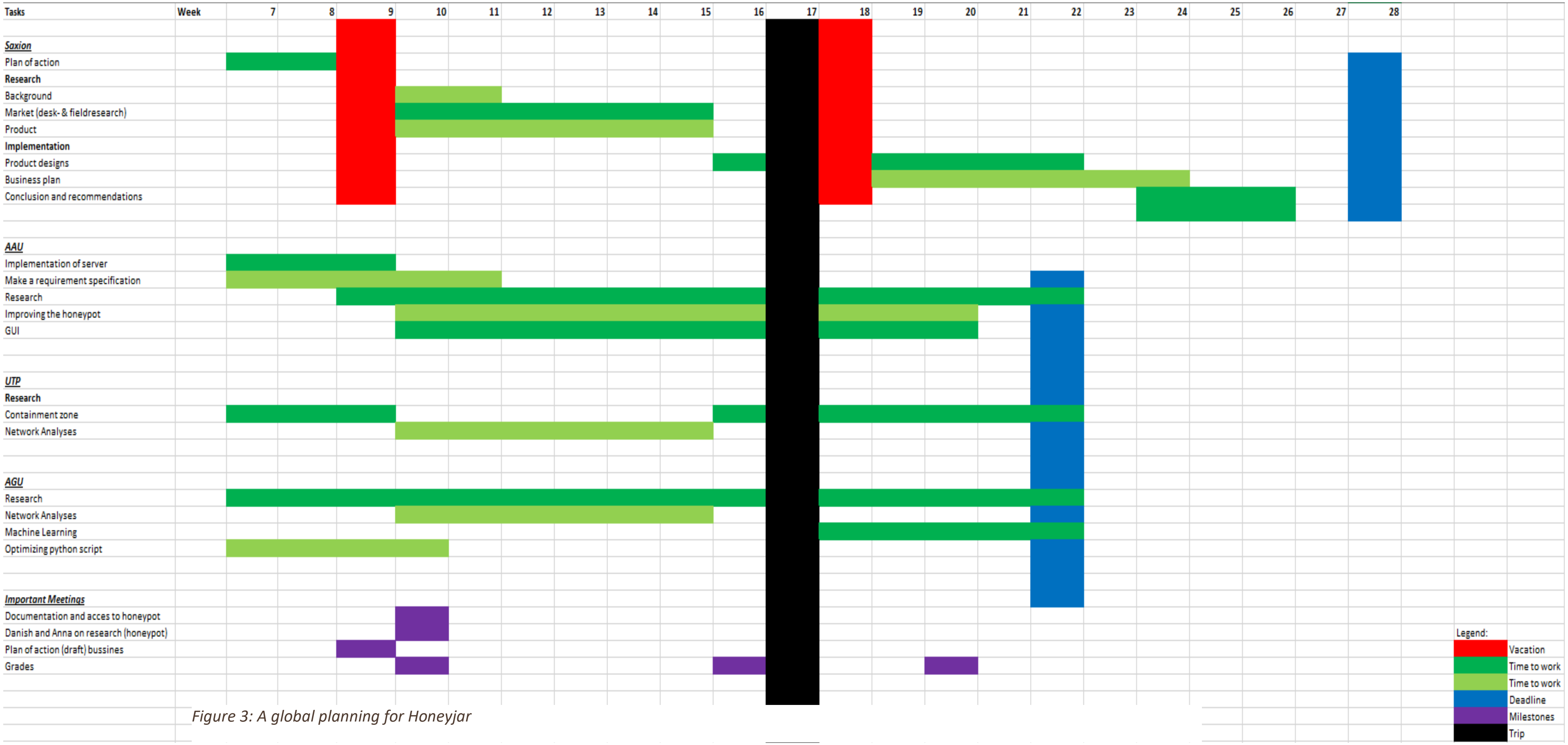
1. A global planning has been set up in Google Calendar
2. Meetings at least once a week
3. Subjects to discuss during the meeting must be submitted 2 days prior
4. Project members are expected to be able to reply in a window of 24 hours

Risks	Causes	Measures
Financial		
End product not feasible within set investment range	Underestimation of funds needed for end product	Adjust investment range accordingly
Organisatorial		
Project members aren't living up to the contract	Various, could be anything ranging from illness to lack of interest, to miscommunication	Communicate with said project member, ultimately supervisor
End product is not deemed useful by businesses	Misunderstanding of what companies are looking for	Adjust requirements of end product to make it interesting for businesses again
Interested businesses already have a Honeyjar solution in place	The business has a functional Honeyjar system	Add requirements that make our Honeyjar system stand out
Failing to define why businesses should use our Honeyjar	Scope too large	Define a set of key selling points for businesses to get interested
Somone in the project group is unable to continue the project	Someone got sick for a long period of time. Another reason might be personal issues.	Share all the work that has been done and set at least two people on one task.
Technical		
Usage of personal data	Risk of law infringement, depending on country where the data is collected from	Only use personal data from countries where it's usage in a Honeyjar is legal

Malware not attracted to the Honeyjar	Lack of relevant data to attract Malware	More research needs to be done on what attracts malware
Machine learning giving false negatives	Regular data being identified as malware	More research needs to be done on how to correctly authenticate data
Server outage	Any possible reason	Secure data on private back-up curated by Daniel
Our files are removed from the cloud	The service we use has a hack and lost all our files	Back-up all data on a local storage or another cloud service.

11 Appendices

11.1 Appendix I: Planni



Tasks	Week	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
				🟥								⬛	🟥										
Saxion				🟥								⬛	🟥										
Plan of action		🟩	🟩	🟥								⬛	🟥										🟦
Research				🟥								⬛	🟥										🟦
Background				🟥	🟢	🟢						⬛	🟥										🟦
Market (desk- & fieldresearch)				🟥	🟩	🟩	🟩	🟩	🟩	🟩		⬛	🟥										🟦
Product				🟥	🟢	🟢	🟢	🟢	🟢	🟢		⬛	🟥										🟦
Implementation				🟥								⬛	🟥										🟦
Product designs				🟥							🟩	⬛	🟥	🟩	🟩	🟩	🟩						🟦
Business plan				🟥								⬛	🟥	🟢	🟢	🟢	🟢	🟢	🟢				🟦
Conclusion and recommendations				🟥								⬛	🟥						🟩	🟩	🟩		🟦
												⬛							🟩	🟩	🟩		🟦
												⬛											
AAU												⬛											
Implementation of server		🟩	🟩	🟩								⬛											
Make a requirement specification		🟢	🟢	🟢	🟢	🟢						⬛					🟦						
Research				🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	⬛	🟩	🟩	🟩	🟩	🟩						
Improving the honeypot					🟢	🟢	🟢	🟢	🟢	🟢	🟢	⬛	🟢	🟢	🟢		🟦						
GUI					🟩	🟩	🟩	🟩	🟩	🟩	🟩	⬛	🟩	🟩	🟩		🟦						
												⬛					🟦						
												⬛					🟦						
UTP												⬛					🟦						
Research												⬛					🟦						
Containment zone		🟩	🟩	🟩							🟩	⬛	🟩	🟩	🟩	🟩	🟩						
Network Analyses					🟢	🟢	🟢	🟢	🟢	🟢		⬛					🟦						
												⬛					🟦						
												⬛					🟦						
AGU												⬛					🟦						
Research		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	⬛	🟩	🟩	🟩	🟩	🟩						
Network Analyses					🟢	🟢	🟢	🟢	🟢	🟢		⬛					🟦						
Machine Learning												⬛	🟩	🟩	🟩	🟩	🟩						
Optimizing python script		🟢	🟢	🟢	🟢							⬛					🟦						
												⬛					🟦						
												⬛					🟦						
Important Meetings												⬛					🟦						
Documentation and acces to honeypot					🟪							⬛											
Danish and Anna on research (honeypot)					🟪							⬛											
Plan of action (draft) bussines				🟪								⬛											
Grades					🟪						🟪	⬛			🟪								
												⬛											

Legend:	
🟥	Vacation
🟩	Time to work
🟢	Time to work
🟦	Deadline
🟪	Milestones
⬛	Trip

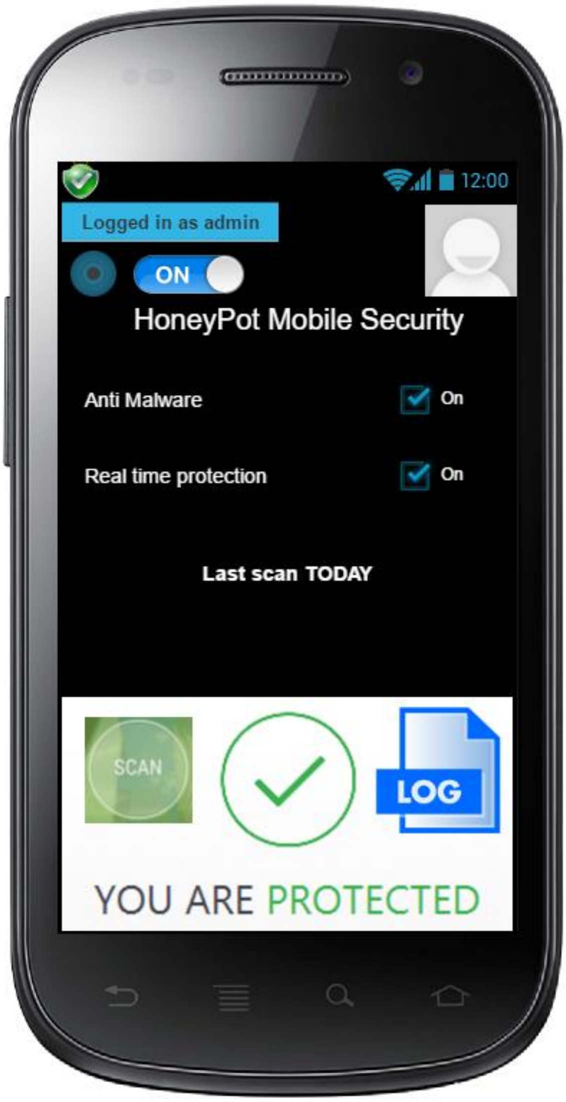
Figure 3: A global planning for Honeyjar

11.2 APPENDIX II: GUI MODEL

In the illustrated example is a mockup of a possible GUI for our Honeyjar-system. This demo features various functionalities;

1. On/off switch for the Honeyjar
2. User login
3. User profile
4. Information on latest scan
5. Checkboxes to turn functionalities of the Honeyjar on or off.

It is worth reiterating that this is a demo, meaning that the end result will likely differ from this mockup. However, this mockup is an excellent way of showcasing the possibilities of our system.



Logged in as admin

HoneyPot Mobile Security

Anti Malware ☑ On

Real time protection ☑ On

Last scan TODAY

YOU ARE PROTECTED

Figure 4: GUI Demo for Android

11.3 Appendix III: Information Flow Diagram

To illustrate an overview of the possible information flows, the following diagram has been made.

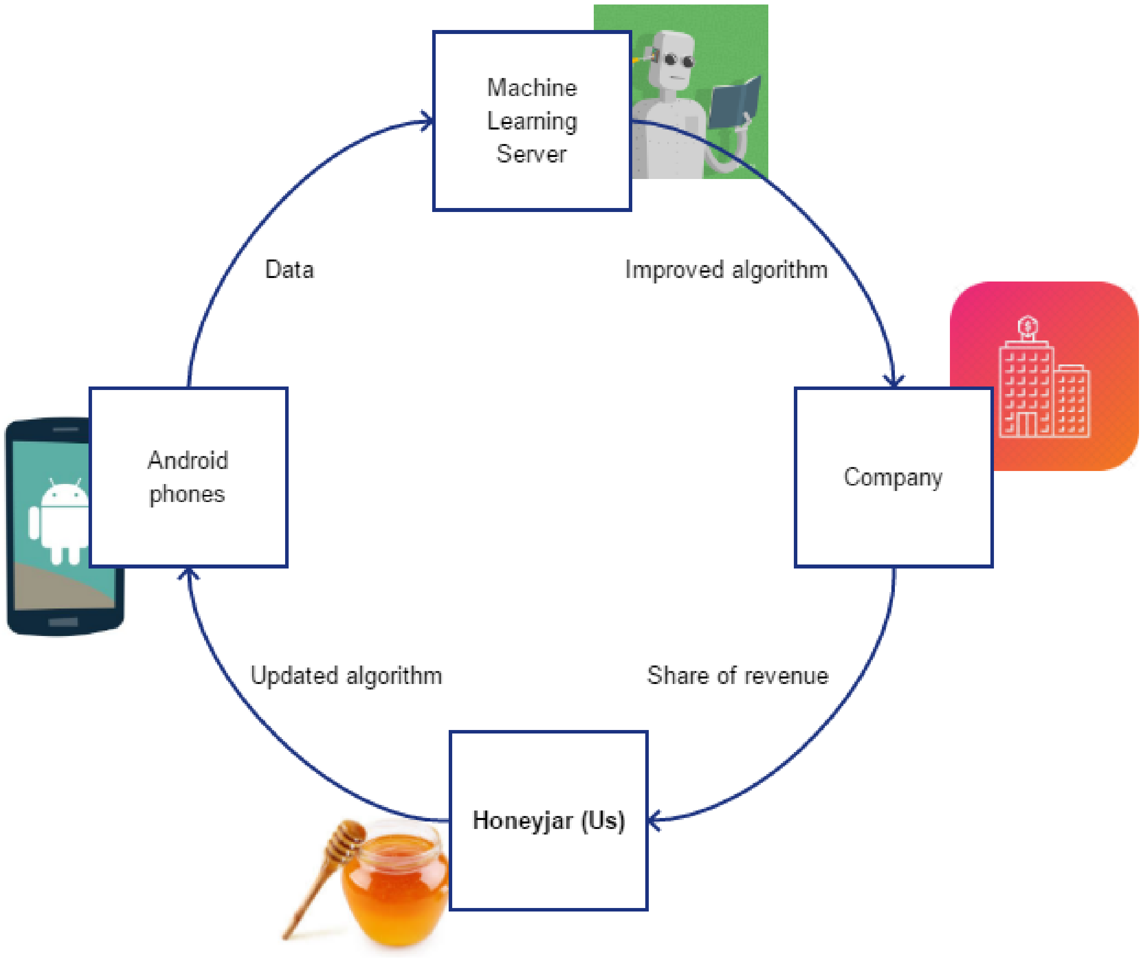


Figure 5: Information flow Diagram

Say that the Android phones function as a Honeypot. They collect data (containing malware), and the app on the Android feeds this to the Machine Learning Server we, as Honeyjar, provided. The algorithm gets improved, giving the company to whom Honeyjar sold the Machine Learning algorithm to information on the collected malware. Their customers pay the company on a subscription basis. For providing the Machine Learning algorithm, we as Honeyjar get a share of the revenue, also based on a subscription basis – however this time it's the company that pays Honeyjar on, for example, a monthly basis. The algorithm is now updated, and can be fed back into the Android phones, repeating the entire process.