2018

EPIC Report: Honeyjar

Saxion University of Applied Sciences

Abdullah Gül University

Aalborg University

1 CONTENTS

2 EARLY PROCESS

2.1 WHAT DID WE WORK ON?

The work to create the HoneyJar will be split into 3 different parts, first one being a honeypot system. A honeypot system is a safe virtual environment where experiments and live testing can take place. It is where the potential malware will get injected into and analyzed to determine whether or not it is actually malicious. When the analysis is done the files will pass through the system or be contained within the virtual environment, where it will be used to improve the security further thanks to the second part of the HoneyJar system, the machine learning.

At the core of the HoneyJar is an algorithm based on machine learning. The algorithm is the part that determines what files are malicious and vise versa. The algorithm being built on machine learning has the capability to improve if provided with the necessary data, which it will get from the analysis done by the honeypot system.

As the third part of the HoneyJar is the business model. The business perspective is what is going to make the system possible, as no matter how good something is it is useless if no one uses it. By making sure the product is marketed correctly it will allow it to sustain it as its own product.

2.2 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.

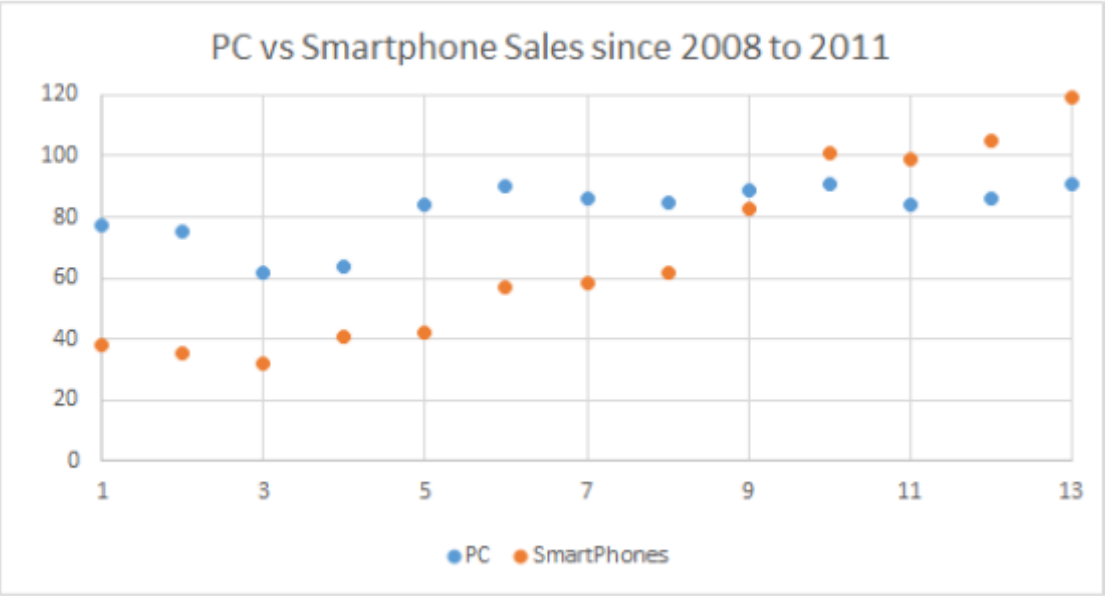


Figure 1: Smartphone sales vs PC sales

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.

2.3 GOAL FOR THE INDIVIDUAL PARTS

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.

3 APPROACH

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

3.1 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

3.1.1 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.

3.2 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. By having a meeting leader, a secretary and an agenda planned the day before, the meetings became more structured which helped getting as much as possible out of every meeting.

4 JOINT WORK

Riga

To start off the collaborative working process a seminar in Riga planned by EPIC. Here the participants would meet up and have courses on teamwork and planning, that would keep work efficient down the road.

Tuesday February 13

08.50	Departure to university
09.00-09.15	Briefing of the day
09.15-11.00	Assessment workshop and mutual expectations in groups
11.00-12.30	Presentation techniques workshop
12.30-13.00	Lunch
13.00-17.00	Problem analysis: Working in groups + supervisor meetings
17.00-17.15	Evaluation and programme for tomorrow
17.15-19.00	Social/cultural activities
19.00	Dinner TBA

Wednesday February 14

08.50	Departure to university
9.00-9.15	Briefing of the day
9.15-11.00	Entrepreneurship workshop
11.00-12.00	Workshop on online collaboration tools
12.00-12.30	Introduction to Agile Software Development (1)
12.30-13.00	Lunch
13.00-14.00	Introduction to Agile Software Development (2)
14.00-15.00	Project planning lecture and introduction of templates
15.00-17.00	Working in groups + supervisor meetings
17.00-17.15	Evaluation and programme for tomorrow
17.15	Quality committee meeting
19.00	Dinner TBA

The plan epic made for Riga lasted every day from 08:50-19:00 on average. The first couple of days were focused around meeting with the group that we would be working together with and learning how to do so efficiently. The remaining 3 days would be a combination of learning Agile Software Development and working in our given groups. When working in groups our prime objective was to lay a plan that everyone could agree on and that would allow everyone to be a part of the HoneyJar. From the very get go we had agreed to separate the project into different parts that together would form a final project. We did this due to everyone having to turn something different at every university and as such we saw it as being the optimal choice.

4.1 HOW WAS THE PROJECT SPLIT UP?

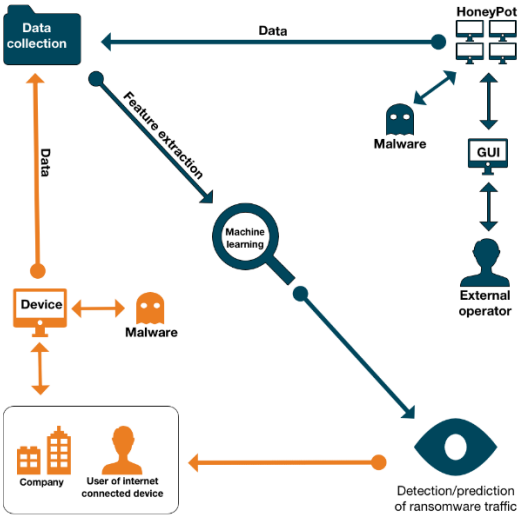
The work to create the HoneyJar will be split into 4 different parts, first one being a honeypot system that the AAU group is responsible for. A honeypot system is a safe virtual environment where experiments and live testing can take place. It is where the potential malware will get injected into and analysed to determine whether or not it is actually malicious. When the analysis is done the files will pass through the system or be contained within the virtual environment, where it will be used to improve the security further thanks to the second part of the HoneyJar system, the machine learning.

At the core of the HoneyJar is an algorithm based on machine learning developed by AGU. The algorithm is the part that determines what files are malicious and vice versa. The algorithm being built on machine learning has the capability to improve if provided with the necessary data, which it will get from the analysis done by the honeypot system.

As the third part of the HoneyJar is the business model made by SAX university. The business perspective is what is going to make the system possible, as no matter how good something is it is useless if no one uses it. By making sure the product is marketed correctly it will allow it to sustain it as its own product.

The fourth part is to improve on the architecture and search for additional features in order to improve the HoneyJar at its very core. This part is handled by (Anna uni)

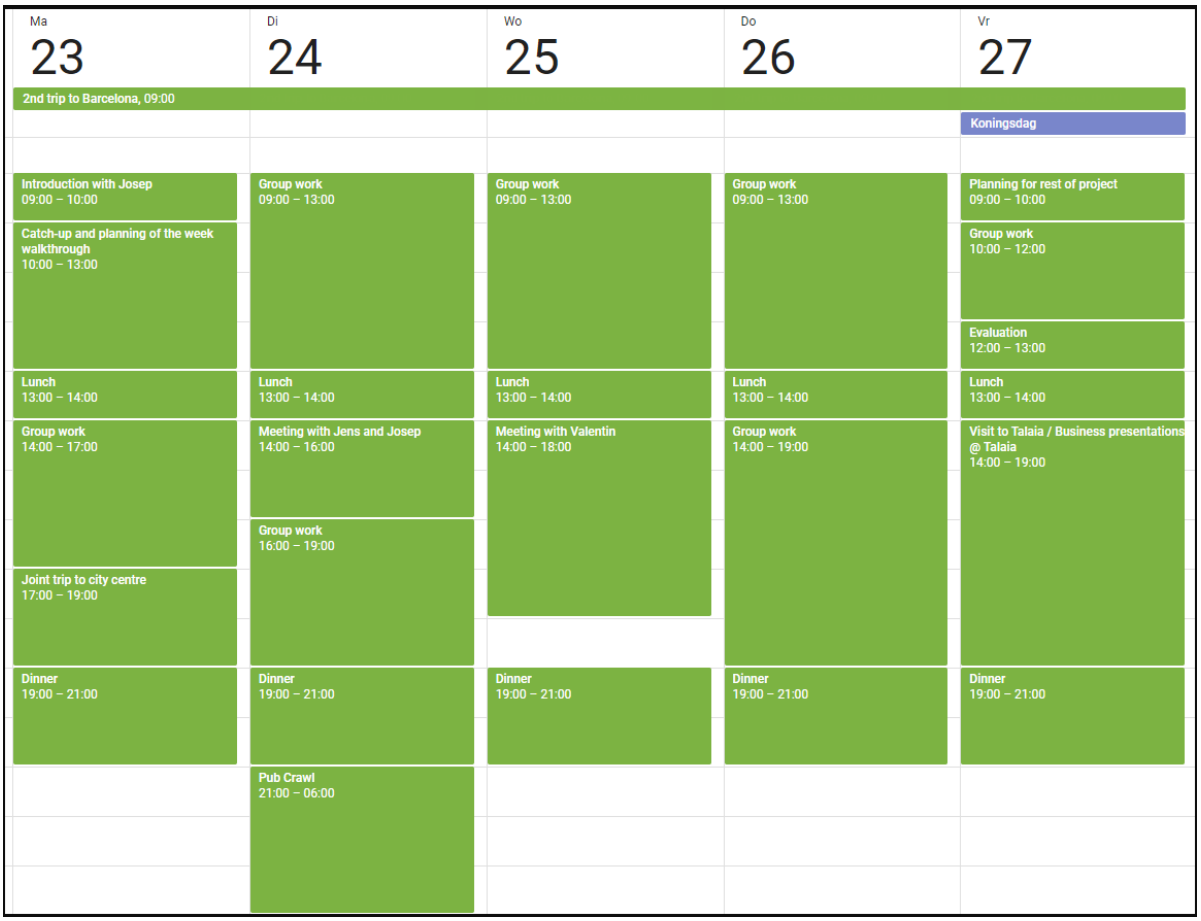
It was later illustrated in the following picture.



These tasks were made taking into account that the group consists of 4 different universities. So while the HoneyJar is a collaborative effort to make, it is clearly split up into manageable pieces, making it more efficient to work with.

4.2 BARCELONA

Having spent more then the majority of the time available for the project, EPIC had a second seminar planned this time in Barcelona. The biggest change from the Riga seminar was not the weather, but the fact that we had to make the time schedule for the trip on our own, with the only requirement being 8 hours of work a day. After a brainstorm and a short discussion, a plan was created.



The reason for the plan ending out like it did was mostley based on experience from Riga. In Riga the group work times were planned to work for small groups of 3-5 people, but being a bigger group means it takes longer to make decisions and as a result there was not enough time to make it work properly for a big group. This problem was not present in Barcelona as there were no courses to take the time off groupwork.

The main reason for the seminar being in Barcelona was because the company Talaia Network's headquarter are located there. Talaia had a representative present at the EPIC seminar in Riga where he functioned as a sparring partner and supervisor.

4.3 TALAIA

The meeting with Talaia networks did not just bring great insight and knowledge to the technical part of the project, but also to the business side of things. Talaia being a company that is already successful within the cypher security industry had a lot to bring to the table no matter what they were asked. It was through their input that the final business model were created and also the final version of our networks analysis system.

Overall was Barcelona a great experience overall with plenty of input to call it not only a fun but also productive trip that really contributed to the EPIC HoneyJar end product.

5 RESULTS

Having reached the end of the project these are the results produced by the EPIC HoneyJar group.

5.1 FINAL JOINT RESULT

- A Business Model for the Honeyjar
  The Saxion students made a Business Model for the sellability of the system, through desk research and field research.
- A basic Honeyjar system
  The basic groundwork for a Honeyjar system has been set up by the Danish group.
- A working Machine Learning Algorithm
  The student from Turkey has developed a Machine Learning algorithm to enhance the Honeyjar.

6 EVALUATION

Reflecting on work when it is done is important to make sure it has been a learning experience. So, this chapter will be a reflection on the overall picture of the project. Every university has their own subsection to write in, so they can express their individual experience.

6.1 COLLABORATION

- Shared
  The collaboration process consisted mainly of weekly meetings but also 2 seminars. For the weekly meetings they mainly functioned as an update to all the groups on what progress had been made in by each participating member. The other use that came out of these meetings was arrangement of subsequent meetings where the actual cooperative work would be done.

- AAU
  Interne at AAU we would be working mostly in groups of 2 where we would work as sparring partners with one another. We think that was a good way of working as you were less likely to get stuck on a problem do to having a qualified person to consult with. For the international cooperation the only regular meeting time was every Wednesday where a structured meeting would take place. But we could always just contact the others over discord if we had urgent matters.

- SAX
  The collaboration part was essential to this project, and we are happy to say that communication went flawless for the most part. All group members were proactive, reachable, and respectful when communicating – in person as well as through the virtual meetings.

- AGU
  There is no doubt that collaboration between group members was pretty good even though group members are from different countries, we managed to keep in touch all times from the beginning of the project by implementing regular meetings on every Wednesdays. The meetings created responsibilities on each group member which pushed us further.

6.2 PLANNING

- Shared
  At the start of the project in Riga, a planning covering the entire scope of the project was made. In the first few weeks this planning was followed accurately, though as time progressed the group realized it had to be more agile and set new goals on-the-go. This didn't turn out to be any problem as the group anticipated that it had to be agile. Having to be agile was a good learning point for the group.

- AAU
  At AAU we made use of a physical timeplan on our blackboard that would be based on the gantt chart we created online to keep track of tings more specific. We later in the project started using Trello which is just a simpler way to keep track of tasks. We started using it

because we found our self not checking up on the gantt chart as it took a long time to reschedule thing in gantt.

- SAX
  At first we were following the set up planning religiously, only to find out that during the course of the project we had to be agile. So in the end we adjusted a lot.

- AGU
  Plans made according to meetings on Wednesdays, so the plans assigned according to work load of responsible person on a part of the project. Those plans changed time to time when unexpected events happened such as emergency, exam and conference times.

6.3 CHALLENGES

- Shared
  The major challenge in the group during the project was the constant rescoping of the project. This is partly because the assignment allowed for free interpretation, as well as the HoneyJar group consisting of a lot of members – each with their own opinions and visions. Another very influential factor was when one project member that didn't fit in the group. Dealing with this member's lacking of her results depleted a lot of the group's time.

- AAU
  Being as big of a group as we are at AAU, a lot of ideas on how to improve the project comes pretty often, however we cant simply change things

- SAX
  Finding common ground on what to make was a challenge at times.

- AGU
  There were several challenges for this project, the main one is luck of time, indeed there should be dedicated time period to be able to work on preparing datasets, algorithms and implementations on data. However, we tried to manage all our courses besides this project at the same time which caused some collisions and undesired delays on implementation on machine learning part. Another challenge was luck of labelled datasets on malicious activities in network, we tried to prepare them, yet bureaucratic issues raised which also created unwanted delays. Furthermore, training process of a machine algorithms requires good datasets, which we do not have yet, to decrease false-positives. However, we managed to extract features to label data and we communicated with Canadian Institute for Cybersecurity (CIC) for datasets which they may have on malicious activities. Now, received network data in PCAP format is under investigation and creating CSV files for machine learning algorithms.

6.4 HOW DID WE OVERCOME THEM?

- Shared
  We compromised and downscaled the project in order to compensate for the lost manpower. Keeping something realistic within sight was important for us as we wanted to have at least a part of a product in the end, not just theory.
- AAU

By keeping the communication between the universities consistent we were able to rescoping the project, so it felt like it was manageable again.

- SAX
  This challenge was overcome by giving this person her own assignment, outside of the main group. Meeting in Barcelona for a second time was very fruitful as it allowed the group to refocus and plan the ending of the project.

- AGU
  Existing datasets and articles on malicious network data, which are collected from virtualized environment, investigated and significant features which specify malicious activities extracted, machine learning algorithms such as Support Vector Machine, One-class SVM, perceptron and those algorithms are investigated according to Canadian Institute for Cybersecurity (CIC) datasets , to be able to understand which algorithms used in which case.

6.5 What did we learn from that?

- AAU
  Having to work with people from other universities is a fun but challenging thing to do. It requires significantly better coordination and planning as the people simply are less assessable when they are so far away. There are also more things to consider, like wanting to make a change in the project. However, before that can happen you need to receive an ok from all other participants as, they have as much to say about the project as anyone else.

- SAX
  Communicating more efficiently, and making compromises.

- AGU
  People who are from different cultures, experiments, areas, countries and universities may have success if they have enough time on a purpose. It would create some problems in time however, those problems can be overcame by power of collaboration. Indeed, having such a diversified people on a project creates different perspectives on a problem.

6.6 How did EPIC help the process?

- Shared
  The facilitations EPIC made in Riga and Barcelona were essential to the proceedings of this project, as well as the continuous support from EPIC-delegate J.M. Pedersen. It's safe to assume that if the group never met in person, the end results would be very different.

- AAU
  By having teamwork exercises from day 1 and having good supervisors to help us get a clear plan of what we would make and how to do so.

- SAX
  Constantly supervising and giving feedback. The seminars really helped out to meet the group members in person.

- AGU

  We faced an opportunity to meet with experienced people in field of cyber security and machine learning from industry and different universities to get responses from their perspective about the project. It brought experiments which cannot be gained in a university education system, furthermore this project taught that how research and business sides of a project works.

6.7 RECOMMENDATIONS

- Shared
  What we'd recommend for future EPIC collaborations/projects/seminars is to use the material that was made by previous EPIC iterations for further development, i.e. our HoneyJar-project. We think that all project groups established a basic product during this year, and it would be interesting to see the further development of these products.

  Other than that, our findings are that EPIC should be more assertive in dropping non-contributing members from project groups once a certain point is reached, as they only slow project work down.

- AAU
  Be prepared to spent some extra time discussing pretty much everything that requires all the members to be decided. Make sure to stay active on your chosen communication platform, and don't just got there for working purposes. Having a good relationship with the other people on the collaboration helps not just avoid fights, but also makes the process more enjoyable as a whole.

- SAX
  Make sure all project members are contributing and if not apply appropriate measures. Also we recommend using previously made materials during the EPIC seminars.

- AGU
  To be able to make fast communication in technical part of a project, all group members should have some solid background on dedicated project topic, otherwise collisions might raise up during project collaboration and further development. Indeed, even though group members have solid background, they might choose to do not work, then those people should be notified several times if they do not consider notifications then all relations of those members should be removed from group and related course which is taken locally for each member in the group.