Interview transcript /w Valentin Carela 25.4.18

**Attendees:**

Valentin Carela

Jens Pedersen

Ahmet Turkmen

Morcel el Ouahbi

Robert Nedergaard

**Security**

1. How big of an issue has cybersecurity become in the last couple of years? How do you notice that the most?

Nowadays security is basic. Before it was only basic for big companies. Nowadays even the smaller companies have noticed it’s crucial for their services. Now the governments and the states are looking after the data of the company and users. It was a hot topic, but nowadays it’s even more important. A single attack was able to destroy big companies at once, like amazon, twitter, facebook.

Before someone attacked Playstation, irrelevant services, but nowadays also governments etc.

1. What type of malware is currently in your experience the most popular?

I would say there are 2 different branches. DDOS. And the other one is related to the exploit data from users and/or companies. We’ve seen cases that focus on a specific company. It was a tool designed specifically to hack that company. I think those are the most common ones. Denial of Service is a problem because everyone is aware of it, but the data that leaks malware is sneaky. It goes unnoticed. It moves very fast. So I would say that the most popular one is Spyware. However it’s not the most dangerous one. I’m going to give you some categories of malware. On the one side you have viruses that you can combine with software, i.e. a torrent client.

Another category is worms. It’s a piece of software that goes over the network, infecting other machines on the same network. You can steal any data, credentials. Another category would be Trojan Horse that is installed manually by the user. This enables you to reach a computer outside your network, like a virtual machine with windows but actually with a real computer. This goes unnoticed.

Then you have spyware, that enables addons to other software, what they do is extract information from your devices and share it with someone else, again credentials, and data. Now we have Grey worms, its technically not malware, for example Google and Facebook, they ask you for permission to use your data. The difference between this and spyware is that they give you a better ‘’service’’ like ads, attuned to the user. Also botnets. Ransomware is probably related to worms, but might be it’s own category really.

1. How do you make sure you keep up to date in an environment that moves very quickly.

Usually for companies it’s being one step behind, because we’re preventing. Sometimes we don’t say as we try to stay on the same level. What we do is usually do research about what the universities are doing, what’s relevant in conferences, always up to date, always implementing new proposals that are state of the art.

In talaia’s case we try to be agile, we can modify things fast and it’s easy for us to be on the edge. Sometimes companies willingly don’t keep up because of finances.

SDM software defined networks. Separate the routers and switches, instead you have one machine that forwards packets, and then there’s a brain that manages all these packages. What happened here, was that they already spent all their money building super expensive machines. If you do a Honeyjar like this you want to keep hardware the same, the system should be modular. Better even, use software. Much easier to manage and implement. Plug and play. So instead of paying thousands of euro’s for machines in the network use software, this can be implemented anywhere. Hardware is unneccesary in this case. It’s a complexity you want to avoid.

1. What makes a Honeypot attractive?

Make it easily deployable, isolated. I.e. there are solutions like enefzen, that cost 2 weeks only for installing the tool. It doesn’t always work. Very unreliable. It would also be useful if you could introduce some malware in the honeypot. Something like a demo. Imagine that you can install different VM’s with malware inside, you can try any combinations. This makes it very attractive. One of the key points is to keep it simple. Other functionalities that should be in the honeypot is to check what have been the most important malware attacks in the last 3 years for android, and ID them. Install these in the specific VM’s. This is actual. “Remember this recent big attack?” This triggers the company.

**Android**

1. How embedded have mobile phones become in your company?

In our case it’s not important. Because we focus in a completely different area. Our target is usually ISP’s and company and we focus on the monitoring of them. We don’t focus on differentiating between phone and computer traffic. However, we use our phone for operational things, like SLACK, Whatsapp, Email. I have a console with a VPN to connect with the machine in Talaia. It’s not comfortable but it’s useful for emergencies. I don’t recommend to use your cellphone for work. I could do my work without my phone. Nowadays I don’t trust anyone so I try to keep as little personal information on my phone.

1. How big of a risk have mobile phones become concerning cybersecurity, in your experience.

I’m cynical. Because people don’t realize how much information is stored in their phone. For my mother that’s receiving a Whatsapp message with like a discount code, having her log in to facebook, allowing permissions. It happens very quickly. For example Facebook was finding out what the best place for a political meeting was. Depending on the information. Facebook takes advantage of this, and as well as fake news. News presented in a very specific way. There are a lot of subliminal controlling factors in facebook.

1. How much manpower, realistically, would be needed to operate a Honeyjar system like ours? (*compare with talaia here)*

It depends. If you just focus on malware just 1 guy is fine. If you allow the user to install whatever they want so they can evaluate, it can be useful for others too. Maybe 2 or 3. Companies are not employing people for this specific topic. Maybe in a company that develops games for Mobile phones.

**Interface**

1. Is there anything you’d like to advice us, having had experience in Security for a long time.
2. What is the most valuable information for CEO’s to gather from a Honeypot? (*For example in exported docs)*
3. What information would you like to have visible on the main screen of your Honeyjar system?
4. What type of interface design would you prefer? *(Operating system, Interface structure, Most important functionality, Things you definitely don’t want to have included in the program)*
5. What do you think is the best way to generate the most up-to-date data on Malware and why do you think so?
6. If you were a customer of our product, what would you want to see on the dashboard? What information is so important that you need to see it quickly?

* What would change for someone working with the system on a daily basis?
* What would change for someone like a CEO?
* What would change for someone like an IT manger?

**Target audience**

1. If we take your application, what type of customer do you usually have? Is it big businesses or more small businesses?

**Network Security**

Q about research:

First I’m not an expert, but I’m going to explain you what we know. So you were talking before about using deep learning. The good thing is that when there is a Machine Learning problem, try to extract all the features that you and from the data that you have, even if it’s not important for you. The Machine learning model will tell what information is relevant to the ML system. Don’t try to understand be relation between the distribution size of the malware because that’s going to be tough. See it is a black box, input and output. Rest is not relevant.

What is important is to try to insert features that are not dependant of the rest. i.e. Fashion and female would be dependant probably. With the data you have show a list of all the features and apply models for output. Compare these.

**Q about Labeled data:** There are 2 approaches. Unsupervised ML and supervised. In our case what would work best is supervised ML. Honeyjar is going to provide us data. The supervised ML will collect data from the internet, you put it in the model and the model creates some clusters. Manually inspect these clusters. If you see common parameters you can tell where its coming. Try to find similar parameters in the non-common data.

**Q about encrypted traffic**: Should we consider encrypted traffic or not?

The trend is that everything is moving to encrypted, but it’s more difficult to detect. Maybe not worth to do.

**Q about Android:** What is the point of Android in this case? It’s the same as Android. What’s the reason behind Android phones? Depends on what you want and what you can do. The start is the Honeyjar. Android is not as common as other systems. It’s a hot topic so it’s a good environment to test. This makes work easier, cos it focuses on 1 point.

Cloning Joy on local server is a good idea. Importing multiple PCAP files, if something is wrong it will make an analysis of that part. Taking those and modifying them and combining them in an effeicient way, is this a good idea? (Straffor)

1. What is the relationship between byte distribution and packets' behaviour?
2. What is the relationship between sequential packet length and packets' behaviour?
3. What’s the number of flows?
4. If there is no real labeled and unlabeled data then should we consider deep learning on the packets that we are working on it.
5. Should we consider encrypted traffic directly or not ?
6. Do you think that we are at right direction by checking http requests and getting some responses from API? (-there will be 15 sec delay and which is huge delay -)
7. Is it ok to use frameworks and modify some functions of existing open source tools according to our needs ?
8. The extracted features are related with DdoS attacks and we want to identify malicious activity but the features are NOT so much related with the identfying malicious activity.

**Feedback on GUI**

**Danish guys:**

Server security. All the same VM’s are going to show the exact same thing because they’re all on the same server.

**Saxion:**

Maybe you can present data that is related. Detection model is made in Machine Learning. I don’t think it’s the right approach. You cannot differentiate traffic inside the honeypot. This is a second step. First you need to get the data and use it for learning process. If you use ML to ID what is malicious and what is not.

Should monitor HDD, network. What are the processes that are running, what’s being used. Once you enable malware and differences appear in the logs you can differentiate the malware. Focus on the process.

Geographical location/Type of company. An idea is to use information gathered from customers. Or we can directly sell it. Subscription based for example. But then you have to get the permission from the users. Giving them access to all the data from the customers is valuable. Give a price for the deployment of the Honeyjar and allow the customers to join using subscriptions.

One of the challenges is good accuracy with Machine Learning. We’re getting too many False positives, this is a drawback – and it’s specific to the company. We’re closing the gap the main channels of the seminar project today. A system can be trained for a specific kind of company to deal with false positives.