Poster: Obfuscation Revealed - Using Electromagnetic Emanation to Identify and Classify Malware

Duy-Phuc Pham, Damien Marion, Annelie Heuser

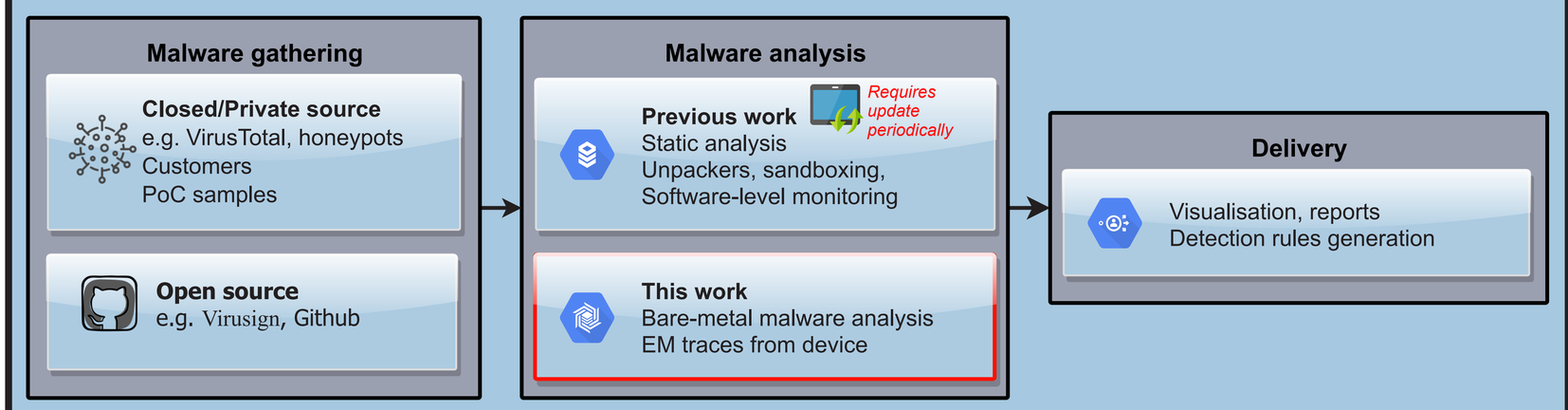
Poster: Obfuscation Revealed - Using Electromagnetic Emanation to Identify and Classify Malware

Duy-Phuc Pham[1], Damien Marion[1], Annelie Heuser[1]
[1] *Univ Rennes, CNRS, IRISA, France*

1 - Abstract

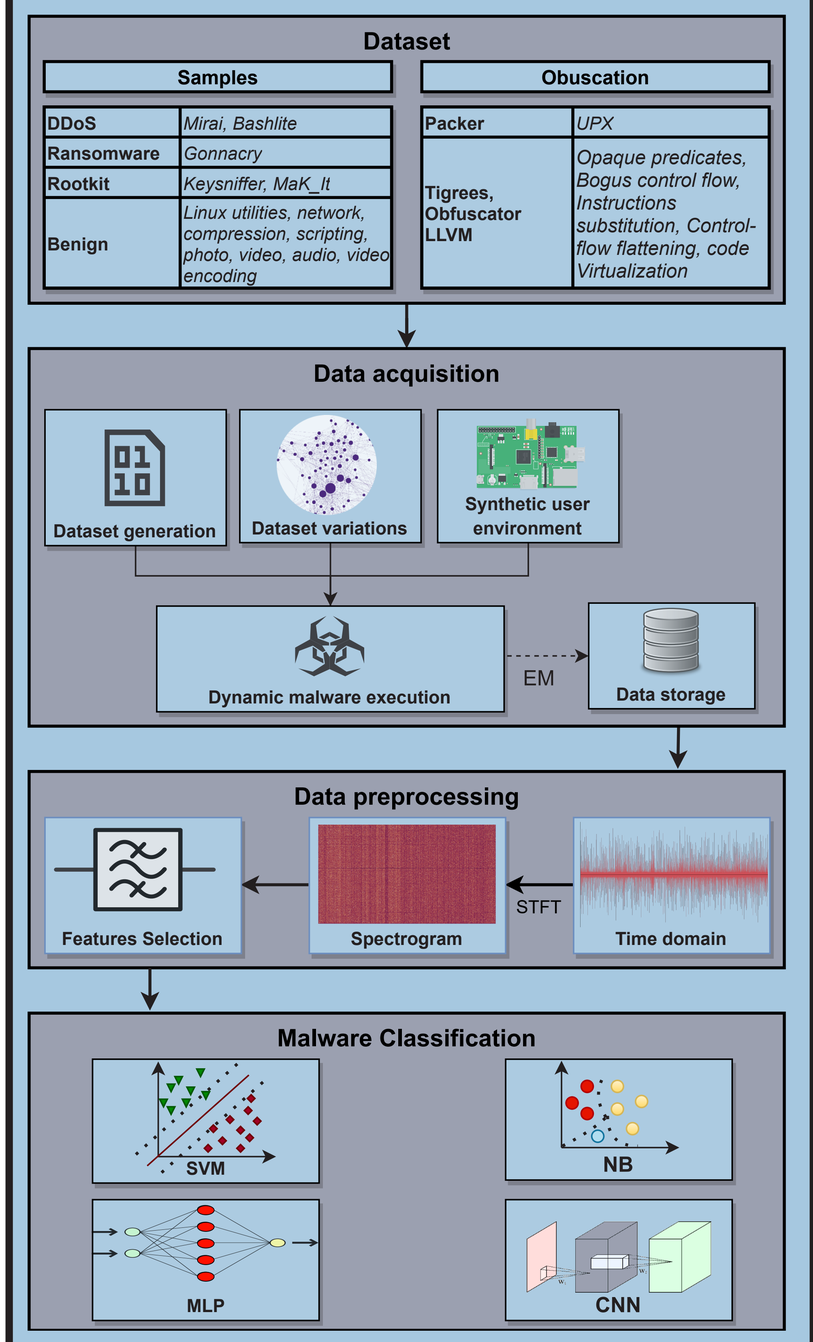
In this poster we present a novel approach of using side channel information to identify the kinds of malware threats that are targeting IoT devices. Although in the presence of obfuscation techniques that can prevent static or symbolic binary analysis, a malware researcher may obtain detailed information about malware type and identification using our method. It operates by leveraging side channel by electromagnetism rather than software-layer malware analysis. By capturing 96,000 measurement traces from an IoT system infected with different malware samples, we can obtain this information without altering the actual hardware. As a result, it can be implemented without any overhead, regardless of the resources available. Furthermore, our method has the advantage of non-trivial for malware writers to avoid. By collecting EM traces and later analyzing them for patterns, we were able to extract information from IoT devices. We were able to distinguish malware families based on side-channel knowledge without being able to see what exact hardware was involved. We were able to predict three generic malware forms (and one benign class) with a 99.89% percent accuracy in our tests. Even more, our results show that we are able to classify altered malware samples with unseen obfuscation techniques during the training phase, and to determine what kind of obfuscations were applied to the binary, which makes our approach particularly useful for malware analysts. **keywords:** ***Malware analyst, side-channel analysis, obfuscated malware classification, electromagnetic leakage, deep-learning, machine-learning.***

2 - Malware analysis model

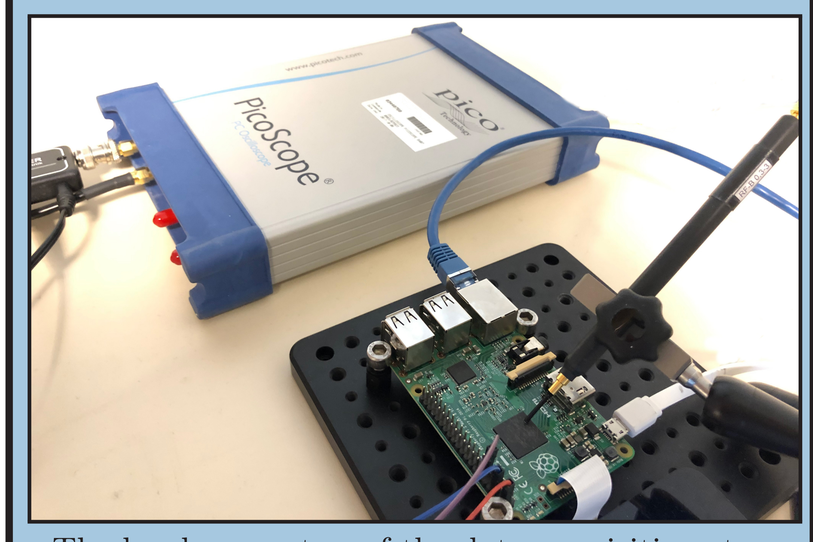


Traditional malware analysis workflow and our approach.

3 - Proposed EM framework



Proposed malware classification framework.



The hardware setup of the data acquisition step consists of a H-Field probe placed 45°above the processor connected to the Picoscope.

4 - Machine learning & Deep learning classification

Convolutional network (CNN), multiplayer perceptron (MLP), support vector machine (SVM), and Naive Bayer (NB) are used for malware classification with linear dimensionality reduction (LDA) to preprocess.

		MLP			CNN			LDA + NB			LDA + SVM		
Scenarios	#	**AC**	**RC**	**PR**	**AC**	**RC**	**PR**	**AC**	**RC**	**PR**	**AC**	**RC**	**PR**
Type	4	99.53	99.71	99.73	**99.89**	99.94	99.93	97.77	98.65	98.49	97.84	98.65	98.64
Family	6	97.73	88.57	88.54	**99.32**	96.71	96.68	96.37	87.16	87.08	96.53	87.52	87.52
Virtualization	2	94.82	94.19	94.70	**96.20**	96.42	95.61	89.63	89.00	88.89	89.45	88.65	88.78
Packer	2	92.50	92.26	92.25	**92.63**	92.80	95.60	78.87	78.40	78.18	79.07	78.51	78.38
Obfuscation	7	71.18	70.09	70.15	**81.62**	81.01	80.97	60.71	59.5	59.56	61.06	60.44	60.58
Executable	35	72.76	73.04	74.08	**81.73**	82.73	82.84	69.55	70.84	70.58	69.85	70.33	70.37
Novelty (family)	5	92.42	94.79	94.32	**99.38**	99.57	99.43	98.90	99.09	98.96	98.86	99.02	99.03

The accuracy (AC), recall (RC) and precision (PR) obtained with MLP, CNN, LDA + NB and LDA + SVM applied on several scenarios. Bold numbers indicate the highest accuracy (%) on the testing set per scenario.

Type classification

This scenario gives us a 4-classes classification problem: ransomware, rootkit, DDoS and benign. All of the models are very efficient for this problem ($> 97\%$ accuracy). We can observe that CNN is slightly more accurate than MLP, NB, and SVM.

Family classification

In this scenario we classify into the malware family plus benign class, which gives 6 classes: *bashlite*, *mirai*, *gonnacry*, *keysniffer*, *maK_it*, and benign. CNN gives the highest accuracy with 99.32%, but also MLP and ML provide results $> 96\%$.

Executable classification

This is a straightforward executable recognition situation in which the model is attempting to profile the binary exactly. It translates to a 35-class classification problem (including benign samples). We distinguish the family, variants, and obfuscation where CNN is the most efficient at 81.73 % (vs a random guess of only 3.23%).

Virtualization & packer identification

We evaluate whether the binary is covered with virtualization or packed which results in two-class detection problems. We see that virtualization is slightly easier to detect than packing, with CNN having the best accuracy (96.20% vs. 92.63%)

Obfuscation classification

CNN is more efficient in classifying between obfuscation techniques, it achieves 81.62% (vs a random guess of 14.29%). This result indicates again that our methodology is able to extract and compare malicious behavioral features.

Novelty (family)

It is very common in real-world malware detection to come across unknown variants. Even if the models are projecting unknown (obfuscated) variants, CNN reach 99.38%. LDA+NB and LDA+SVM also perform very well ($> 98\%$).

5 - Conclusion

- Our method involves no changes to the target system, with no software tracking, exact triggering, or extra device overhead. To be applicable to practical IoT systems in the wild, we use real world malware and multi-processor hardware running a functional Linux OS in our tests.
- Our approach consists of preprocessing the measured electromagnetic emanation of the device by selecting the most relevant features (frequency bands) over time, followed by classification using neural network and machine learning models.
- We put together a number of scenarios, each of which represents a real-world malware detection: malware classification by type and family, exact malware executable profiling, virtualization and packer discovery, obfuscation classification, and classification of unseen obfuscated variants.