A Framework for Detecting Mobile Ransomware

Amirhossein Gharib, Arash Habibi Lashkari, and Ali A. Ghorbani

Canadian Institute for Cybersecurity (UNB), University of New Brunswick (UNB)

Abstract

Ransomware has become one of the main cyber-threats for mobile platforms and in particular for Android. The number of ransomware attacks are increasing exponentially, while even state of art approaches terribly fail to safeguard mobile devices. In this paper, DNA-Droid, a two layer detection framework is proposed. It benefits of a dynamic analysis layer as a complementary layer on top of a static analysis layer. The DNA-Droid utilizes novel features and deep neural network to achieve a set of features with high discriminative power between ransomware and benign samples. Moreover, Sequence Alignment techniques are employed to profile ransomware families. This helps in detecting ransomware activity in early stages before the infection happens. The DNA-Droid is tested against thousands of samples. The experimental results shows high precision and recall in detecting even unknown ransomware samples, while keeping the false negative rate below 1.5%.

Ransomware?

What is it?

A type of malicious software designed to block access to a computer system until a sum of money is paid.

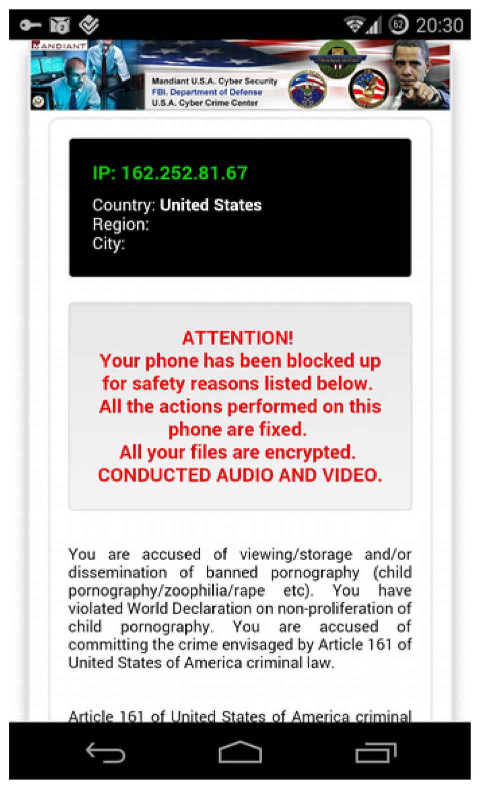


Fig 1. An infected Android device

Is it serious?

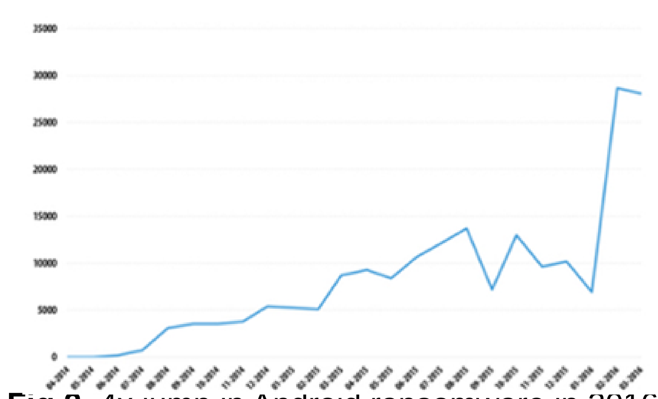


Fig 2. 4x jump in Android ransomware in 2016 and still growing.

Why?

- Android popularity and lack of defense.
- Easy to use crypto-currencies such as Bitcoin.
- Easy to use Android crypto libraries.

System Architecture

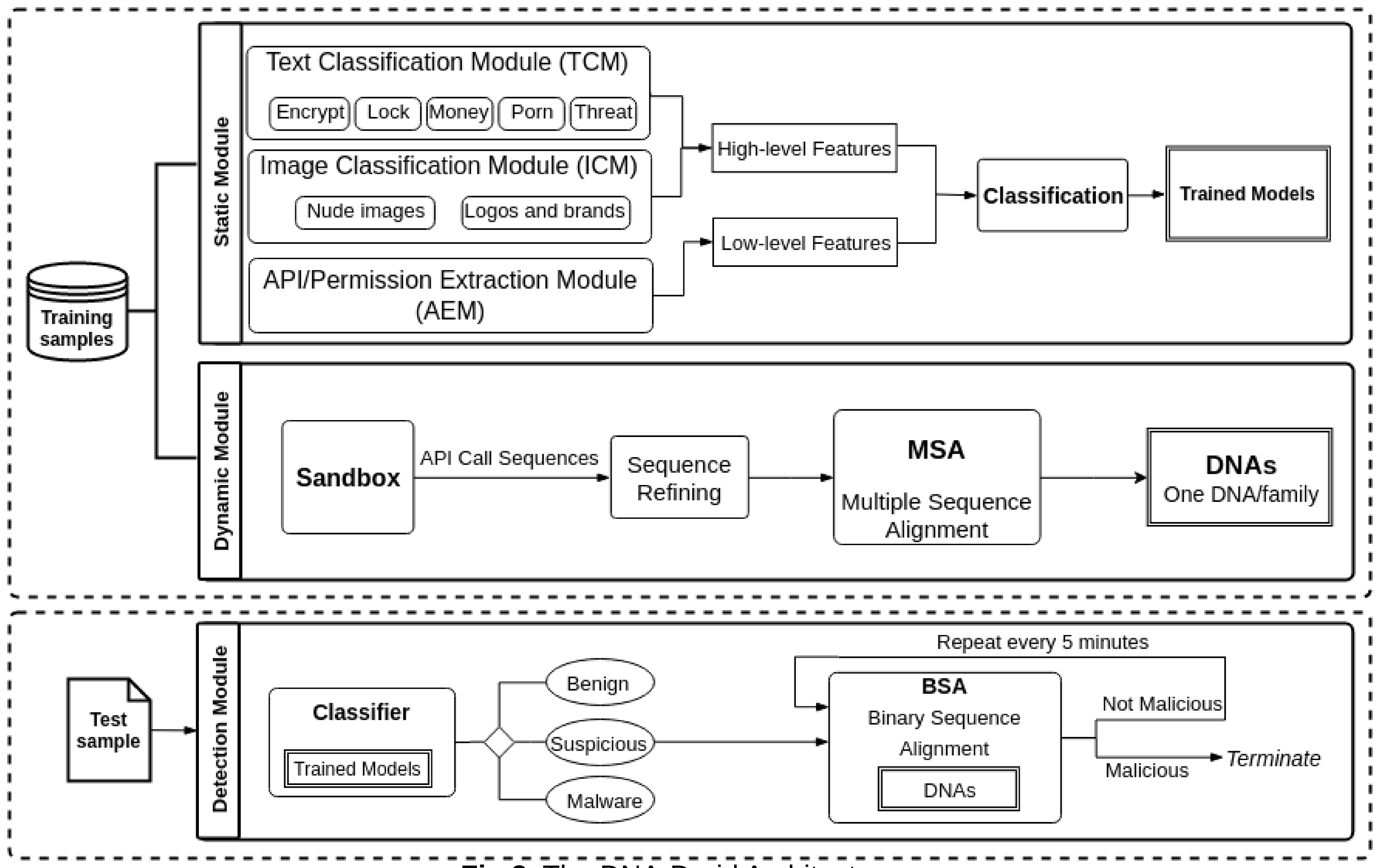


Fig 3. The DNA-Droid Architecture

Feature Learning

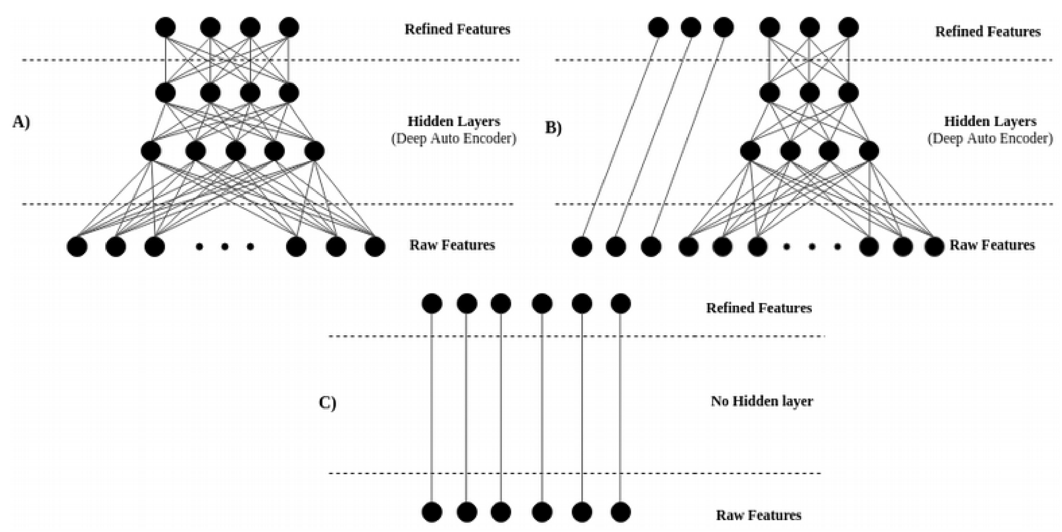


Fig 4. The three design models that are used to reduce and learn new features using Deep Auto Encoder.

$$N_{hiddenNeurons} = \frac{N_{samples}}{\alpha * (N_{inputNeurons} + N_{outputNeurons})}$$

Formula 1. Defines the number of Neurons for a hidden layer. Alpha is a scaling factor indicating model generalization.

How to detect Ransomware before infection?

Is it infected? No backup? Too late…

We propose a two layers detection framework, containing the three major components: **static analysis module**, **dynamic analysis module** and **live detection module**. The detection module first attempts to quickly scan the incoming samples and score them statically. Further analysis is then enabled only for the suspicious samples.

- **Static module** includes three sub-components for evaluating different aspects of an app.
- Three **deep learning designs** are explored to **reduce** and **learn** new features.
- Dynamic behavior is defines as an **API call sequence.**
- Dynamic module profiles malware families based on the API call sequences, and produces a **DNA** for each family.
- In live detection phase, run-time behavior of a suspicious sample is continuously compared with the families' DNA and will be terminated if the sample is matched with a DNA.

Ransomware Visualization

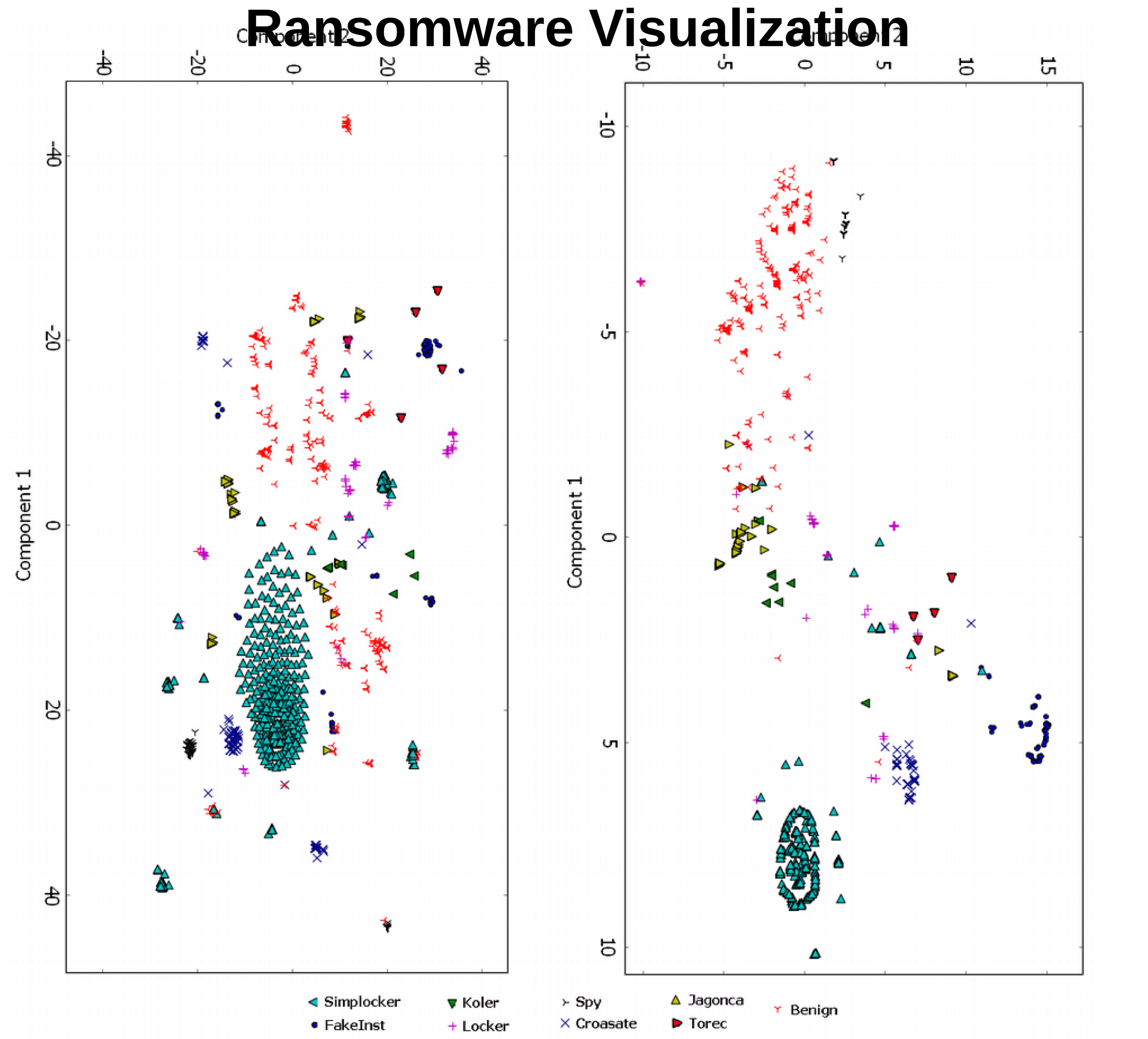


Fig 5. The left image shows the T-SNE transform of raw features with error rate of 0.376, while the right image shows the T-SNE transform of refined features with error rate of 0.084.

Framework Evaluation

Table 1. Static detection performance of the DNA-Droid

Design	Classifier	Accuracy	Precision	Recall	FP
A	*NB*	94.7%	95.5%	94.7%	1.5%
	SVM	96.1%	96.0%	96.1%	3.2%
	RF	96.6%	96.7%	96.6%	1.1%
	AB	68.2%	57.8%	68.2%	13.5%
	DNN	97.1%	98.0%	97.1%	**0.1%**
B	*NB*	93.0%	92.3%	90.0%	3.8%
	SVM	96.2%	96.0%	96.2%	3.6%
	RF	98.0%	97.5%	98.0%	0.7%
	AB	68.9%	58.9%	68.9%	13.1%
	DNN	**98.1%**	**98.1%**	**98.1%**	0.5%
C	*NB*	84.1%	86.2%	84.1%	6.3%
	SVM	88.5%	87.9%	88.5%	6.9%
	RF	90.2%	92.0%	96.5%	4.0%
	AB	65.4%	53.5%	65.4%	14.9%
	DNN	86.6%	85.9%	8.66%	4.1%

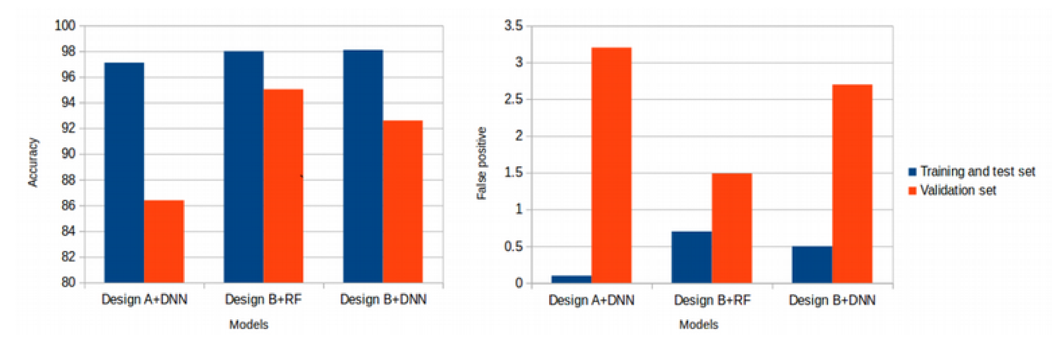
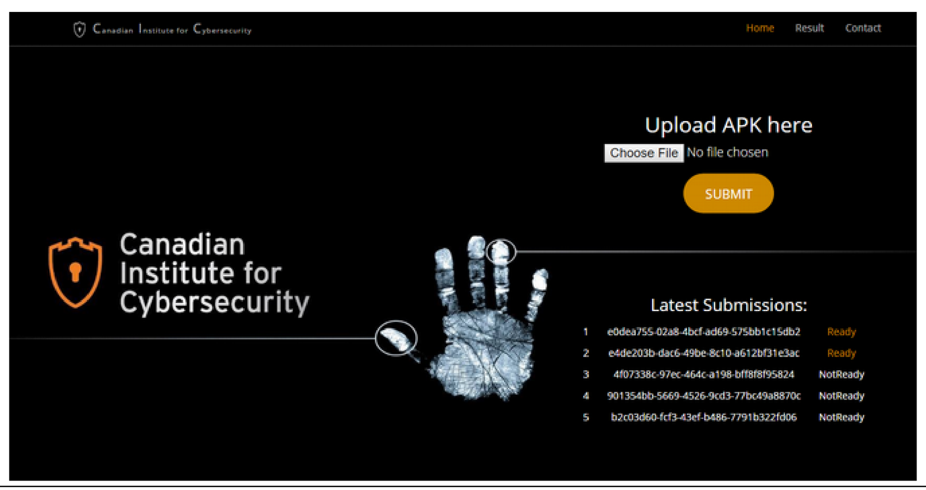


Fig 6. The comparison of the top three models by using the validation set.

Contributions

- ✓ Introduce **novel features** with high discriminative power; making the DNA-Droid capable of **recognizing unknown** ransomware samples.
- ✓ Investigate the performance of **Deep Auto Encoder** to reduce and **learn new features**.
- ✓ Utilize Binary and Multiple Sequence Alignment (**MSA**) techniques to analyze dynamic system call sequences.
- ✓ Release a **publicly available** fully automated Android sandbox that is able to report the **sequence of API calls** as a web service.



Conclusion

The experimental results show that the DNA-Droid is able to discriminate between ransomware and benign samples with high precision and that it **outperforms** state of the art approaches. It shows a high capability to detecting ransomware activity in **early stages** before the infection happens.