CELL-PHONE CLASSIFICATION: A CONVOLUTIONAL NEURAL NETWORK APPROACH EXPLOITING ELECTROMAGNETIC EMANATIONS

Baki Berkay Yilmaz, Elvan Mert Ugurlu and Alenka Zajić

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, 30332 USA

Milos Prvulovic

School of Computer Science
Georgia Institute of Technology
Atlanta, GA, 30332 USA

ABSTRACT

In this paper, we propose a methodology to identify both the brand of a cell-phone, and the status of its camera by exploiting electromagnetic (EM) emanations. The method composes two parts: Feature extraction and Convolutional Neural Netwotk (CNN). We first extract features by averaging magnitudes of short-time Fourier transform (STFT) of the measured EM signal, which helps to reduce input dimension of the neural network, and to filter spurious emissions. The extracted features are fed into the proposed CNN, which contains two convolutional layers (followed by max-pooling layers), and four fully-connected layers. Finally, we provide experimental results which exhibit more than 99% classification accuracy for the test signals.

***Index Terms*—** Security, Classification, Convolutional Neural Network, Electromagnetic Emanations

1. INTRODUCTION

A side channel is an unintentional source of information which can leak confidential data [1]. Many attacks have been performed to exfiltrate sensitive information by exploiting the emanations. These attacks, called side channel attacks, are established on systematical changes while executing a script, a program, etc. Some examples of these attacks exploit distinct features while signing different bits of a cryptosystem, and are based on temperature [2, 3], timing [4, 5, 6], cache-misses [7, 8], acoustic signals [9], and power consumption [10, 11]. These attacks either require direct access to the targeted device or have a limited bandwidth. However, attacks based on EM emanations require only close proximity [12]. Moreover, they can take advantage of larger bandwidth which can result in higher throughput [13, 14].

The potential of the EM based side channel attacks is demonstrated by Eck [15] and Kuhn [16] when they reconstruct images on a video display unit by capturing the emanated EM signals from some distance. Although these advancements are made by experimenting on relatively simple devices, they have motivated researchers to monitor different units of more sophisticated devices. Detecting the current status of a camera could be a very good example for monitoring relatively more complex devices because many public places enforce no camera policy, i.e. museums, cinemas, theaters, etc. However, easy access to cameras through smartphones make the policy enforcement much more difficult. To address this issue, a supervised learning model which utilizes k-Nearest-Neighbor algorithm is proposed in [18] to classify phone status only when its camera is active. In this paper, we propose a methodology to classify the phone brand and status of a camera even when the camera is inactive. The method contains two modules: Feature extraction and CNN model. We first extract features by averaging STFT outputs of the measured signal, and then utilize these features in the proposed CNN model. We demonstrate that classification for the test measurements exceeds 99% accuracy rate.

The organization of the paper is as follows: In Section 2, we explain the feature extraction method and introduce the CNN model, Section 3 provides the experimental results and discussion, and conclusions are drawn in Section 4.

2. METHODOLOGY FOR PHONE BRAND AND/OR CAMERA STATUS IDENTIFICATION

As the mobile devices get more complex, exfiltrating information becomes more challenging because of higher operating clock frequency, coupling of emanated signals from various components, etc. It has been shown that there are many sources in a smartphone which can leak information [17]. In this respect, we investigate different patterns in the frequency domain when the status of a camera changes. An example of the received signal is given in Fig. 1 for the cases when rear-camera is on, front-camera is on, and both cameras are off. One of the main observations is that some frequency components are activated when the cameras are active. Moreover, the activated frequencies are different for rear and front cameras. However, the last observations could be misleading

This work has been supported, in part, by NSF grant 156399, and DARPA LADS contract FA8650-16-C-7620. The views and findings in this paper are those of the authors and do not necessarily reflect the views of NSF, and DARPA.

since it is shown that the same frequency components are activated for both rear and front camera activities in some devices [18]. The first approach that comes into mind is to track the frequencies which are activated when cameras are on. This approach can be inefficient as the sample size increases, and can cause loss of buried information existing in the measured signal. Moreover, it is almost impossible to differentiate distinct phones when both cameras are off.

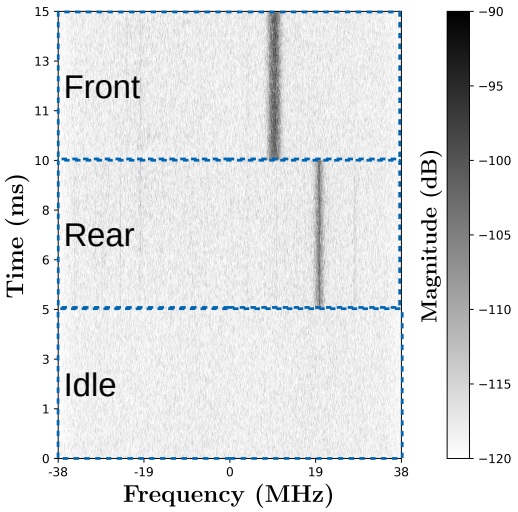


Fig. 1. Spectrogram of the received signal when the camera is idle, rear camera is active and front-camera is active for ZTE.

To obviate the difficulties, we first need to extract features that can be utilized to cluster existing classes in the measured data. However, measurements that are obtained by measuring devices with high sampling rates are generally very large, and contain many frequency components which are not relevant to camera activity. To decrease the dimension of the input signal and weaken the undesired frequency components due to other sources, we apply STFT averaging introduced in [18]. Let T_M and T_S be the measurement and sampling time of the measuring device, respectively. Therefore, the number of samples taken for each measurement can be written as $I_S = T_M/T_S$. Assuming O_S is the number of non-overlapping samples between consecutive STFT operations, the total number of STFT operation can be written as $\Xi = \texttt{floor}\left(\left(I_S - \mathcal{F}\right)/O_S + 1\right)$ where $\mathcal{F}$ is the FFT size for the STFT window. Therefore, the input vector $\mathbf{m}_i$ for the i^{th} measurement after feature extraction can be written as

$$\mathbf{m}_i[k] = \sum_{n=1}^{\Xi} \left| X_n^i[k] \right| \tag{1}$$

where $k \in \{0, 1, \cdots, \mathcal{F}-1\}$, and

$$X_n^i[k] = \sum_{\xi=0}^{\mathcal{F}-1} \mathbf{\Theta}_i\left[\xi + (n-1)O_S\right] \exp\left(-j2\pi k\xi/\mathcal{F}\right) \tag{2}$$

where $\mathbf{\Theta}_i$ is the measured raw signal. After extracting the features, our next goal is to classify the signals as accurate as possible. In this respect, we propose to utilize a CNN model given in Fig. 2.

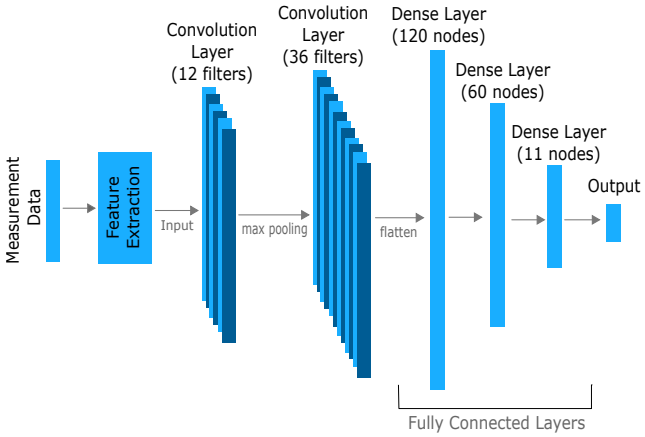


Fig. 2. CNN model for classification of camera status and phone brand.

The model contains two convolutional layers and four fully connected layers. Each convolutional layer is followed by a max-pooling layer. The kernel size of the max-pooling layer is set to 10 with stride of 10. The stride for the convolutional layers is kept as one while the kernel size is 10. The input size for the CNN model is equivalent to $\mathcal{F}$ because $\mathbf{m}_i$ is the vector containing the extracted features from the measurements. Dense layers are followed by a ReLU layer except the output layer, where we apply softmax function. The size of the output layer is set based on the number of considered clusters.

3. EXPERIMENTAL RESULTS

The last step is to validate whether the feature extraction method and the proposed CNN model can cluster existing classes accurately. For that we first introduce the experimental setup given in Fig. 3. We use a signal analyzer (Keysight

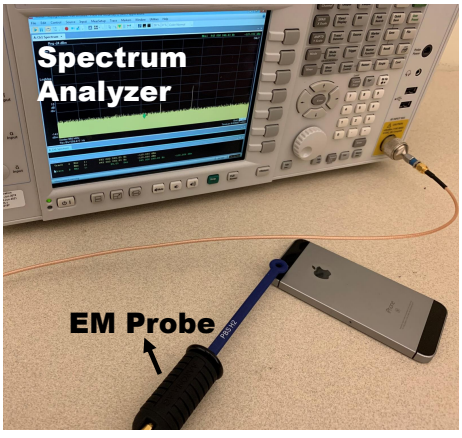


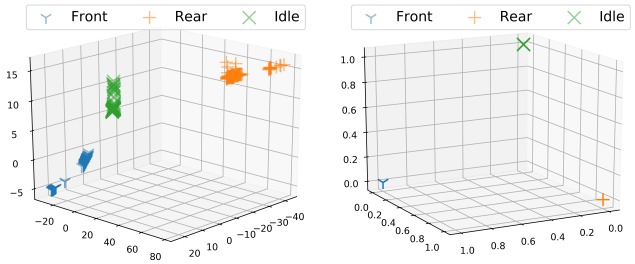
Fig. 3. Experimental setup for measurements.

MXA N9020B), and a near-field magnetic probe (AAronia PBS H2) which is placed on top of the rear camera. The considered phones are ZTE ZFive with 1.4 GHz clock frequency (Quad-Core), Alcatel Ideal with 1.1 GHz clock frequency (Quad-Core), iPhone SE with 1.85 GHz clock frequency (Quad-Core), and Samsung Centura with 800 MHz clock frequency (Single-Core). The received signal is downconverted by 990 MHz and the bandwidth is set to 30 MHz.

Before demonstrating the experimental results, we provide the steps that we follow to cluster the brand and the camera status of a phone:

- Collect the training signal for T_M seconds for each phone and camera status with the experimental setup given in Fig. 3. Set the sampling rate of device, T_S, and the considered bandwidth to be the same for all measurement to prevent any inconsistency.
- Determine the `FFT` size and apply the feature extraction method described in Section 2.
- Modify the output layer of the CNN model so that the number of nodes is equivalent to the number of distinct training classes.
- Train the CNN model by utilizing softmax as the loss function of the model.
- Collect new signals to test the trained model.

We first investigate whether differentiating the status of the camera of a given phone is possible. We collect 200 signals for any camera status. Half of the measurements are kept for testing. We set T_M and T_S as 0.5 ms and 1.3×10^{-5} ms, respectively. The number of non-overlapping samples, O_S, and `FFT` size, $\mathcal{F}$, are selected as 256 and 4096, respectively. These parameters are also used in the rest of the paper unless otherwise stated.



(a) (b)

Fig. 4. Values of output layer a) before, and b) after applying softmax operation.

Since there are at most three possible outcomes for a given phone, i.e., rear-camera is on, front-camera is on, or both cameras are off, we set the size of the output layer as three. We achieve 100% accuracy rate for each considered phone. The outcome of the output layer for iPhoneSE is given in Fig. 4. In Fig. 4a, we plot the outcomes before applying softmax function, and in Fig. 4b, we plot the normalized outcomes with softmax function. Each axis in these figures represents the outcome of the considered neuron. We observe that although both figures demonstrate perfect clustering for the camera status, the outcome of the softmax function makes it clear why we achieve 100% accuracy rate for each considered phone.

Being motivated by the results of the previous experiment, we modify the CNN model so that the classification of the phone brands are possible irrespective of camera status. The number of the output nodes is increased to four which is equivalent to the number of the tested devices. The confusion matrix for the testing measurements is given in Fig. 5. We observe that there is only one inaccurate classification for the testing signals. Here, we need to note that Samsung Centura has only rear-camera, therefore, the number of test signals is less than the other brands (100 testing signal is missing corresponding to the front-camera).

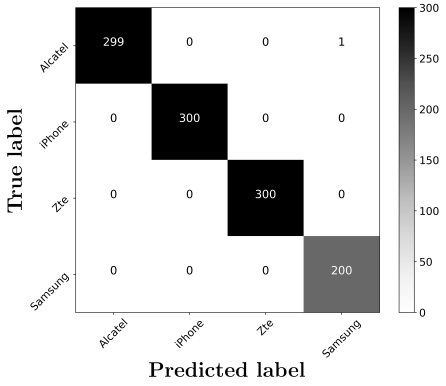


Fig. 5. Confusion matrix for the test data to classify the brand of phones.

Brand classification results reveal that clustering the brand even with no camera activity is possible. Therefore, the next goal is to identify both the brand and camera status of the measurements. In this regard, we first increase the node number of output layer to eleven (which is equivalent to the number of brand and camera status combinations). Compared to the previous experiment, each class has the same number of measurements for training and testing. The confusion matrix for the experiment is given in Fig. 6. We observe that out of 1100 measurements only two of them are misclassified. One of the inaccurate classification is when two phones are idle, and the other one is when rear-camera or front-camera of the same phone is active. Both errors are because of the high similarity between the corresponding classes. For example, if we consider Alcatel, we observe that the same frequency components are activated with different power levels when rear or front camera is active separately [18].

The inaccurate classification can be corrected if we in-

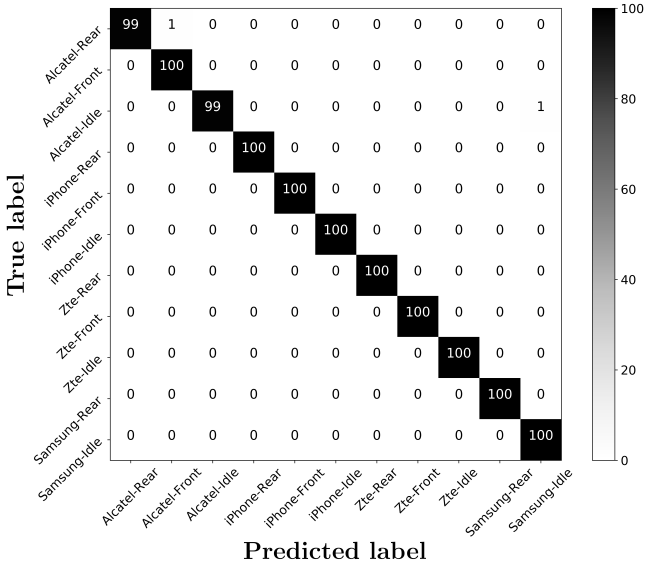


Fig. 6. Confusion matrix for the test data to classify both the brand and camera status of phones.

crease training size or decrease the number of clusters. To investigate, we have disregarded the measurements corresponding to idle status and work on the data that rear or front camera is active, hence, seven different clusters. Again, we only modify the number of nodes at the output layer as seven. With the modified network, we accurately label all test signals. Then, we apply the same methodology only to the classes where the phones are idle. However, we still obtain the same inaccurate classification even with the reduced CNN model where the output layer contains only four nodes. We provide all the results for the experiments in Table 1 to demonstrate the strength of the proposed methodology in this paper. Forth column of the table contains the numbers of test data for each experiment, and the fifth column presents the number of classification errors.

Table 1. Experimental Results for the proposed CNN model.

# of Classes	Phones	Camera	# of Test Data	# of Error
3	Alcatel	All	300	0
3	ZTE	All	300	0
3	Samsung	Rear & Idle	200	0
3	iPhone	All	300	0
4	All	Camera status independent	1100	1
11	All	All	1100	2
7	All	Rear & Front	700	0
4	All	Idle	400	1

The proposed CNN model overcomes the difficulties faced with the methodology given in [18] because differentiating the brands of the phones is possible even when the phones are idle. This is mostly because the proposed model can reveal non-linear information embedded within the measurement signals. Moreover, T_M is ten times smaller which shows that the proposed model is more resilient to the irrelevant frequency components because better classification results are obtained in a shorter measurement period.

Finally, we investigate the effect of FFT size, $\mathcal{F}$, and number of training, $\mathcal{T}$, on the accuracy rate of the proposed model. We consider three different training lengths where $\{\mathcal{T} \in \{10, 50, 100\}\}$, and three FFT sizes where $\{\mathcal{F} \in \{256, 1024, 4096\}\}$. The experimental results with different combinations of $\mathcal{T}$ and $\mathcal{F}$ are given in Fig. 7.

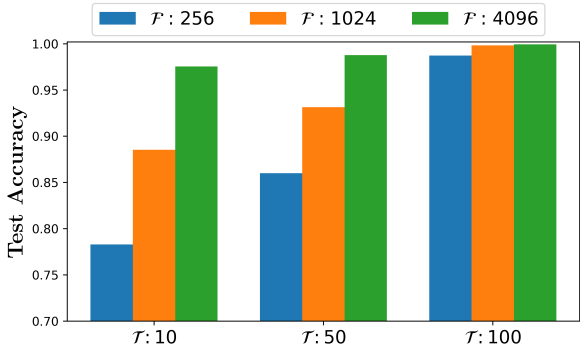


Fig. 7. Effect of FFT size ($\mathcal{F}$) and training length ($\mathcal{T}$) on the test accuracy.

We observe that the accuracy gets better as both FFT size and the training size increase. Increasing the window size of the STFT operation increases the resolution of the signal in the frequency domain, therefore, it provides more information to the proposed model. However, the effect of increasing the window size degrades as the training length increases, and vice versa. However, if both parameters are small, the accuracy rate decreases dramatically. Hence, to obtain better accuracy rates, a proper window size and training length have to be selected.

4. CONCLUSION

A CNN model is implemented to identify both the brand of a phone, and the status of its camera. Before feeding input to the neural network, we extract features by averaging STFT magnitudes of the measured signal, which helps to reduce the input dimension of the neural network, and to filter spurious emissions. The proposed CNN model contains two convolutional layers, which are followed by max-pooling layers, and four fully-connected layers. We achieve more than 99% classification accuracy in the test phase, and even perfect classification when the camera status of a single phone is considered. The results reveal the severity of leakages due to EM emanations, hence, we believe some countermeasures are required to develop to prevent attacks based on these information emissions.

5. REFERENCES

[1] D Agrawal, B Archambeult, J R Rao, and P Rohatgi, "The EM side-channel(s)," in *Proceedings of Cryptographic Hardware and Embedded Systems - CHES 2002*, 2002, pp. 29–45.

[2] Michael Hutter and Jorn-Marc Schmidt, "The temperature side channel and heating fault attacks," in *Smart Card Research and Advanced Applications*, vol. 8419 of *Lecture Notes in Computer Science*, pp. 219–235. Springer International Publishing, 2014.

[3] J. Brouchier, T. Kean, C. Marsh, and D. Naccache, "Temperature attacks," *Security Privacy, IEEE*, vol. 7, no. 2, pp. 79–82, March 2009.

[4] Dan Boneh and David Brumley, "Remote Timing Attacks are Practical," in *Proceedings of the USENIX Security Symposium*, 2003.

[5] P Kocher, "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems," in *Proceedings of CRYPTO'96, Springer, Lecture notes in computer science*, 1996, pp. 104–113.

[6] W Schindler, "A timing attack against RSA with Chinese remainder theorem," in *Proceedings of Cryptographic Hardware and Embedded Systems - CHES 2000*, 2000, pp. 109–124.

[7] David Gullasch, Endre Bangerter, and Stephan Krenn, "Cache games–bringing access-based cache attacks on aes to practice," in *Security and Privacy (SP), 2011 IEEE Symposium on*. IEEE, 2011, pp. 490–505.

[8] Yukiyasu Tsunoo, Teruo Saito, Tomoyasu Suzaki, Maki Shigeri, and Hiroshi Miyauchi, "Cryptanalysis of des implemented on computers with cache," in *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 2003, pp. 62–76.

[9] Daniel Genkin, Adi Shamir, and Eran Tromer, "Rsa key extraction via low-bandwidth acoustic cryptanalysis," in *Annual Cryptology Conference*. Springer, 2014, pp. 444–461.

[10] P Kocher, J Jaffe, and B Jun, "Differential power analysis: leaking secrets," in *Proceedings of CRYPTO'99, Springer, Lecture notes in computer science*, 1999, pp. 388–397.

[11] Elke De Mulder, Siddika Berna Örs, Bart Preneel, and Ingrid Verbauwhede, "Differential power and electromagnetic attacks on a FPGA implementation of elliptic curve cryptosystems," *Computers & Electrical Engineering*, vol. 33, no. 5-6, pp. 367–382, 2007.

[12] Alenka Zajić and Milos Prvulovic, "Experimental demonstration of electromagnetic information leakage from modern processor-memory systems," in *IEEE Transactions on Electromagnetic Compatibility, Volume: 56, Issue: 4,*, 2014, p. 885893.

[13] B. Berkay Yilmaz, R. Callan, M. Prvulovic, and A. Zajić, "Quantifying information leakage in a processor caused by the execution of instructions," in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, Oct 2017, pp. 255–260.

[14] B. B. Yilmaz, M. Prvulovic, and A. Zajić, "Electromagnetic side channel information leakage created by execution of series of instructions in a computer processor," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 776–789, 2020.

[15] Wim Van Eck, "Electromagnetic radiation from video display units: an eavesdropping risk?," *Computers & Security*, vol. 4, no. 4, pp. 269–286, 1985.

[16] Markus Guenther Kuhn, *Compromising emanations: eavesdropping risks of computer displays*, Ph.D. thesis, University of Cambridge, 2002.

[17] R. Callan, A. Zajić, and M. Prvulovic, "FASE: Finding amplitude-modulated side-channel emanations," in *2015 ACM/IEEE 42nd Annual International Symposium on Computer Architecture (ISCA)*, June 2015, pp. 592–603.

[18] Baki Berkay Yilmaz, Elvan Mert Ugurlu, Milos Prvulovic, and Alenka Zajić, "Detecting cellphone camera status at the distance by exploiting electromagnetic emanations," in *Military Communications Conference (MILCOM), MILCOM 2019-2019 IEEE*. IEEE, 2019.