Fake it till you make it: Data Augmentation using Generative Adversarial Networks for all the crypto you need on small devices

Naila Mukhtar[1] , Lejla Batina[2], Stjepan Picek[3], and Yinan Kong[1]

[1] Macquarie University, Australia
{naila.mukhtar,yinan.kong}@mq.edu.com
[2] Radboud University Nijmegen {lejla@cs.ru.nl}
[3] Delft University of Technology, The Netherlands {picek.stjepan@gmail.com}

Abstract. Deep learning-based side-channel analysis performance heavily depends on the dataset size and the number of instances in each target class. Both small and imbalanced datasets might lead to unsuccessful side-channel attacks. The attack performance can be improved by generating traces synthetically from the obtained data instances instead of collecting them from the target device. Unfortunately, generating the synthetic traces that have characteristics of the actual traces using random noise is a difficult and cumbersome task. This research proposes a novel data augmentation approach based on conditional generative adversarial networks (cGAN) and Siamese networks, enhancing in this way the attack capability. We present a quantitative comparative machine learning-based side-channel analysis between a real raw signal leakage dataset and an artificially augmented leakage dataset. The analysis is performed on the leakage datasets for both symmetric and public-key cryptographic implementations. We also investigate non-convergent networks' effect on the generation of fake leakage signals using two cGAN based deep learning models. The analysis shows that the proposed data augmentation model results in a well-converged network that generates realistic leakage traces, which can be used to mount deep learning-based side-channel analysis successfully even when the dataset available from the device is not optimal. Our results show potential in breaking datasets enhanced with "faked" leakage traces, which could change the way we perform deep learning-based side-channel analysis.

Keywords: Machine learning-based Side-channel Attacks · ASCAD · Elliptic Curves Cryptography · Data Augmentation · Signal Processing.

1 Introduction

Recently, deep learning-based attacks have been extensively studied for improving the profiling side-channel analysis (SCA) attacks, see, e.g. [MPP16, KPH+19, ZBHV19]. Profiling attacks are the class of side-channel attacks in which the adversary is assumed to have access to the target device's open copy. Then, the attacker uses that copy of a device to build a strong profiling (deep learning) model. With it, in the second phase of the attack, the attacker infers the secret information, e.g., the secret key from the target device based on the profiling model and measurements of some physical activity of the target device while cryptographic implementation is running on it. The deep learning model's performance can suffer if enough data is not provided during the training phase. What is more, using deep learning may not even make sense if not enough data is available. This lack of availability could be a consequence of implemented countermeasure prohibiting

collecting a large number of traces or due to the evaluation setup [PHG19]. Additionally, it is rather common to use side-channel leakage models like the Hamming weight or Hamming distance, which will result in an imbalanced class scenario [PHJ+18].

Deep learning is data-hungry. Not providing enough data can mean that either we do not reach the full potential of a certain method or, in more extreme cases, that the method shows very poor performance. One common reason for it is overfitting, where the deep learning model learns how to model the noise, making it difficult to generalize to the previously unseen examples. To fight against it, researchers commonly use techniques like 1) hyperparameter tuning - to find architectures that are better tuned for the task and thus have less capacity to overfit, 2) regularization - to lower the complexity of a neural network model during training, and thus prevent the overfitting, and 3) data augmentation - to provide additional (synthetic) examples, utilizing the capacity of a model better, but to also regularize the model with noise inherent to the synthetic examples. Each of those techniques has its advantages and drawbacks, and they are also successfully applied to the side-channel domain. Interestingly, while hyperparameter tuning and regularization (e.g., dropout, L2 regularization) are commonly used in SCA, data augmentation received less attention, despite very good preliminary results. One possible reason lies in the difficulty of clearly visualizing how a successful synthetic side-channel measurement should look (something much simpler in, e.g., image classification domain).

Cagli et al. proposed the first data augmentation setup for deep learning-based SCA to counter the effects of clock jitter countermeasure [CDP17]. Still, the authors do not consider scenarios where the number of measurements is significantly limited. Picek et al. presented the results with traditional machine learning data augmentation techniques and concluded that Synthetic Minority Over-sampling Technique (SMOTE) could aid in data generation, resulting in improved attack performance [PHJ+18]. Differing from us, the authors used the Hamming weight leakage model, which resulted in an imbalanced dataset. In our work, we used intermediate values resulting in more classes and more challenging analyses. Luo et al. used a mixup data augmentation technique where new synthetic examples are based on randomly selected combinations of existing examples [LZW+21]. The authors conducted experiments for several datasets and leakage models and obtained mostly similar behavior for the original and mixup traces.Generative Adversarial Networks (GAN) is another popular data augmentation technique that is widely used in the image processing domain for generating fake images [LHY+20], which significantly improves the machine learning model's performance [GPAM+14]. Only one existing study presents the realization of using GAN-generated fake signals as leakage signals for machine learning-based side-channel analysis [WCL+20]. However, the authors use more profiling traces and the HW leakage model (which will result in fewer classes), making their work significantly different from ours. What is more, this work misses providing a detailed analysis of the GAN network before using it for the fake data generation.

GAN's performance for generating fake images/signals depends on the generator's progressive learning based on the discriminator's response. The design and selection of a GAN play an important role in generating realistic leakage signals. A well-convergent GAN network will generate traces carrying relevant/significant features similar to the original data samples. Designing a GAN-based model with optimum convergence or equilibrium point is one of the greatest challenges for generating fake signals [KAHK17] that contain the characteristics of real leakage traces. Several techniques, presented for fake image generation, can help achieve convergence, including feature matching [SGZ+16], conditional GAN (cGAN) [MO14], and semi-supervised learning [SGZ+16]. In our work, we generated 50% traces for each class ($256\times150\times2$ real and fake traces for AES and $16\times150\times2$ real and fake traces for ECC), making the setting very challenging. The selection of a well-convergent cGAN based network is an important step that should be thoroughly explored before data generation. Moreover, we provide a detailed insight on how cGAN generates

relevant data and works for side-channel analysis. We also present a fast convergent GAN network by utilizing the configurations/characteristics of the Siamese network. Our approach is inspired by the fake image generation presented in [HLSC18].Our presented generic model can generate fake data for various leakage datasets, including symmetric and public-key algorithm implementations. We provide a comparative analysis with the existing simple dense layer-based cGAN network used for leakage generation.

Specifically, we list our contributions as follows:

- We present a layered approach for generating the 1-dimensional fake signals for machine learning-based side-channel analysis (ML-SCA). Our approach combines Siamese network and Conditional GAN characteristics with an extra model loss monitoring layer introduced to detect the model convergence. The visual representation of the loss function of the proposed data augmentation technique helps in analyzing the well-converged GAN model. A well-converged network helps in generating indistinguishable fake traces that will give the same insights to the data as that of the original signals. With this, we showcase the relevance of "real vs. fake" data and exhibit successful attacks with synthetic data that could impact various real-world use cases and security applications.
- We provide a comparative analysis exhibiting the fake leakage trace datasets' effect on the side-channel model training performance. These fake leakage traces are generated from various converging points during the proposed Siamese-cGAN model training, which helps to analyze the importance of the well-converged models.
- The proposed Siamese-cGAN model can be generalized to any leakage dataset (from varying cryptographic algorithm implementations). To demonstrate this, we have trained our proposed model on datasets containing either symmetric or public-key algorithm implementations, using two different neural networks for generator and discriminator. Best performing neural networks are further selected for analysis. While there are a few results showing the benefits of data augmentation for symmetric-key implementations, to the best of our knowledge, this was not done before for public-key implementations.
- We provide a comparative analysis of our proposed Siamese-cGAN model with the existing used cGAN model [WCL+20] (named as $cGANModelA$ in this study) for generating fake leakage traces. [1]
- The performance of the Siamese-cGAN data augmentation model is evaluated by applying the actual machine learning-based side-channel attack on the generated leakage traces using four different neural network architectures (one multilayer perceptron (MLP) and two Convolutional Neural Networks (CNNs) for symmetric algorithm implementation leakages and one CNN for public-key implementation leakages). Our results show that the fake data samples generated from the well-convergent model combined with 50% real data successfully recover the secret with similar efficiency as real data traces alone. What is more, for the ASCAD case, the key rank suggests even improved results for the dataset consisting of real and fake traces. We emphasize that the goal of our approach is not only to improve the attack performance in scenarios where there are enough real measurements for a successful attack. Rather, we envision it for constrained settings where more measurements than available are needed to break the target. We think here of implementations that randomize the secret (key) after a number of algorithm's execution such that the adversary can collect only a limited number of traces for the analysis.

The rest of the paper is organized as follows. Section 2 gives an overview of the profiled and machine learning-based side-channel attacks, cGAN background, machine learning, and cryptographic algorithms used for analysis. In Section 3, we provide an overview of

[1] We note that the provided details were not sufficient to ensure the reproducibility of the results, so we did our best to infer the used architecture from the description.

related works. Section 4 explains the layered GAN approach for leakage traces generation for ML-SCA, while Section 5 discusses the experiments conducted along with the results. Finally, Section 6 concludes the paper and discusses possible future research directions.

2 Preliminaries

In this section, we discuss the profiled side-channel attacks, (conditional) generative adversarial networks (GANs), data augmentation, and machine learning techniques we investigate in this paper. Finally, we briefly discuss the datasets we use.

2.1 Profiled Side-channel Attacks

The profiled attack represents the most powerful side-channel attack where the adversary has access to the target device's open copy. There are two phases of the attack: the profiling phase and the attack phase. In the profiling phase, the adversary creates a profile of the device with all the possibilities for the data leakages and then uses that profile (template in the case of template attack [CRR03] or machine learning model) in the attack phase to distinguish/predict the unknown secret information [WOS14].

Commonly, profiled attacks are divided into classical ones like template attacks [CRR03] and stochastic model [SLP05], and machine learning-based attacks [LBM14, MPP16]. Machine learning-based side-channel attacks (ML-SCA) follow the same steps as the classical profiling attacks. More precisely, they have two phases: training phases (profiling phase) and testing phase (attack phase). The adversary can train the model with the leakage examples collected from the identical copy of a target device and then evaluate the trained model using previously unseen examples.

2.2 Generative Adversarial Networks (GANs)

The generative adversarial networks were first introduced by Goodfellow et al. in 2014 [GPAM+14]. Since then, many variations of GAN have been proposed, including Conditional GAN (cGAN), Deep Convolutional GAN (DC-GAN), Information Maximizing (InfoGAN), and Stacked GAN (StackGAN) [RMC16, MO14, CDH+16, ZXL+17, BDS18].

A generative Adversarial Network (GAN) is a neural network architecture for training generative model which is then capable of generating plausible data. It consists of two neural networks adversarial models, discriminator D and generator G. Generator G network generates the fake data with random noise input z, and discriminator D network discriminates between real and fake data. Real data is labeled as '1' and artificially generated (fake) data is labeled as '0'.

The discriminator's task is to distinguish the real and generated data instances, whereas the generator's task is to improve the model based on the feedback from the discriminator. GANs are based on the concept of a zero-sum non-cooperative game where one network (discriminator) is trying to minimize the loss, and the other network (generator) is trying to maximize the loss (this min-max problem as given by equation 1), meaning the discriminator is trying to distinguish and classify the data instances as real or fake. However, the generator is trying to generate data traces that are alike. This makes it hard to find a good convergence point. GAN converges when both D and G reach a Nash equilibrium, meaning that one (D/G) will not change its actions anymore, no matter what opponent (D/G) does. This is the optimal point which adversarial loss in GANs aims to optimize.

$$E_x[log(D(x))] + E_z[log(1 - D(G(z)))]. \tag{1}$$

There are different techniques used to improve the performance of GANs by improving the convergence. One of such techniques is Conditional GAN (cGAN).

2.3 Conditional Generative Adversarial Networks (cGANs)

GANs help in generating plausible data, but there is no way of controlling what kind of random data they will generate. To handle the issue, GAN can be conditioned with some extra data to control the information being generated. GAN with an extra condition is called Conditional Generative Adversarial Network (cGAN). The extra data in cGANs is the class label of the data samples. Using class labels for cGAN training has two main advantages: firstly, it helps in improving the GAN performance, and secondly, it helps in generating the specific target class samples.

In the training process of the cGAN model, the generator network generates data based on the label and the random input and tries to replicate the actual distribution in the real data samples. The generated samples are of the same structure as that of the input labeled data. In the next step, the real and the generated/fake data are given as input to the discriminator. The discriminator is first trained with the original/real labeled data samples and is then trained with the fake generated data samples. Similar to GANs, the discriminator's task is to separate the fake from real data samples, which helps the generator is generating better (realistic) samples [MO14].

2.4 Data Augmentation

Data augmentation is an umbrella term representing various techniques used to increase the amount of data by adding (slightly) modified copies of already existing examples or newly created synthetic examples from existing data. Data augmentation can act as a regularization technique and will help reduce overfitting. While data augmentation can be applied to any domain, most of the results and techniques were developed for data augmentation in image classification [SK19].

2.5 Machine Learning Algorithms

Based on the deep learning-based side-channel attacks (DL-SCA) performance on various cryptographic algorithms [CDP17, WPB19, MMKA18], we tested our newly generated datasets using two state-of-the-art machine learning algorithms: multilayer perceptron (MLP) and Convolutional Neural Network (CNN). MLP and two variations of CNN [BPS+20, ZBHV19] are used to evaluate the AES dataset, and one CNN architecture [WPB19] is used to evaluate the ECC dataset.

2.5.1 Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) belongs to a class of feed-forward neural network algorithms consisting of one input layer, one or more hidden layers, and one output layer. Each layer consists of nodes/neurons, except the input layer. These nodes utilize a nonlinear activation function to learn the patterns in the data. MLP uses supervised learning-based backpropagation to change the weights on the connections during the data processing. A batch of data is presented to these fully connected networks during each epoch for learning.

2.5.2 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN or ConvNet) is a class of deep learning neural networks and is principally based on convolutions. A CNN architecture consists of an input and an output layer and few hidden layers. Hidden layers are usually convolutional layers having an activation function, followed by pooling layers that help reduce the dimension

of data. CNN has the capability of learning the patterns from noisy side-channel leakages without any preprocessing [CDP17].

2.6 Cryptographic Algorithms Under Evaluation

For our analysis, we investigate the performance of our proposed model on two publicly available datasets. One dataset corresponds to the implementation of Advanced Encryption Standard (AES) and the other to the Elliptic Curve Cryptographic (ECC) implementation.

- The ASCAD dataset ([BPS+20]) is the first dataset that acts as a basis for comparative analysis of deep learning-based side-channel analysis (DL-SCA). The traces are collected from the masked AES-128 bit implementation on an 8-bit AVR microcontroller (ATmega8515). The leakage model is first-round S-box (Sbox$[P(i)_3 \oplus k^*]$) where the third byte is exploited (as that is the first masked byte). There are 60 000 total traces along with the metadata (plaintext/ciphertext/key). These traces are further split into two datasets, one for training (profiling) consisting of 50 000 and the other for testing (attacking), consisting of 10 000 traces. Each trace consists of 700 features. The labels are stored in a separate file.

- The publicly available ECC dataset ([WPB19],[ecc19]) consists of power consumption traces collected from a Pinata development board (developed by Riscure). The board is equipped with a 32-bit STM32F4 microcontroller (with an ARM-based architecture), running at the clock frequency of 168 MHz and having Ed25519 implementation of WolfSSL 3.10.2. The target is the profiling of a single EC scalar multiplication operation with the ephemeral key with the base point of curve Ed25519. The 256-bit scalar/ephemeral key is interpreted as slices of four nibbles. Hence there are a total of 16 classes/labels in the dataset. The dataset has a similar format as ASCAD. The database consists of two groups of traces; profiling traces and attack traces. Each group further consists of "TRACES" and "LABELS". Each raw trace consists of 1 000 features, representing the nibble information used during the encryption. Profiling and attack traces groups consist of n_p and n_a tuples, with a corresponding label for each trace. In total, there are 6 400 traces, out of which 80/20 are kept for profiling ($n_p = 5\,120$) and attacking ($n_a = 1\,280$).

The profiling traces from both datasets are used to train the cGAN models. Half of the real traces per class are kept in the final dataset, along with the generated data samples. The attacking traces are used for evaluating the performance of the DL-SCA model trained with the new dataset.

It should be noted that the purpose of this research is to analyze the effect of artificially generated features in data traces for environments where the adversary has an additional constraint on collecting leakage traces to form a dataset. The presented methodology can be extended to produce and test the fake leakage traces for any other cryptographic algorithms.

2.7 Siamese Neural Network

Siamese neural network (also called twin/identical neural network) is an artificial neural network architecture that consists of two similar neural networks (having the same weights and network parameters) and is capable of processing two different input vectors to produce comparable output vectors [KZS15]. The two neural networks are feed-forward multilayer perceptrons that work in tandem and are trained in a backpropagation manner. The idea behind Siamese neural network is not to learn to classify the labels but to learn to discriminate between the input vectors. Hence a special loss function, contrastive loss, or Triplet Loss is used for the training of the network [LSP21]. For training the network, the pairs (x_i, x_j) of input vectors are prepared; few pairs consist of similar vectors, and few

pairs consist of dissimilar vectors. The similar vector pair is labeled as y='1', whereas the dissimilar pair is labeled as y='0'. Each pair is fed to the Siamese network, and distance is computed to check the similarity. The output vectors from each network are compared using cosine or Euclidean distance and can be considered as a semantic similarity between projected representation of the input vectors [Chi21].

3 Related Works

Data augmentation represents a set of techniques to reduce overfitting and improve the supervised machine learning task (commonly classification). Data augmentation is a well-researched topic, mostly framed in the context of image data augmentation [SK19]. While there are multiple ways to divide the data augmentation techniques, a common one is on techniques transforming the input, deep learning techniques (where our work also belongs), and meta-learning.

Data augmentation in SCA is used to improve side-channel attack performance and can be put in the same general direction as, e.g., works exploring how to improve hyperparameter tuning. As data augmentation increases the amount of data, it is commonly considered in the deep learning perspective as there, very large datasets are beneficial.

There are significantly more works considering symmetric-key cryptography and deep learning-based SCA than public-key cryptography. What is more, we are not aware of any works exploring the perspective of data augmentation for public-key cryptography.

3.1 Deep Learning-based SCA on AES

The first investigation that uses convolutional neural networks for side-channel attacks on AES is conducted by Maghrebi et al. [MPP16]. This work represents a significant milestone for the SCA community as it demonstrated how deep learning could be used without feature engineering and efficiently break various targets. [2]

Cagli et al. investigated how deep learning could break implementations protected with jitter countermeasures [CDP17]. This work is highly relevant as it introduced data augmentation to the SCA domain. The authors used two data augmentation techniques: shifting (simulating a random delay) and add-remove (simulating a clock jitter). Picek et al. investigated how reliable are machine learning metrics in the context of side-channel analysis. Their results showed that machine learning metrics could not be used as sound indicators of side-channel performance [PHJ+18]. Additionally, as the authors used the Hamming weight leakage model that results in class imbalance, they utilized a well-known data balancing technique called SMOTE, showing that the attack performance can be significantly improved. Kim et al. explored how to design deep learning architectures capable of breaking different datasets [KPH+19]. Additionally, they used Gaussian noise at the input to serve as a regularization factor to prevent overfitting. Luo et al. investigated how mixup data augmentation can improve both CPA and deep learning-based side-channel attacks [LZW+21].

Next, several works aimed at improving neural network performance by providing a systematic approach for tuning neural network architectures. Zaid et al. were the first to propose a methodology to tune the hyperparameters related to the convolutional neural network size (number of learnable parameters, i.e., weights and biases) [ZBHV19]. Wouters et al. [WAGP20] further improved upon the work from Zaid et al. [ZBHV19] by showing how to reach similar attack performance with even smaller neural network architectures. Rijsdijk et al. used reinforcement learning to design in an automated way

[2] We note earlier works are also using neural networks like multilayer perceptron, but the results were in line with other machine learning techniques, and researchers commonly used feature engineering to prepare the traces.

small convolutional neural networks that perform well [RWPP21]. Following a different approach to improving the attack performance, Wu and Picek showed how denoising autoencoder could be used to reduce the effect of countermeasures [WP20]. The first work that considers the usage of GANs (more specifically, cGANs) for the SCA domain is made by Wang et al. [WCL+20]. While this work shows the potential of GANs in SCA, the results indicate that a large profiling set is required to construct useful synthetic data.

3.2 Deep Learning-based SCA on Public-key Cryptography

There are several works using template attack (and its variants) to attack public-key cryptography, see, e.g., [MO08, HMH+12, BCP+14, OPB16]. Lerman et al. used a template attack and several machine learning techniques to attack an unprotected RSA implementation [LBM14]. Carbone et al. used deep learning to attack a secure implementation of RSA [CCC+19]. The authors showed that deep learning could reach strong performance against secure implementations of RSA. Weissbart et al. showed a deep learning attack on EdDSA using the curve Curve25519 as implemented in WolfSSL, where their results indicate it is possible to break the implementation with a single attack trace [WPB19]. Weissbart et al. considered deep learning-based attacks on elliptic curve Curve25519 protected with countermeasures and showed that even protected implementations could be efficiently broken [WCPB20]. Perin et al. used a deep learning approach to remove noise stemming from the wrong choice of labels after a horizontal attack [PCBP21]. The authors showed that protected implementations having an accuracy of around 52% after a horizontal attack could reach 100% after deep learning noise removal. Zaid et al. introduced a new loss function called ensembling loss, generating an ensemble model that increases the diversity [ZBHV21]. The authors attacked RSA and ECC secure implementations and showed improved attack performance.

4 Proposed Approach

This section explains the proposed layered GAN model used for selecting the optimal model for generating leakage signals from random noise.

4.1 Data Splitting

The leakage data traces L are collected from the device while AES or ECC algorithms encryptions E are performed using the secret key K or scalar/ephemeral key K, respectively. The labeled collected traces are then divided into two sets: Training ($D_{Training}$) and Testing ($D_{Testing}$). The training set is used for training the Siamese-cGAN model, which then produces the fake traces. The newly generated dataset (real+fake traces) is used to train the DL-SCA model, and the testing set is used for evaluating the performance of the SCA model. For a fair evaluation of the trained Siamese-cGAN model, the testing set is never shown to the network during the Siamese-cGAN model training process.

For the rest of the paper, "cGAN models" or "Siamese-cGAN models" refer to the model used for generating data. However, "DL-SCA models" refer to the deep learning models, which are used to evaluate the performance of the generated data by applying profiling side-channel attacks.

4.2 Siamese-cGAN Model for Data Augmentation

In contrast to the standard GANs, conditional GANs (cGANs) perform conditional generation of the fake data based on the class label rather than generating signals blindly. The labels c of the data traces/instances are used to train GANs in a zero-sum or adversarial

manner to improve the learning of the generator (G). As mentioned before, the generator's G task is to generate the leakage signals that carry similar properties as the original traces using the random noise z and the latent space input ls. The discriminator's D task is to distinguish real and fake signals. In the cGAN training process, first, the discriminator D is trained with the labeled real data traces T_{real}, and then the discriminator D is trained with the fake generated signals $G(z)$ or T_{Gen}. The objective function for cGAN is given by Eq. (2).

$$E_{x\sim T_{Real}}[log(D(x|c))] + E_z[log(1 - D(G(z|c)))]. \tag{2}$$

In our proposed design of Siamese-Conditional Generative Adversarial Network ($Siamese-cGAN$), we are combining cGAN with the Siamese network concept. Siamese network is an architecture in which two identical/twin networks carrying the same weights are trained with two different inputs. In the proposed model, two generators $G1$ and $G2$ take two random input noise vectors $z1$ and $z2$, and generate fake signals $G(z1)$ and $G(z2)$, respectively, in a Siamese fashion. Both the generators share the same network weights, and only the input is different. The discriminator D is first trained with the labeled real data T_{real} and then with the fake data T_{gen}, originating from the two twin generator networks $G1$ and $G2$.

As mentioned before, convergence is a challenging issue in training GANs. In some cases, the model converges and then starts diverging again; that is, it forgets its learned examples. Several techniques include memory-based learning, to handle such scenarios [WHL+19]. Training the model simultaneously with the random noise from two sources can help obtain a better-converged model. Moreover, to analyze the impact of convergence, we introduced another layer in the two-step cGAN model to analyze the model performance for leakage traces. This layer monitors the real traces loss L_{real}, generated traces loss L_{gen}, and GAN model Loss L_{GAN}.

Let $D_{GAN\rightarrow R}$ represent the loss difference between L_{GAN} and L_{real}, and $DGAN \rightarrow G$ represent the loss difference between L_{GAN} and L_{real}, then the average of loss differences over last t iterations will be given by:

$$\frac{1}{t}\sum_{i=1}^{t}(|D_{GAN\rightarrow R}| + |D_{GAN\rightarrow G}|). \tag{3}$$

The Siamese-cGAN model stops training when the average model loss $Loss_{Avg}$ over the last t iteration is less than the average loss over the last $t * 2$ iterations. The trained Siamese-cGAN model is then used to generate the n_g fake traces T_{gen}, containing features similar to the original signals. T_{gen} and T_{real}, having n_g and n_r instances respectively, are combined together to form a resultant dataset T_GAN. This dataset is then used to train the machine learning model to analyze the generated dataset's performance with ML-SCA. A test set is set aside for a fair evaluation before adding the generated traces into the training dataset. The test set is never shown to the neural network during training. The proposed Siamese-cGAN specific for ML-SCA is shown in Figure 1.

4.3 cGAN Models for Discriminator and Generator

In the proposed Siamese-cGAN model, the selection of generator and discriminator plays a vital role. The authors in [WCL+20] recommended using a generator and discriminator with fully connected dense layers only, without any complex layers. However, in this study, we explore the possibility of using fully connected dense layers as well as the convolutional layers in discriminator. The reason for evaluating CNN layer-based network is that it has provided better fake images generation in image processing [RMC16]. Hence, for evaluating the trained Siamese-cGAN model performance, two different networks are used

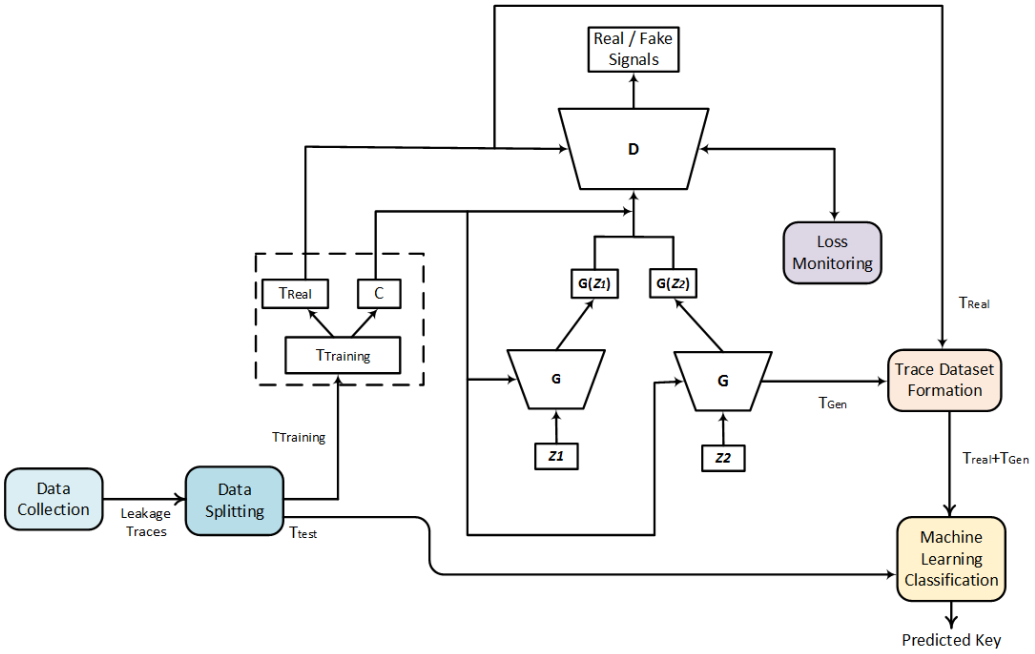


Figure 1: Proposed Siamese-cGAN architecture for ML-SCA

for the generator and discriminator. These networks are denoted as *Model A* and *Model B*. *Model A* is based on two fully connected layers for the generator and the discriminator. However, *Model B* (CNN-based) has a more complex architecture with four fully connected and one CNN layer. Batch normalization, LeakyRelu, and dropout layers are introduced, which help achieve a more stable model and help avoid overfitting. Additionally, we tested both *Model A* and *Model B* generators and discriminators with and without Siamese settings to analyze the improvements introduced by using the Siamese configuration.

Figures 2 and 3 show the structure of both the models. We use the same generator and discriminator model architectures for generating data samples for both symmetric and public-key leakages. The only parameter that needs to be changed is the size of dense layers, which changes based on the number of classes per target dataset. The size in each subsequent layer is doubled from the previous layer. That is, layer 1, layer 2, and layer 3 have a size equal to the number of classes, the number of classes*2, and the number of classes*4, respectively. The convolutional layer has a tanh activation function. The parameter details for the generator network are given in Table 1, while for discriminator, we use a dense layer (512), LeakyReLU, and Dropout set at 0.4.

Table 1: Generator Architecture Details

Hyper-parameter	Value
Input shape	(700,1) for AES, (1 000,1) for ECC
Fully Connected layer 1	number of classes
Fully Connected layer 2	number of classes*2
Fully Connected layer 3	number of classes*4
Dropout rate	0.4

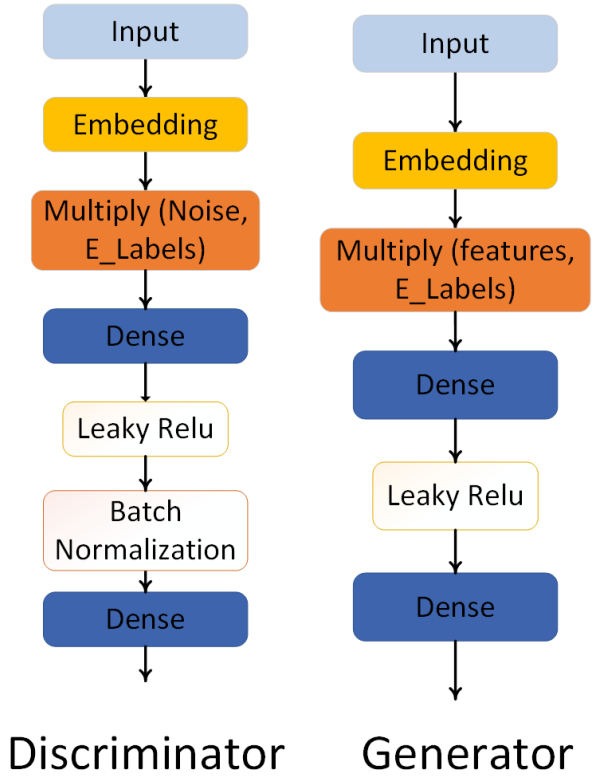


Figure 2: $Siamese - cGANModel\ A$

5 Experiments and Results

In this section, we present the experimental setup and the results of applying our proposed methodology on the publicly available datasets.

5.1 DL-SCA Evaluation Model Architectures

As mentioned in Section 2.5, we use MLP and CNN architectures to evaluate the performance of the newly generated datasets using the cGAN model. The state-of-the-art DL-SCA model architectures for evaluating the original publicly available datasets ([BPS+20, ZBHV19, WPB19]) are used in the same setting except for the batch size and epochs, which are varied between 50-200 to see the impact on the results.

For evaluating the ASCAD dataset, the first architecture is an MLP-based network from [BPS+20]. The first CNN architecture is denoted as ASCAD-CNN1 ([BPS+20]), while the second one is denoted as ASCAD-CNN2 ([ZBHV19]). For evaluating the ECC dataset performance, we use the deep learning architecture presented in [WPB19]. Note that we use the existing state-of-the-art DL-SCA model architectures to allow fair comparison. What is more, the goal of this work is not to find new deep learning architectures but to enhance the performance of the existing architectures.

The performance of the DL-SCA trained models with the newly generated datasets is evaluated using two commonly used evaluation metrics: key rank and accuracy. Accuracy represents the number of correctly classified examples divided by the total number of examples. Key rank is the position of the correct key guess in the key guessing vector. More precisely, this vector contains the key candidates in decreasing order of probability,

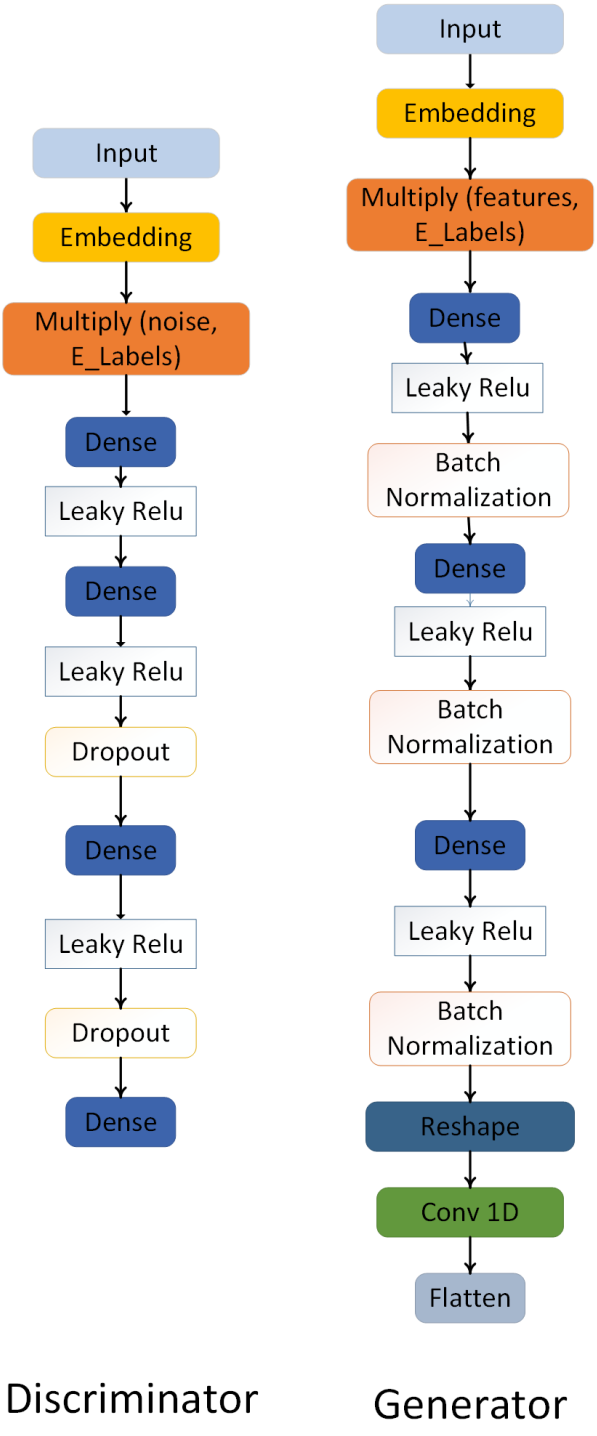


Figure 3: *Siamese – cGANModel B*

and finding the position of the correct key indicates the effort required by the attacker to break the target.

5.2 Experimental Setup

For our experiments, the proposed GAN models have been developed and trained using Keras and Tensorflow libraries. The models are trained to generate the fake data on a common computer equipped with 32Gb RAM, i7-4770 CPU with 3.40GHz, and Nvidia GTX 1080 Ti. The $Siamese - cGANModels$ training took less than 5 minutes to train and less than 5 minutes to generate 150 fake traces for 256 classes in total. Obviously, the time required for fake data generation will vary depending upon the number of classes and the number of generated fake traces per class.

We divide our experimental analysis into two sections.

- Analysis-1: first, in Section 5.3, we provide the visual representation of the trained cGAN models over 1 000 epochs. Visual representation helps identify the convergent model, which is then further selected for analysis in the second phase of analysis. We trained four cGAN models, out of which two are existing and two are newly proposed, based on the generator and discriminator as explained in Section 4.3. Details of the cGAN models are given in Table 2. We also provide the model convergence-based comparison of our proposed Siamese-cGAN-based model with the existing cGAN models [WCL+20].
- Analysis-2: second, in Section 5.4, we provide the machine learning-based side-channel analysis of two datasets; dataset containing real leakage signals only and the dataset consisting of both real and fake leakages, by training with two neural networks (MLP and CNN). For this analysis, we also show the comparison of generating leakage signals from the non-converging network and a converging network. To achieve this, we generated fake signals from various points while training the Siamese-cGAN network. More precisely, we generated the signals when the model converged the best and generated the signals when the model was the least convergent (initial epochs). Converging details of each model are given in the respective sections. This comparison is performed to highlight the fact that not any cGAN can be selected blindly. *Only a well-convergent network will generate traces that are more alike in characteristics to that of original real traces.*

Table 2: cGAN Model Details

Model Name	Description
$cGAN\ Model\ A$	Model without CNN and Siamese network
$cGAN\ Model\ B$	Model with CNN but without Siamese network
$Siamese - cGAN\ Model\ A$	Model without CNN but with Siamese network
$Siamese - cGAN\ Model\ B$	Model with CNN and with Siamese network

5.3 Analysis-1: Existing and Proposed GAN-based Approaches

In this section, we perform a convergence-based comparative analysis of our approach with the existing cGAN models. The existing cGAN networks (without layered Siamese-cGAN setting) are denoted as $CGAN\ Model\ A/B$, whereas our proposed models are denoted as $Siamese - cGAN\ Model\ A/B$. For this analysis, all four cGAN networks are trained with the $D_{Training}$ dataset and the GAN model loss for both AES and ECC, as shown in Figures 4 and 5, respectively. Figure 4 *a* and *b* presents the real, fake, and GAN loss for training with $cGAN\ Model\ A$ and $cGAN\ Model\ B$ without Siamese settings, respectively. Next, Figure 4 *c* and *d* presents the real, fake, and GAN loss for training with $Siamese - cGAN\ Model\ A$ and $Siamese - cGAN\ Model\ B$ with the Siamese setting. Similarly, Figure 5 presents all four cGAN training models' loss for the ECC dataset.

$Siamese - cGAN$ and $cGAN$ models, for the ASCAD and ECC dataset analysis, are trained for 1 000 epochs, and history is recorded every ten epoch. Hence x-axis is

scaled by 10. It can be seen that the proposed $Siamese - cGAN\ Model\ B$ architecture (based on CNN) provides the best loss convergence of real, fake, and GAN models as the models converge around 100-150 epochs and 700-1 000 epochs for AES and ECC datasets, respectively.

This shows that at this convergence point, the generator started generating traces that are similar in characteristics to the real traces, making it harder for the discriminator to discriminate between real and fake. Existing $cGAN\ Model\ A$ and $cGAN\ Model\ B$ without Siamese configuration did not converge well in 1 000 epochs. Moreover, for $Siamese - cGAN\ Model\ B$, GAN loss is high and quickly decreases in initial epochs. Hence, the proposed $Siamese - cGAN\ Model\ B$ is the robust solution for generating artificial/fake leakage signals for both algorithms as it converges better and faster than other cGAN models. Interestingly, $Siamese - cGAN\ Model\ A$ performs relatively poorly, indicating that indeed the more powerful neural network architecture was required for this task. In conclusion, from this analysis, $Siamese - cGAN\ Model\ B$ is further selected for generating the fake data in analysis phase 2.

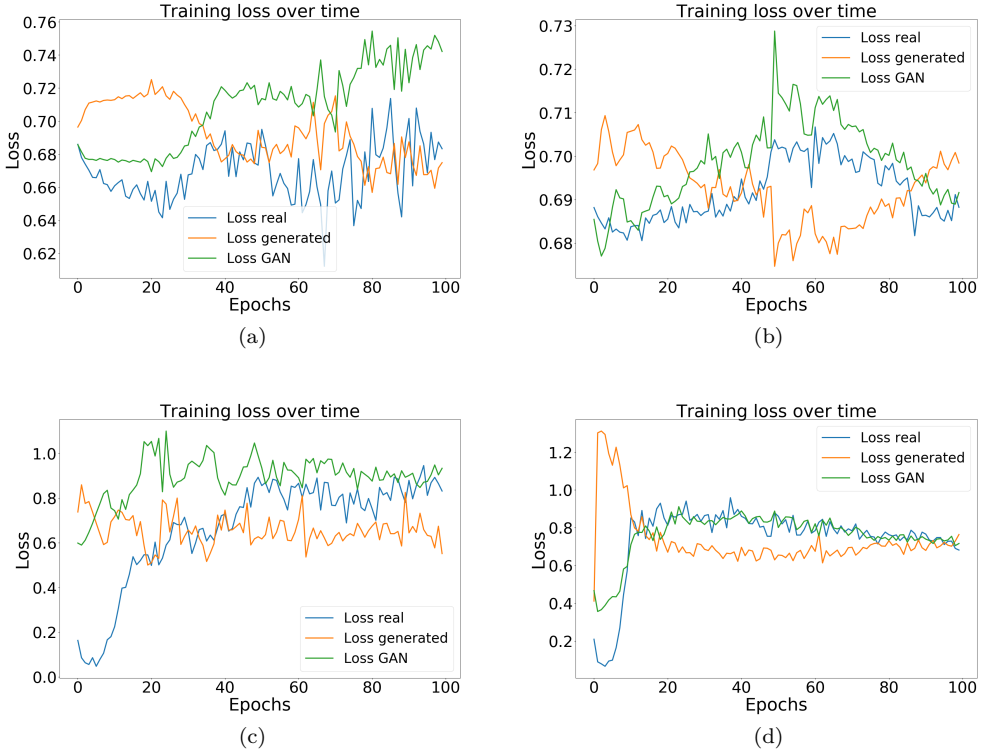


Figure 4: CGAN Model Training Loss for the ASCAD dataset (a) $cGAN\ Model\ A$, (b) $cGAN\ Model\ B$, (c) $Siamese - cGAN\ Model\ A$, (d) $Siamese - cGAN\ Model\ B$

5.4 Analysis-2: Analysis of the Proposed Siamese-CGAN for DL-SCA

Based on the results from the previous analysis, $Siamese - cGAN\ Model\ B$ is selected for further experiments in this section. Now, we perform DL-SCA (using MLP, ASCAD-CNN1, and ASCAD-CNN2) on the newly generated T_GAN (real+fake traces) datasets, generated using $Siamese - cGAN\ Model\ B$. However, for performance comparison, we also conducted analysis on the real dataset (with reduced traces per class). Hence, two

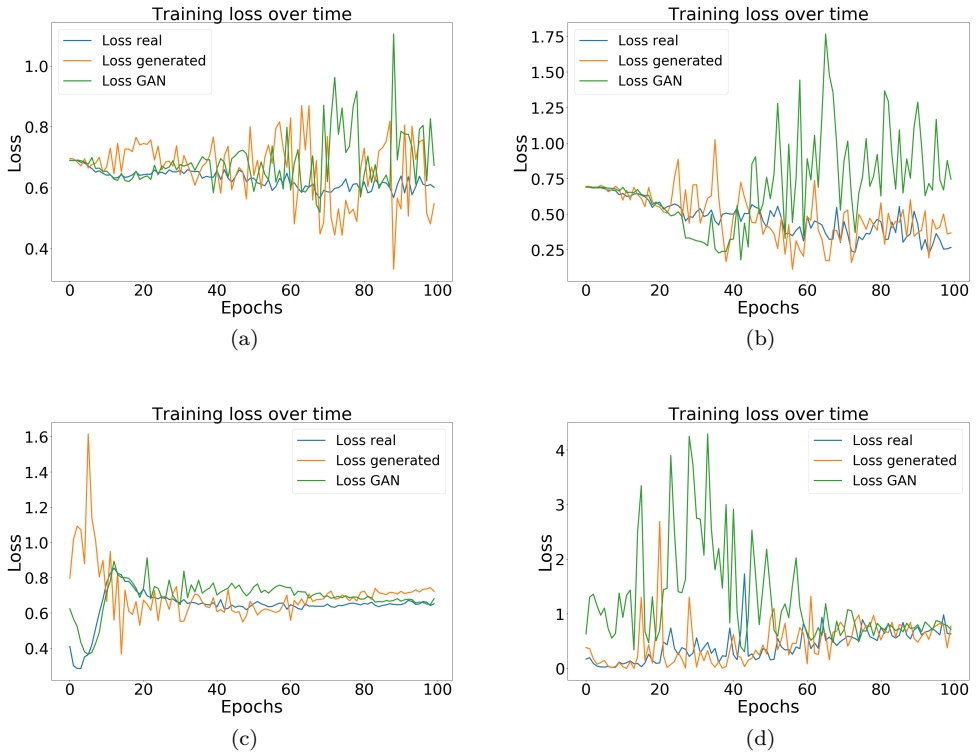


Figure 5: cGAN Model Training Loss for the ECC dataset (a) $cGAN\ Model\ A$, (b) $cGAN\ Model\ B$, (c) $Siamese-cGAN\ Model\ A$, (d) $Siamese-cGAN\ Model\ B$

datasets are formed; one with real traces only and the other (T_GAN) with both real and fake traces. For T_GAN datasets, two further datasets are formed, one for the fake data generated from the well-converged model and the second from the non-convergent model. Finally, we analyze how the key rank of a side-channel attack is impacted by all these different settings.

5.4.1 Analysis on Real Traces

In our experiments, we reduced the size of the ASCAD and ECC leakage datasets intentionally to analyze the effect of the artificially generated data traces on the small size datasets. We selected $n_r = 150$ (for each class) leakage traces from the ASCAD and ECC datasets. The reason for selecting precisely 150 traces per class is because we wanted an equal number of real traces for all the classes. In the ASCAD dataset, class 213 has a minimum number of traces (154 traces), hence 150 is selected. No artificial/fake traces are included in the training dataset, so $n_g = 0$. Hence, total number of traces in AES and ECC datasets are 38 400 (150 traces $\times$ 256 classes) and 24 00 (150 traces $\times$ 16 classes), respectively. For machine learning-based attacks, we used the previously successful machine learning side-channel attack models for AES and ECC in respective studies [BPS+20, ZBHV19, WPB19].

The purpose of using the same deep learning-based models is to show that the artificially generated traces produce the same results as the real traces with the same model architectures. When considering the ASCAD-CNN1 architecture, everything stays the same as in previous studies' analysis except that we perform normalization on the training and testing

data. For normalization, each input variable feature is scaled in the range $[-1, 1]$ by using MinMaxScaler from the Sklearn library. For MLP, 10-fold cross-validation is performed. For ASCAD-CNN2 analysis, in addition to applying normalization, a standardization is added as per the proposed architecture in [ZBHV19], and data is standardized around mean with a unit standard deviation (between 0 and 1) [GBC16, LBOM12]. For ECC, the same model is used for training and testing as proposed in [WPB19].

Figure 6 (a) shows the key rank for the real traces (38 400) dataset for the ASCAD dataset analysis using MLP and 10-fold cross-validation. We compare our results of reduced original traces with the results of MLP_{best} reported in [BPS+20], which is plotted for the trained model on 50 000 traces. We can see a slight deviation though both figures are for the trained model on real traces. Figure 6 (b) shows the key rank on reduced ASCAD dataset trained using ASCAD-CNN2. It shows key rank not approaching zero in the first 1 000 traces. This shows that the reduced dataset did not perform as expected with the existing models.

Figure 7 shows the accuracy for the real traces dataset for ECC dataset analysis using CNN architecture [ZBHV19]. For ECC, the analysis of raw real data traces shows that the private key can be recovered with 100% accuracy using CNN. It should be noted that preprocessing and alignment have not been applied to these datasets.

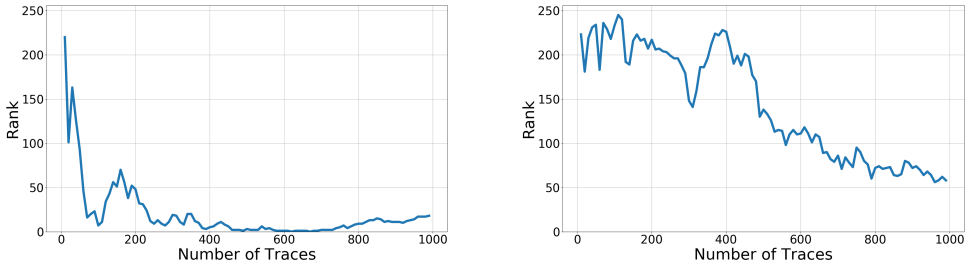


Figure 6: Results for (a) Key rank for the ASCAD dataset having 38 400 profiling traces trained using MLP, and (b) Key rank for the ASCAD dataset having 38 400 profiling traces trained using ASCAD-CNN2

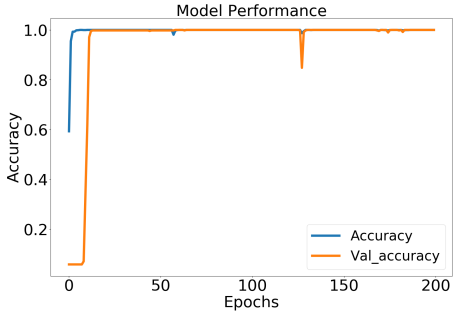


Figure 7: Results for training and validation accuracy for the ECC dataset having 2 400 profiling traces

5.4.2 Analysis on Real and Generated traces Dataset with the Maximum and Minimum Convergence

For the maximum convergence analysis, the training dataset consists of an equal proportion of the real traces and the artificially generated fake traces, that is, $n_r = 150$ and $n_g = 150$. Fake leakage signals are generated for the epochs during which the cGAN-Siamese model achieves maximum convergence. For the minimum convergence analysis, we combined the real traces with the artificially generated traces in equal proportion, the same as for the maximum convergence case. However, traces are collected for the epochs during which the GAN model showed minimum convergence.

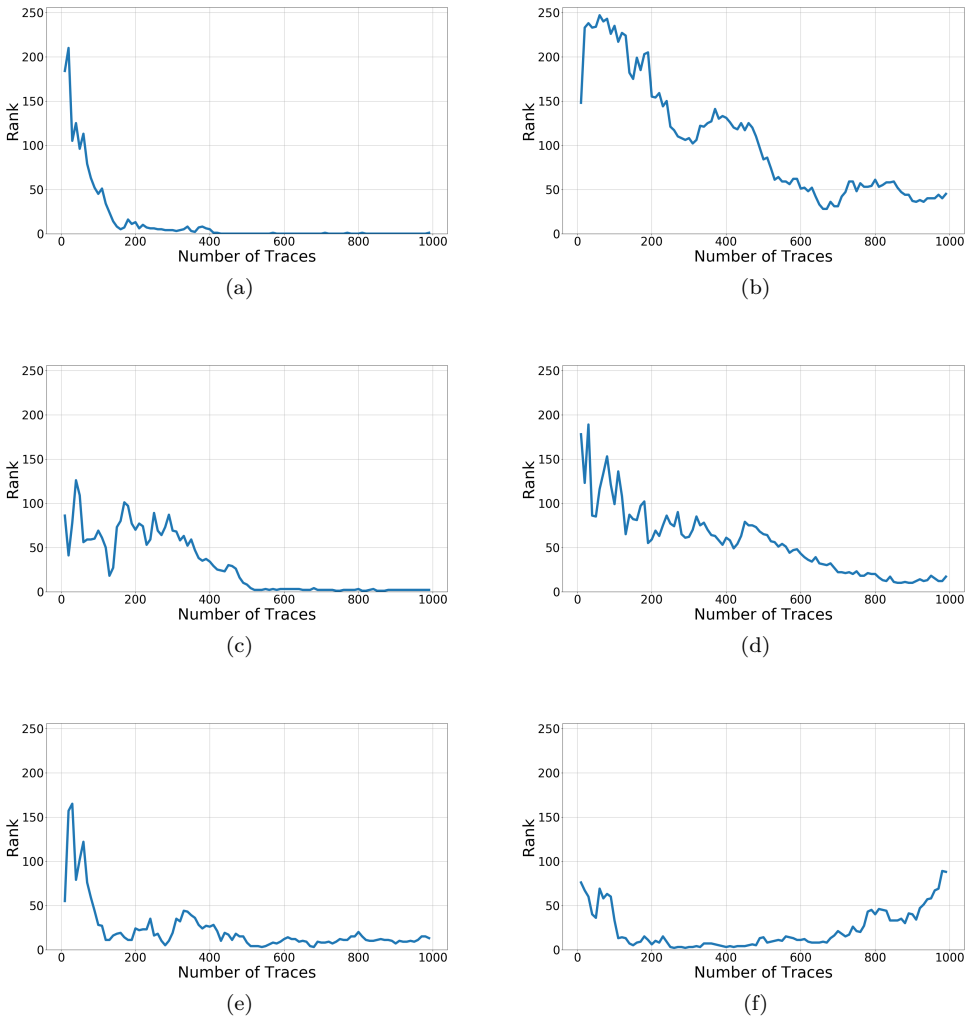
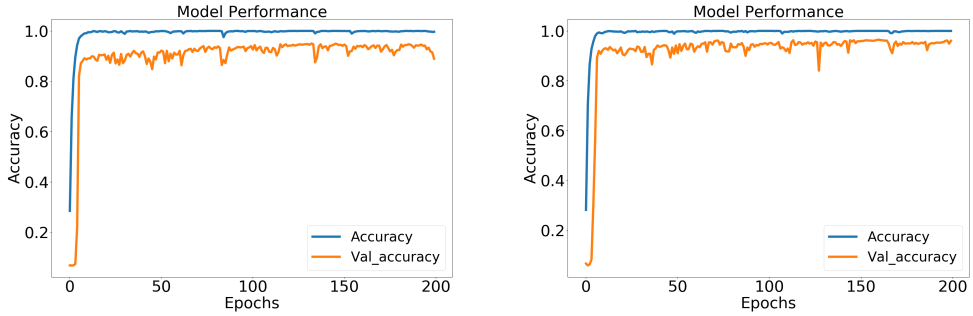


Figure 8: Key rank for the ASCAD dataset for (a) Maximum convergence $cGAN - Siamese\ Model\ B$ using MLP, (b) Minimum convergence $cGAN - Siamese\ Model\ B$ using MLP, (c) Maximum convergence $cGAN - Siamese\ Model\ B$ using ASCAD-CNN1, (d) Minimum convergence $cGAN - Siamese\ Model\ B$ using ASCAD-CNN1, (e) Maximum convergence $cGAN - Siamese\ Model\ B$ using ASCAD-CNN2, (f) Minimum convergence $cGAN - Siamese\ Model\ B$ using ASCAD-CNN2

Figure 8 shows the key rank for both maximum and minimum convergence for all three

DL-SCA models and the ASCAD dataset. We notice that generating fake traces from the maximum convergence point has a significant impact on key rank. The maximum convergence is achieved around 100-150 epochs for $Siamese - cGAN\ Model\ B$. Hence, data traces are generated around those epochs for analysis. It is observed that with the generated traces, all models (MLP, ASCAD-CNN1, and ASCAD-CNN2) gave the best performance, and the secret key can be obtained efficiently. *Thus, we can conclude that the artificially generated traces contain significant information that improved the ML-SCA performance.*



(a) (b)

Figure 9: Training and validation accuracy on the ECC dataset collected from (a) Maximum convergence point, (b) Minimum convergence point

The minimum convergence is observed around initial epochs, so artificial 150 traces per class are generated around this point and are combined with the real 150 traces to train the DL-SCA model. Observe that with minimum convergence, the DL-SCA attack model shows key rank is not stable, as it reaches zero in certain cases and starts increasing again as can be seen from Figures 8*b* and 8*f*. However, for Figure 8*d* (trained with ASCAD-CNN1), it appears to reach a key rank of zero near 1 000 traces, so more investigation is required to assess this case properly.

Figure 5 shows the GAN convergence curve for the ECC dataset. The model trained with the proposed $Siamese - cGAN\ Model\ B$ shows a better convergence as compared to the other three cGAN models' losses. The traces with the maximum convergence analysis are generated around epoch 700-1 000 (70-100 scaled in Figure 5), and traces for the minimum convergence are generated around 20-30 epochs. The performance accuracy is high after adding artificial traces. The trained model with the artificial traces generated with the convergent model (Figure 9a) shows accuracy greater than 97% using CNN, which is nearly the same accuracy as achieved on the real traces. However, the trained model with artificial traces, with the least convergent model, shows around 90% accuracy. While this performance is still good, we note that it cannot be compared with the performance on the real dataset. *Hence, the maximum convergent model generates the artificial traces that are more alike in characteristics to the real leakage traces and helps in training an efficient model for profiling side-channel analysis.*

5.5 Discussion

Based on the conducted experiments, we can draw some more general observations:

- GANs (more precisely, conditional GANs) represent a viable option for constructing synthetic side-channel traces. To improve the performance of GANs, it is beneficial to use deeper architectures and convolutional layers.

- A combination of a Siamese network and a cGAN can further improve the quality of the obtained synthetic examples.
- The procedure of generating fake traces is efficient and can generate hundreds of traces in a matter of minutes.
- It is important to monitor the GAN loss carefully and use the model that minimizes it when crafting synthetic examples.
- The combination of fake and real traces performs well regardless of the applied machine learning-based model. What is more, we see that fake traces can even improve attack performance.
- It is possible to construct synthetic examples for various cryptographic implementations with similar success, i.e., this technique is not limited to a specific cryptographic implementation.

6 Conclusions and Future Work

A dataset of leakage traces with an insufficient number of traces can pose a significant problem for accurate attack modeling using machine learning-based side-channel analysis. For such scenarios, data augmentation using a Generative Adversarial Network (GAN) can be useful. In this work, we proposed a layered architecture (Siamese-cGAN) based on cGAN and Siamese network that presents a well-convergent model to generate the artificial traces that have similar characteristics to the real traces. We performed two sets of analyses. In the first set of analyses, we run the experiments and present a visual comparative analysis between the performance of the proposed model and the existing cGAN based models for leakage signal generation. For this analysis, two neural network-based models (MLP and CNN) have been used for modeling the generator and the discriminator networks. The best model is selected based on the comparative analysis. In the second set of analyses, we evaluated the generated fake dataset by applying the actual machine learning-based side-channel attack on the leakage datasets from the exiting AES and ECC algorithm implementations. Four state-of-the-art neural network architectures (one MLP and three CNNs) are used for this evaluation. We also provided a comparative analysis of the dataset's performance consisting of data generated from the well-convergent network and data generated from the non-convergent network.

The proposed Siamese-cGAN model performed better than the existing simple cGAN models for both existing symmetric and asymmetric datasets. The quantitative analysis results show that the well-converged Siamese-cGAN network produces fake leakage traces similar to the real collected traces. Hence, they enable a better machine learning-based model for side-channel attacks. We also observed that the CNN-trained models performed better than MLP for the key recovery. We conclude that leakage traces/instances with significant contributing features can be efficiently generated. However, selecting a fully converging model might vary for each cryptographic algorithm.

In future work, it would be interesting to explore what are the limits of our approach from the perspective of the number of synthetic traces. Indeed, our results indicate that 150 traces per class are more than sufficient to construct convincing synthetic data. Understanding what the minimum required number of traces is would allow its proper evaluation of the viability. Furthermore, we used only the intermediate value leakage model, which results in more classes but also balanced measurements per class. We plan to evaluate different leakage models like the Hamming weight model that results in imbalanced data but also have fewer classes. Finally, it would be interesting to explore if the GAN-based approach could generate measurements that would help reduce the effect of portability for deep learning-based SCA [BCH+20]. There, instead of synthesizing measurements that resemble the traces from the profiling device, the task would be to generate measurements resembling those from the device under attack.

References

[BCH+20] Shivam Bhasin, Anupam Chattopadhyay, Annelie Heuser, Dirmanto Jap, Stjepan Picek, and Ritu Ranjan Shrivastwa. Mind the portability: A warriors guide through realistic profiled side-channel analysis. In *27th Annual Network and Distributed System Security Symposium, NDSS 2020, San Diego, California, USA, February 23-26, 2020*. The Internet Society, 2020.

[BCP+14] Lejla Batina, Łukasz Chmielewski, Louiza Papachristodoulou, Peter Schwabe, and Michael Tunstall. Online template attacks. In Debdeep Mukhopadhyay Willi Meier, editor, *Progress in Cryptology - INDOCRYPT 2014 - 15th International Conference on Cryptology in India, New Delhi, India, December 14-17, 2014, Proceedings*, volume 8885 of *LNCS*, pages 21–36. Springer, 2014. http://cryptojedi.org/papers/#ota.

[BDS18] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. *CoRR*, abs/1809.11096, 2018.

[BPS+20] Ryad Benadjila, Emmanuel Prouff, Rémi Strullu, Eleonora Cagli, and Cécile Dumas. Deep learning for side-channel analysis and introduction to ASCAD database. *J. Cryptogr. Eng.*, 10(2):163–188, 2020.

[CCC+19] Mathieu Carbone, Vincent Conin, Marie-Angela Cornélie, François Dassance, Guillaume Dufresne, Cécile Dumas, Emmanuel Prouff, and Alexandre Venelli. Deep learning to evaluate secure RSA implementations. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2019(2):132–161, Feb. 2019.

[CDH+16] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 2172–2180. Curran Associates, Inc., 2016.

[CDP17] Eleonora Cagli, Cécile Dumas, and Emmanuel Prouff. Convolutional neural networks with data augmentation against jitter-based countermeasures. In Wieland Fischer and Naofumi Homma, editors, *Cryptographic Hardware and Embedded Systems – CHES 2017*, pages 45–68, Cham, 2017. Springer International Publishing.

[Chi21] Davide Chicco. *Siamese Neural Networks: An Overview*, pages 73–94. Springer US, New York, NY, 2021.

[CRR03] Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi. Template attacks. In Burton S. Kaliski, çetin K. Koç, and Christof Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2002*, pages 13–28, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[ecc19] Database for eddsa., 2019. https://github.com/leoweissbart/MachineLearningBasedSideChannelAttackonEdDSA.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[GPAM+14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proceedings of the 27th International Conference on*

Neural Information Processing Systems - Volume 2, NIPS'14, page 2672–2680, Cambridge, MA, USA, 2014. MIT Press.

[HLSC18] Chih-Chung Hsu, Chia-Wen Lin, Weng-Tai Su, and Gene Cheung. Sigan: Siamese generative adversarial network for identity-preserving face hallucination. *CoRR*, abs/1807.08370, 2018.

[HMH+12] Johann Heyszl, Stefan Mangard, Benedikt Heinz, Frederic Stumpf, and Georg Sigl. Localized electromagnetic analysis of cryptographic implementations. In Orr Dunkelman, editor, *Topics in Cryptology – CT-RSA 2012*, volume 7178 of *LNCS*, pages 231–244. Springer, 2012.

[KAHK17] Naveen Kodali, Jacob Abernethy, James Hays, and Zsolt Kira. On convergence and stability of gans, 2017.

[KPH+19] Jaehun Kim, Stjepan Picek, Annelie Heuser, Shivam Bhasin, and Alan Hanjalic. Make some noise. unleashing the power of convolutional neural networks for profiled side-channel analysis. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, pages 148–179, 2019.

[KZS15] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. 2015.

[LBM14] Liran Lerman, Gianluca Bontempi, and Olivier Markowitch. Power analysis attack: An approach based on machine learning. *Int. J. Appl. Cryptol.*, 3(2):97–115, June 2014.

[LBOM12] Yann A. LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. *Efficient BackProp*, pages 9–48. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[LHY+20] Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang, and Arun Mallya. Generative adversarial networks for image and video synthesis: Algorithms and applications. *CoRR*, abs/2008.02793, 2020.

[LSP21] Maria Leyva-Vallina, Nicola Strisciuglio, and Nicolai Petkov. Generalized contrastive optimization of siamese networks for place recognition. *CoRR*, abs/2103.06638, 2021.

[LZW+21] Zhimin Luo, Mengce Zheng, Ping Wang, Minhui Jin, Jiajia Zhang, and Honggang Hu. Towards strengthening deep learning-based side channel attacks with mixup. Cryptology ePrint Archive, Report 2021/312, 2021. https://eprint.iacr.org/2021/312.

[MMKA18] Naila Mukhtar, Ali Mehrabi, Yinan Kong, and Ashiq Anjum. Machine-learning-based side-channel evaluation of elliptic-curve cryptographic fpga processor. *Applied Sciences*, 9:64, 12 2018.

[MO08] Marcel Medwed and Elisabeth Oswald. Template attacks on ECDSA. In Kyo-Il Chung, Kiwook Sohn, and Moti Yung, editors, *Information Security Applications*, volume 5379 of *LNCS*, pages 14–27. Springer, 2008. https://eprint.iacr.org/2008/081/.

[MO14] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.

[MPP16] Houssem Maghrebi, Thibault Portigliatti, and Emmanuel Prouff. Breaking cryptographic implementations using deep learning techniques. In *International Conference on Security, Privacy, and Applied Cryptography Engineering*, pages 3–26. Springer, 2016.

[OPB16] Elif Özgen, Louiza Papachristodoulou, and Lejla Batina. Classification Algorithms for Template Matching. In *IEEE International Symposium on Hardware Oriented Security and Trust, HOST 2016, McLean, VA, USA, 2016 (to appear).*, 2016.

[PCBP21] Guilherme Perin, Lukasz Chmielewski, Lejla Batina, and Stjepan Picek. Keep it unsupervised: Horizontal attacks meet deep learning. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2021(1):343–372, 2021.

[PHG19] S. Picek, A. Heuser, and S. Guilley. Profiling side-channel analysis in the restricted attacker framework. *IACR Cryptol. ePrint Arch.*, 2019:168, 2019.

[PHJ+18] Stjepan Picek, Annelie Heuser, Alan Jovic, Shivam Bhasin, and Francesco Regazzoni. The curse of class imbalance and conflicting metrics with machine learning for side-channel evaluations. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2019(1):209–237, Nov. 2018.

[RMC16] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[RWPP21] Jorai Rijsdijk, Lichao Wu, Guilherme Perin, and Stjepan Picek. Reinforcement learning for hyperparameter tuning in deep learning-based side-channel analysis. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2021(3):677–707, Jul. 2021.

[SGZ+16] Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *CoRR*, abs/1606.03498, 2016.

[SK19] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), July 2019.

[SLP05] Werner Schindler, Kerstin Lemke, and Christof Paar. A stochastic model for differential side channel cryptanalysis. In Josyula R. Rao and Berk Sunar, editors, *Cryptographic Hardware and Embedded Systems – CHES 2005*, pages 30–46, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[WAGP20] Lennert Wouters, Victor Arribas, Benedikt Gierlichs, and Bart Preneel. Revisiting a methodology for efficient cnn architectures in profiling attacks. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2020(3):147–168, Jun. 2020.

[WCL+20] Ping Wang, Ping Chen, Zhimin Luo, Gaofeng Dong, Mengce Zheng, Nenghai Yu, and Honggang Hu. Enhancing the performance of practical profiling side-channel attacks using conditional generative adversarial networks, 2020.

[WCPB20] Léo Weissbart, Łukasz Chmielewski, Stjepan Picek, and Lejla Batina. Systematic side-channel analysis of curve25519 with machine learning. *Journal of Hardware and Systems Security*, 4(4):314–328, October 2020.

[WHL+19] Chenshen Wu, Luis Herranz, Xialei Liu, Yaxing Wang, Joost van de Weijer, and Bogdan Raducanu. Memory replay gans: learning to generate images from new categories without forgetting, 2019.

[WOS14] Carolyn Whitnall, Elisabeth Oswald, and François-Xavier Standaert. The myth of generic dpa... and the magic of learning. In Josh Benaloh, editor, *Topics in Cryptology – CT-RSA 2014*, pages 183–205, Cham, 2014. Springer International Publishing.

[WP20] Lichao Wu and Stjepan Picek. Remove some noise: On pre-processing of side-channel measurements with autoencoders. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2020(4):389–415, Aug. 2020.

[WPB19] Leo Weissbart, Stjepan Picek, and Lejla Batina. One Trace Is All It Takes: Machine Learning-Based Side-Channel Attack on EdDSA. In Shivam Bhasin, Avi Mendelson, and Mridul Nandi, editors, *Security, Privacy, and Applied Cryptography Engineering - 9th International Conference, SPACE 2019, Gandhinagar, India, December 3-7, 2019, Proceedings*, volume 11947 of *Lecture Notes in Computer Science*, pages 86–105. Springer, 2019.

[ZBHV19] Gabriel Zaid, Lilian Bossuet, Amaury Habrard, and Alexandre Venelli. Methodology for efficient cnn architectures in profiling attacks. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2020(1):1–36, Nov. 2019.

[ZBHV21] Gabriel Zaid, Lilian Bossuet, Amaury Habrard, and Alexandre Venelli. Efficiency through diversity in ensemble models applied to side-channel attacks: – a case study on public-key algorithms –. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2021(3):60–96, Jul. 2021.

[ZXL+17] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5908–5916, 2017.