Power of Prediction: Advantages of Deep Learning Modeling as Replacement for Traditional PUF CRP Enrollment

Amir Ali Pour, David Hely, Vincent Beroulle, Giorgio Di Natale

Power of Prediction: Advantages of Deep Learning Modeling as Replacement for Traditional PUF CRP Enrollment

Amir Alipour*, David Hely*, Vincent Beroulle*, Giorgio Di Natale**
*Univ. Grenoble Alpes, Grenoble INP, LCIS F-26000 Valence
**Univ. Grenoble Alpes, Grenoble INP, CNRS, TIMA 38000 Grenoble, France
*{firstname.lastname}@lcis.grenoble-inp.fr,
**giorgio.di-natale@univ-grenoble-alpes.fr

***Abstract*—Physically Unclonable Functions (PUFs) have been addressed nowadays as a potential solution to improve the security in authentication and encryption process of Cyber Physical Systems. The research on PUF is actively growing due to its potential of being secure, easily implementable and expandable, using considerably less energy. Depending on the application, the size of a PUF Challenge-Response Pair (CRP) set can be different. Applications that demand frequent use of PUF, require enrollment of a very large set of CRPs per PUF unit. This for resource constraint ecosystems, especially in IoT edge device authentication, can become a challenge. In this work our aim is to put spotlight on the prediction power of trained Neural Network models, constructed using Deep Learning techniques, to replace the traditional usage of CRP set tables. Potentially, the trained Neural Networks for that purpose require very small set of CRPs per PUF unit for enrollment (training) and can predict a considerably larger set of CRPs for a given PUF. Different implementation of Neural Network based PUF authentication potentially exist, to which we point out and explain the pros and cons. In addition, we will also discuss other benefits of conducting Deep Learning for enrollment, such as being resilient to instability of PUF CRP, and a Deep Learning based PUF enrollment can be utilized to implement robust key generation for encryption.**

Keywords—Physically unclonable Function (PUF), Deep Learning, Challenge Response Pair (CRP), Neural Network (NN), Enrollment

I. INTRODUCTION

The popularity of Physically Unclonable Functions (PUFs) has been reflected in the abundance contributions in academia and companies conducting this technology to provide a robust secure infrastructure for authentication [1], [2] and/or key generation [3] .

PUF in digital electrical and electronic devices refers to functional components, either intrinsic or implemented, on hardware, which have exactly the same architecture, but output differently, compared to one another, based on the same input [4]. The characteristic of PUF are based on a Challenge-Response mechanism. PUF architectures usually have a vector input, which is the challenge, and has an output vector, which is the response of the PUF. For a given Challenge, there is of course a Response. Noting that it is expected in theory that the Response to a Challenge is unique per PUF unit. By this definition then, the Challenge-Response Pair (CRP) can be used to authenticate a PUF enabled device [4].

This work is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02).

To employ PUF, a designer should be certain that the intrinsic behavior of the PUF, per PUF unit, satisfies their application. To do so, parameters such randomness, uniqueness, diffuseness and the reliability of the collected CRP corresponding implemented PUF hardware should be studied[5], [6]. Commonly, the PUF Challenge space is large enough to build any CRP, there exists a considerable chance that a Response for a given Challenge is unstable. Therefore, the selection of CRPs for an optimal CRP set per PUF unit should be conducted delicately, to uphold the optimal value for the mention parameters. This happens when the two physical quantities (that are compared to generate the response) have very similar values, and metastability happens in the comparator. These responses are not useful in the protocols for authentication, and even less for generating secret keys. Therefore, each CRP is tested multiple times or dedicated procedures must be used [7], [8] to prove stability. And finally, the unstable CRPs are not considered in the final CRP set.

The number of CRPs, measured and stored, is also an important factor, which differs based on the type of PUF, and also the application using the PUF [9], [10]. Generally, there are two types of PUFs, strong PUFs and weak PUFs. The main difference between these two types is the number of stored CRP, wherein for strong PUF, the number of stored CRP is considerably higher. There are many applications existing that rely on using Strong PUFs, and in such applications preserving the identity of PUF units is a common concern. For instance, it has been discussed in [11] that reusing the same Challenge-Response Pair, raises the likelihood of attacks such as Replay attack and Impersonation attack. Thus, to prevent this, PUF CRPs in such systems are preferred to be used only once.

The case of measuring large number of CRPs is also relevant for characterizing the behavior of PUF units under different environmental situations, for which the measurement of CRP subsets under different environmental states, such as varied temperature, or the effect of aging, is conducted, which then leads to construction of an even larger CRP dataset [12], [13]. Measuring this amount of knowledge could be a challenge, especially for cases in which PUF units targeted for authentication are highly populated. Moreover, another challenge could be with the storage of the CRP sets and managing them during authentication [14]. Especially if the population of PUF devices for authentication is high and demanding on CRP extraction, either for authentication or encryption purposes is frequent and high.

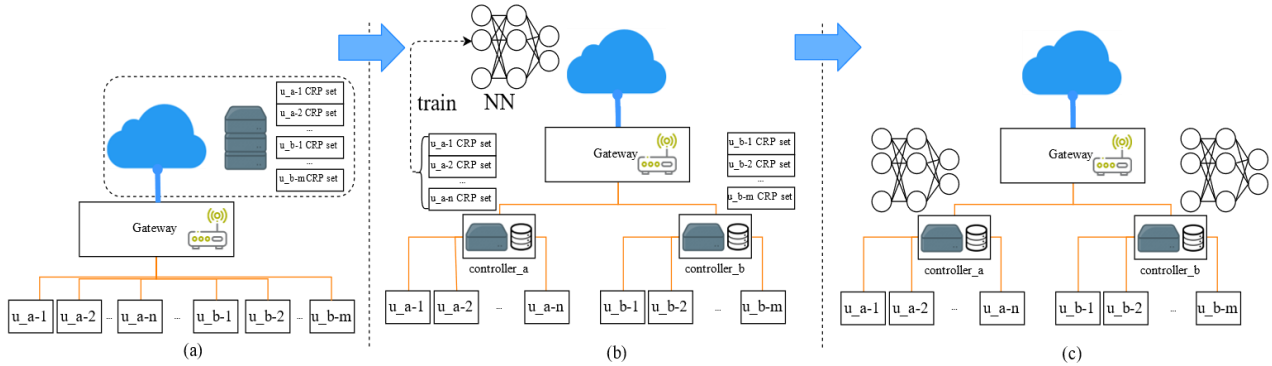

Fig. 1. Examples of schematics of two Architecturally Hierarchical IoT system with edge devices and authenticators, (a) with central authentication and (b) with distributed authentication. (c) is distributed authentication where Neural Networks (NN) instead of CRP tables are used for authentication.

Fig. 1 is an example of an IoT, where authentication of edge devices is distributed rather centralized. And it is ever closer to the edge of the ecosystem. Part of the definition of such topology is that the authenticators themselves are defined relatively resource constraint. Therefore, needing a careful usage of energy and storage on memory.

While the CRP storage of PUF is not mentioned to be of a concern in early stages of the development of the technology, with rapid growth of population of devices relying on PUF technology for their authentication and encryption purposes, especially with increasing complexity of PUF architecture, the CRP dataset enrollment and storage will face its own challenges. We believe at that time the challenges would be some as we mentioned in the following, which in turn lead to increase in time of CRP dataset enrollment and size of CRP datasets:

- Increased complexity of PUF architecture
- Inclusion of different operational states
- Aiming for a reliable, hence stable CRP set

To prevent such challenges to take effect, a new technique should be used to replace traditional PUF enrollment. In other words, it could be potentially more effective to construct a data-model of the PUF unit, instead of a huge CRP dataset (which we refer to as M set). The reference model is built using a subset of the CRP dataset with all possible CRPs of a given PUF unit, which we refer to as the *M* set. However, the reference model then can reflect the larger set, possibly all within the *M* set. To consider this reference model an eligible substitution, we assume that the construction of it requires considerably smaller number of CRPs, compared to that of a CRP set needed according to a given application which utilizes the PUF. Also, we should make sure that deploying this model in an actual ecosystem is also practically acceptable, in terms of computation overhead and memory storage.

To this end, we propose using Deep Learning modeling as an efficient modeling technique. We made this proposal based on our observation of number of recent works that conducted Deep Learning modeling to clone PUFs [15]–[17]. In most of these works, DL models are assumed to be trained by an adversary using leaked information, and so far, the models have achieved high accuracy in cloning the PUF.

However, in our case, if DL modeling is to be conducted, further optimization is required to reach to the highest accuracy. That, however, can be achieved in our case, since we are on the manufacturers side thus having access to enough amount of knowledge to further increase the optimality of our DL models.

In this work we will discuss the potential of Deep Learning Modeling used for PUF CRP enrollment. We will discuss different implementations of PUF, such as SRAM based PUF and Arbiter PUFs, as well as their application and how Deep Learning can potentially be used to facilitate enrollment for each architecture, as well as other benefits of enrollment based on Neural network models.

The rest of this work will be as followed. Section II presents the preliminaries of this work. In section III we will talk about how deep learning can in various ways be conducted to model PUF on an authenticator system, and what would be the benefits of doing that compared to the traditional way of storing CRP tables. In section IV we will discuss also the challenges we see forward with our proposed methods and how we also think of solutions that can overcome these challenges.

II. Perliminaries

Section II.A describes the basic structure of Deep Learning briefly, with a deeper look into the Neural Network architectures which are potentially fit for our context, i.e., to be used for modeling PUF. Section II.B presents some of the commonly used and discussed PUF architectures.

A. *Deep Learning models*

Deep Learning is a subset of Machine learning and is defined based on the structure and behavior of Artificial Neural Networks (ANNs). It's been proven widely that ANNs are capable models for applications requiring complex data classification [18]. The field of ANN models has been generously maturated. Common architectures such as Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Auto Encoders (AEs), etc. have been defined for more than 2 decades and have been so far conducted in various industrial, academical, and medical applications. The basic block of all ANNs is the Neuron. And the idea of ANNs is the composition of a population of Neurons, in layers, which is referred to as Neural Network. Fig. 2. shows the basic structure of one Neuron.

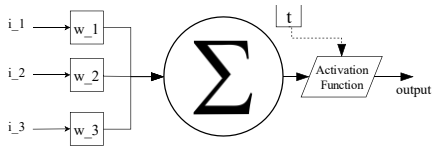

Fig. 2. Diagram showing the functionality of a single neuron.

An Artificial Neuron model is a computational unit, which receives a vector of input *I* of size *n*, and produces *v* which is the dot product of *I* to a vector of weight values *w* of size *n*. The value of *v* is then passed to a function *f ()* called the activation function. *f (v)* will then define the output of the Neuron based the value of *v*. Figure_2 shows a single perceptron. Common Activation functions are ReLu, Sigmoid, Step Function, etc. [18].

A Neural Network is a composition of a group of Neurons, ordered in layers. Various compositions are possible to make different Neural Networks. For instance, CNN is used for feature extraction in images with high dimentiality [19], as well as object detection, etc. Auto Encoder is used for regeneration purposes from noisy data. While in (c), the Recurrent Neural Network is optimal for audio processing, and generally robust for feature extraction with respect to temporal situation.

B. *Common PUF architectures*

The definition of Physically Unclonable Function has been coined for almost two decades [20], [21]. Since then, several architectures, as well as many variations of each has been proposed. SRAM based PUF for instance, is found quite feasible in implementation, due to using the most commonly found hardware component in almost all digital devices, the internal memory [22]. This intrinsic type of PUF is found potential for both authentication and encryption key generation. Another example of a common PUF architecture, is the family of Arbiter PUFs (APUFs) [6]. APUFs are quite feasible in hardware implementation, such as on FPGA. They are low cost, energy efficient, and potentially known as strong PUFs. Yet they are faced with the challenge of being susceptible initially to modeling attacks. On the other hand, there are Ring Oscillator (RO) PUF, which compared to APUFs, are more resilient to modeling attacks, due to their relatively more complicated design, but consecutively also, demand more energy for computation [13]. There exists also other architecturally innovative PUFs such as PUFs based on Neural Networks, but to keep this work compact and relative to the context, we decided to move forward with two of the former mentioned architectures. In the coming section, we will discuss the different possible implementation of Deep Learning modeling for SRAM based PUF and Arbiter PUFs enrollment, respectively, the pros and cons, as well as the benefits it features for PUF enrollment.

III. Implementation

The implementation of an ideally working DL based PUF enrollment can differ relative to the PUF architecture. For instance, let's consider the family of Arbiter PUFs. Quite number of works has been conducted describing DL modeling attacks against this family of PUFs. Wherein the models are commonly architected to replicate a PUF unit behavior, based on a set of leaked CRPs. The results are consequently binary classifiers which predict a response for a given challenge, to replicate a single block of an Arbiter PUF. Following this modeling technique for authentication, an obvious proposal would be to train compact NN models per PUF unit. While the potential exists for predicting a large set of CRPs with a trained NN which requiring only a small set of CRPs to be trained highly accurate, on the other hand, the issue with storage may remain active. Since a model for a PUF unit is saved on an authenticator server. One can say, if the size of the stored trained model on server is considerably smaller than a fixed CRP LUT with the same capacity, then the tradeoff is tolerable.

A better implementation of this technique however can potentially be realized. That would be in following of a recent work by Karimian et al. in [14] which has taken the approach towards using DL modeling for PUF enrollment. This work uses CNNs to classify PUF units directly, by reading their Response, which is a large 2d matrix of all cells of the embedded DRAM. Noting that the response comes from a DRAM based PUF. Therefore, any feature extracted from the response matrix, solely belong to the physically unique characteristic of the DRAM, during training the CNN. Conducting the same method for Arbiter based PUFs will not highly likely result the same way, if classification is based on one CRP. That is since the population of Responses are not enough to extract specific features from. However, to mitigate this we have a proposal.

Our proposal is to implement a classification of PUF units, but taking in a range of CRP, which we refer to as ranged-CRP which is ordered in a 2d matrix. Fig. 3 shows an example of a ranged-CRP frame sent from a PUF unit.

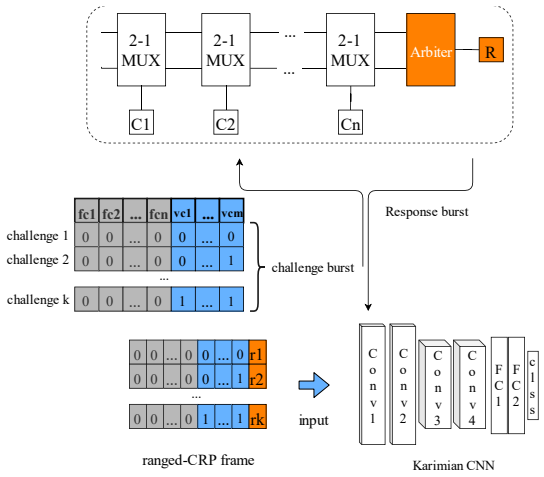

Fig. 3. Implementation of enrollment/authentication using ranged-CRP frame and Karimian CNN [] for classification of PUF units.

The reason why we stress on a range of CRP, rather one, is that in one CRP, any characteristic drawn from it, would majorly be weighed on the Challenge section of the CRP, since the response is only one bit. This would in turn, increase the training process, while needing also a large CRP set on the other hand, to achieve a high accuracy in classification. On the other hand, a ranged-CRP frame can more likely reflect potential characteristics to a PUF unit from the obvious variations in the frame. Since these variations are now comprising multiple response bits, as well as some of the challenge bits. While a large chunk of the frame belongs to the fixed challenge bits, which during training is considered as white space, from machine

learning terms, and consequently less relative to effect classification.

The realization of this idea, is potential not just to reduce the size of stored reference DL model on the authenticator server, compared to the former proposal, but also equally potential to require only a small CRP set to train a highly accurate NN model. We speculate also that the ranged-CRP frame has a lesser complexity compared to that of a 2d memory map extracted from a DRAM, following the work of Karimian et al at [14]. Moreover, we think that even a Multi-Layer Perceptron (MLP) can be used for modeling, which compared to CNNs, are easier to train and less resource demanding.

Following the work of Karimian et al [14], the realization of DL based PUF enrollment has been discussed to be vulnerable to man-in-the middle attacks. That is, as we assume, to be due to the fact that at each time, the same memory frame is sent for authentication, leaving the chance for attackers to read enough variations of the same frame during the transmission, and gradually be capable of model-building a clone of the memory. To mitigate this issue, we propose sending positionally different chunks of memory for classification from a PUF unit to server at each time. For this, we propose the use of challenge vector as address bit to address a chunk of memory and send it to server. In complementary, what is to be addressed further is the questions of how classification can be done if we are not taking in input the entire map of memory, but just a chunk of it. For that we propose the following.

During the enrollment of DRAM PUF units, an entire map of the DRAM is extracted from each PUF unit, formed in a 2d-matrix, and sent to a CNN for training. Note the classification is to identify a PUF unit from its memory map. That is, similar to the work of Karimian et al in[14]. In addition to CNN, an Auto Encoder Neural Network [23] can be used to regenerate a reference memory map per PUF unit. The reference map is also saved on the server.

Furthermore, during authentication, a chunk of memory, which is addressed by a challenge vector coming from server to PUF unit, will be sent from PUF unit to server as response. On the server side, after receiving the response for a given challenge, we multiply or replace each corresponding bit of the chunk on the reference map model, with the bits on the received response. The result is a new map which then is sent to CNN for classification. If classified correctly, the response corresponding the challenge is valid, which means it belongs to the PUF unit. Thus, the PUF is authenticated. Additionally, the potential to only train and classify with the addressed chunk of the memory, without using the reference Image, can be further studied to save more storage.

For this method, we need to carefully consider the size of the memory chunks are extracted. First, we should consider that the memory chunk should not be too small so that its effect is not that impactful to the classification. We also don't want to make it large enough so that attackers who collect the large size memory chunks are potentially capable to build their clone memory model. We also want the memory chunk to be efficiently sized so that for cases of using them only one time, or limited times, we have the capacity, by the proportion of the size of the entire DRAM memory to the size of the extracted memory chunk, that is resulting a large number of possible challenges. This limitation of course from security point of view is beneficial since it leaves very few clues of the entire memory behavior per memory chunk for attackers to exploit on. But the chunk is large enough that, for instance, if it comes from a clone device, it has the impact to disturb the classification. In addition, the definition of the size of the memory chunk is also relative to how frequent an application would want to use the PUF. Fig. 4. shows the implementation of the idea.

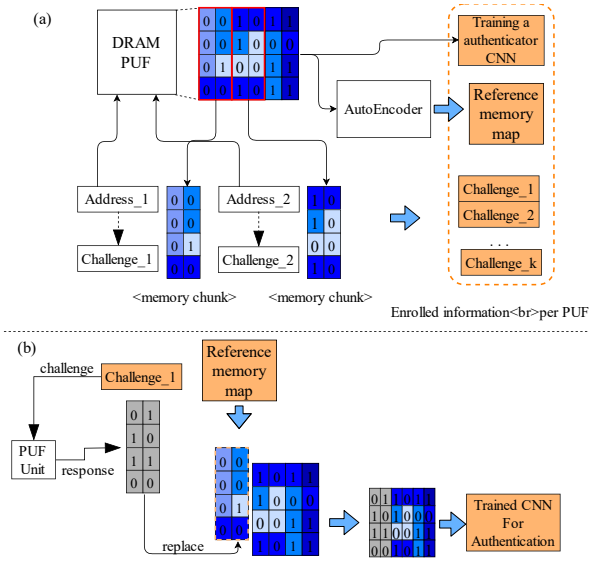

Fig. 4. Implementation of re-mapping response chunk on reference memory map and classifying the new map. Also, the training of CNN based on self-filtered memory frame. (a) is the Enrollment process of a PUF unit, and (b) shows the authentication of a puf unit with the proposed method.

IV. Discussion

In this section we will discuss the existing challenges we found impactful on Deep Learning Modeling for PUF enrollment. The base of the proposal to use DL modeling as a new method to enroll PUF, is training or optimizing a Neural Network model towards accurately predicting all possible responses a PUF model per challenge, or either directly classifying a PUF unit based on its response. In the following, we will discuss about the parameters that are important in providing a robust CRP set for DL based PUF enrollment.

A. Requirements of a DL based PUF CRP set

In practice, during authentication, as mentioned, a Neural Network is trained by a small fraction of CRPs from this space. Different factures here should be taken into consideration. One for instance is, the proportion of set sizes of (Training_set + Validation_set) / (Desired Predictable_set). While training accuracy based on the Training set assures the Neural Network is trained to optimally classify CRPs of the training set, or classifying a PUF unit, high Validation accuracy on the other hand, determines that the network can efficiently predict CRPs, or the correctness of CRPs, outside of the Training set. This, however, is a relative measurement. Potentially, Randomness of the (Training set + Validation set) plays an important role on definition that how far an addressed accurately trained; of course, based on validation accuracy, can predict CRPs outside Trainig_Validation set.

Conclusions

In this work we discussed that this new solution can greatly decrease the number of CRPs during enrollment, due to the prediction power of neural networks. We showed that initially, deep learning modeling has been studied and conducted as an adversary model to clone PUF architectures, and we discussed that highly accurate models can be trained based on only leaked information of PUF, which has been already practically proven in previous works. We then mentioned the work of Karimian et al [14], as being one of the first and resent works addressing the utilization deep learning for DRAM based PUF enrollment. Following their work, we also proposed how the method can be utilized for another type of PUF architecture, the family of Arbiter based PUF. We proposed a technique to extract more physically unique features for a PUF unit, during authentication, so to take maximum benefit of DL based classification. Furthermore, for the work of Karimian et al [14], in order to operate a less leaking transmission of challenge response during authentication, we proposed sending small parts of memory, thus if leaked, lead to no major clue for attackers to enable them to construct a clone model. We then showed how this modification can be adapted to Karimian' CNN based classification model. The future of this work will be for us to perform experiments on these proposed solutions and find the limits of their implementation, as well as practically proving the methods. In our future experiments, our goal is to include more than just the direct response of PUF, but also the environmental and temporal state of enrolled PUF units, in order to build a more robust model for authentication.

References

[1] A. Grigorescu, H. Boche, and R. F. Schaefer, "Robust PUF based authentication," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, Roma, Italy, 2015, pp. 1–6.

[2] W. Che, F. Saqib, and J. Plusquellic, "PUF-based authentication," in *2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Austin, TX, USA, 2015, pp. 337–344.

[3] A. R. Korenda, F. Afghah, and B. Cambou, "A Secret Key Generation Scheme for Internet of Things using Ternary-States ReRAM-based Physical Unclonable Functions," in *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Limassol, Cyprus, 2018, pp. 1261–1266.

[4] R. Maes and I. Verbauwhede, "Physically Unclonable Functions: A Study on the State of the Art and Future Research Directions," in *Towards Hardware-Intrinsic Security*, A.-R. Sadeghi and D. Naccache, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 3–37.

[5] F. Ganji, D. Forte, and J.-P. Seifert, "PUFmeter a Property Testing Tool for Assessing the Robustness of Physically Unclonable Functions to Machine Learning Attacks," *IEEE Access*, vol. 7, pp. 122513–122521, 2019.

[6] M. S. Alkatheiri, Y. Zhuang, M. Korobkov, and A. R. Sangi, "An experimental study of the state-of-the-art PUFs implemented on FPGAs," in *2017 IEEE Conference on Dependable and Secure Computing*, Taipei, Taiwan, 2017, pp. 174–180.

[7] E. Ioana Vatajelu, G. Di Natale, and P. Prinetto, "Towards a Highly Reliable SRAM-based PUFs," in *Proceedings of the 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2016, pp. 273–276.

[8] H. Martin, E.-I. Vatajelu, G. Di Natale, and O. Keren, "On the Reliability of the Ring Oscillator Physically Unclonable Functions," in *2019 IEEE 4th International Verification and Security Workshop (IVSW)*, Rhodes Island, Greece, 2019, pp. 25–30.

[9] A. Babaei and G. Schiele, "Physical Unclonable Functions in the Internet of Things: State of the Art and Open Challenges," *Sensors*, vol. 19, no. 14, p. 3208, Jul. 2019.

[10] Y. Komano, K. Ohta, K. Sakiyama, M. Iwamoto, and I. Verbauwhede, "Single-Round Pattern Matching Key Generation Using Physically Unclonable Function," *Security and Communication Networks*, 2019. [Online]. Available: https://www.hindawi.com/journals/scn/2019/1719585/. [Accessed: 05-Dec-2019].

[11] A. Braeken, "PUF Based Authentication Protocol for IoT," *Symmetry*, vol. 10, no. 8, p. 352, Aug. 2018.

[12] Md. T. Rahman, F. Rahman, D. Forte, and M. Tehranipoor, "An Aging-Resistant RO-PUF for Reliable Key Generation," *IEEE Trans. Emerg. Topics Comput.*, pp. 1–1, 2015.

[13] L.-Y. Chiou, C.-H. Wu, and P.-C. Wei, "A Reliable Delay-Based Physical Unclonable Function with Dark-Bit Avoidance," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, Sapporo, Japan, 2019, pp. 1–4.

[14] N. Karimian, F. Tehranipoor, N. Anagnostopoulos, and W. Yan, "DRAMNet: Authentication based on Physical Unique Features of DRAM Using Deep Convolutional Neural Networks," *arXiv:1902.09094 [cs]*, Feb. 2019.

[15] J.-Q. Huang, M. Zhu, B. Liu, and W. Ge, "Deep Learning Modeling Attack Analysis for Multiple FPGA-based APUF Protection Structures," in *2018 14th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, Qingdao, 2018, pp. 1–3.

[16] M. Khalafalla and C. Gebotys, "PUFs Deep Attacks: Enhanced modeling attacks using deep learning techniques to break the security of double arbiter PUFs," in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Florence, Italy, 2019, pp. 204–209.

[17] S.-J. Wang, Y.-S. Chen, and K. S.-M. Li, "Adversarial Attack against Modeling Attack on PUFs," in *Proceedings of the 56th Annual Design Automation Conference 2019 on - DAC '19*, Las Vegas, NV, USA, 2019, pp. 1–6.

[18] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, p. e00938, Nov. 2018.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[20] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical One-Way Functions," *Science*, vol. 297, no. 5589, pp. 2026–2030, Sep. 2002.

[21] B. Gassend, D. Clarke, M. van Dijk, and S. Devadas, "Silicon Physical Random Functions," in *Proceedings of the 9th ACM Conference on Computer and Communications Security*, New York, NY, USA, 2002, pp. 148–160.

[22] F. Afghah, B. Cambou, M. Abedini, and S. Zeadally, "A ReRAM Physically Unclonable Function (ReRAM PUF)-based Approach to Enhance Authentication Security in Software Defined Wireless Networks," *arXiv:1712.09916 [cs]*, Dec. 2017.

[23] F.-N. Yuan, L. Zhang, J.-T. Shi, X. Xia, and G. Li, "Theories and Applications of Auto-Encoder Neural Networks: A Literature Survey," *Jisuanji Xuebao/Chinese Journal of Computers*, vol. 42, pp. 203–230, Jan. 2019.