Enhancement of Security in Connection establishment for IoT Infrastructure through Adversarial Neural Cryptography using GANs

Basil Hanafi1, Mohammad Ubaidullah Bokhari2

{basilhanafi@gmail.com1, mubokhari@gmail.com2}

Department of Computer Science, Aligarh Muslim University, Aligarh – 202002 India1
Department of Computer Science, Aligarh Muslim University, Aligarh – 202002 India2

Abstract. As we are moving ahead in this Digital Age of Information, the need for fortification for Digital Assets and Data is also increasing day by day. Soon, in the near future, we will be able to see everything connected and controlled by a single large-scale network called the Internet. This whole infrastructure of connected various machines is coined with a term called the Internet of Things. With this versatile connection of Devices, Sensors, and machines, every digital asset an Individual can own will be at risk. The Increase in number of Assets will lead to increase in the number of Threats and Vulnerabilities which will result in a catastrophic future. To avoid such situations, we are in need of some dynamic strategies that can help individuals to protect themselves from such threats. Here we will be discussing, how these assets can be protected in an IoT environment by implementing Security through Adversarial Neural Cryptography using Generative Adversarial Networks. Fortification of this Cryptographic technique is giving momentous results when implemented in IoT devices during establishing connections, without human interference.

Keywords: Adversarial Neural Cryptography, Generative Adversarial Networks, Security, Internet of Things

1. Introduction

Internet of Things is an environment in which several nodes of devices are connected throughout the internet for the primary cause to increase the productivity of any Individual or an Organization. The potential of increasing productivity of devices among IoT infrastructure or an organization lies within the usage and connectivity of the Devices. More usage will lead to more piling up of digital assets and for these set of assets, we will be in need of increasing security among the communication channels of these IoT Infrastructures. With the rapid advancement of technology, the implementation of IoT is increasing with various other relevant fields of IoE and IoNT. This expeditious development will be leading these devices and digital assets to the risk of data theft or hijacking the communication channel.
Several techniques are being used to protect channels involved in IoT infrastructure. As for the case of encrypting the data which is passed from sender to receiver, popular techniques which are being used these days are DES, AES, RSA, etc. With the swift increase in the development of the digital world, there arises a need for the development of some more dynamic and versatile techniques which can help to communicate the devices to maintain the security of the communication channel since its establishment. This technique works over the principal working of Generative Adversarial Neural Networks. Using GANs in this technique, a secure cryptographic way of sending data from the sender to receiver using two Artificial Neural Networks is tried to be attained.

In the year 2002, the notion of neural cryptography was initially developed. Because the underlying process may be as discrete as possible, neural networks' black-box qualities are ideal for cryptography applications. Neural Cryptography is quite a complicated concept to understand and implement, but it is an incredible domain to be developed soon. Due to the wide range of applicability, it is having a lot of potential to provide secure communication in IoT systems that will be governed by AI.

With generative adversarial networks, Abadi and Andersen recently presented the concept of adversarial neural cryptography (GANs). A GAN model is composed of a set of two Neural Networks namely, A generator and A Discriminator. Both Neural Networks have their unique set of responsibilities in the functioning of Generative Adversarial Networks Respectively. The aim of the Generator is to generate new data after getting trained over the training Dataset over some required and limited datasets. Whereas on the other hand, The Discriminator is responsible for discriminating among the data generated by the generator and the real data present in the set of data over which the discriminator is trained in the arrangement mutually. This process is done in phases to attain the required state. In the particular condition of Adversarial Neural Cryptography, Three Neural Networks are implemented in the arrangement to attain the desired state which are named over conventional arrangements of Cryptographic Examples Alice, Bob, and Eve. The plaintext is encrypted through the secret key which is shared by Alice and Bob. Then the ciphertext is sent over the Communication Channel by Alice to Bob, in between this scenario the third neural network comes into play, namely Eve whose primary intention is to extract the plaintext from ciphertext without knowing the secret key. Alice tries to make the extraction of plaintext from ciphertext difficult for Eve, as Alice is taught in an adversarial manner for the context to achieve.

This newfangled perspective of Adversarial Neural Cryptography of making neural networks to learn for performing Cryptographic responsibilities on their own has a lot of Potential in implementing them in the IoT Infrastructure. Although, Certain hurdles can be faced during the implementation of the technique in IoT Systems. In the first place, Adequate training requires a prolonged period which may not result as desired sometimes. In addition to this, GAN just like auxiliary Neural Networks generates values of floating-point as Ciphertext which is not easy to assess for judging the robustness as compared to binary. Also, the Neural Networks were too complicated to be implemented and deployed over IoT devices earlier. IoT devices needed Lightweight Cryptographic Algorithms which are rapid and swift to be deployed in such cases since they can be trained efficiently over a broader range of Neural Cryptosystems. For the sake of Data pre-processing Neural Cryptosystems of a single encryptor, Convolutional Layer is can be associated with other Neural networks in the System. With the recent technological
advancements with time, Lightweight can be replaced with this latest technique of Adversarial Neural Cryptography.

The remnant part of the composition is laid out as: In part II, related research on Adversarial Neural Cryptography is discussed. The methods used in the Present Scenario for Implementing Security in IoT Infrastructure are discussed in Section III. Adversarial Neural Cryptography undermines Security Implementation and Enhancements in IoT Infrastructure in Section IV. In section V, the approaches' study and analysis findings are provided, followed by conclusions and future research prospects.

2. Related Works

Kanter et al. presented a novel two-party key exchange protocol in 2002 based on the postulation for chaotic synchronization, through which two separate chaotic systems can coincide at a single point which can be considered in the ailng states [1]. This arrangement is considered as a set of Neural Networks which will be at arbitrary states while initialization and can result in providing worthwhile information about their respective states with each iteration of update. An extension of this, it was found that if an attacker utilizes the exact Neural Network with the same edification is not able to synchronize with the other participant networks of the arrangement. Shamir et al. used three separate approaches to break the system in the same year.

Following that, chaotic NN was utilized in various publications to provide encryption as well as hash methods [2–4]. All these works simply employed the chaotic NN as a tool for randomization. All of these algorithms, however, were also broken [5–7]. Pseudorandom number generators (PRNGs) were also developed using NNs [8–10]. The parameters of the NNs were used as the fuel for the PRNG in these studies, the uncertain randomness of outcome was evaluated using tools such as the NIST random number generator test suit.

In a well-defined and functional communication channel, Neural Network’s characteristics as an encryptor are transferred to the decryptor for efficient communication as mentioned by Yayik and Kutlu [11]. These Characteristics include weights, the number of neurons, and the number of hidden layers in the concerned Neural Networks. Backpropagation of any neural network with the same parameters may likewise reverse the output of a neural network, as shown in [12] and [13]. Because two synchronized neural networks may safely exchange secret keys in a public channel, neural cryptosystems can also solve the problem of secret key distribution for symmetric encryption methods [14]. As shown in [15], synchronization between the encryptor and decryptor may also be achieved by communicating partial weights of the neural networks such that the adversarial neural network cannot directly utilize the weights. While the use of two neural networks for symmetric encryption has been extensively studied in the literature, it necessitates neural network topologies that feature inverse operations or allow a portion of the neural network parameters to be broadcast via the communication channel. Neural networks having chaotic features can be employed in neural cryptosystems, introducing unpredictability into the neural networks. In instance, to secure user privacy within cloud servers, [16] advocated using radial basis function networks with natural noise simulations as one-to-one personal encryption algorithms. Other network architectures, such as topology-changing neural networks [17] and multi-layer auto-encoder neural networks [18], can be employed in neural cryptosystems. Moreover, the plurality of the mentioned neural
cryptosystems is really not end-to-end encryption methods since the neural networks do not execute at least one element of the cryptographic functions.

Ian Goodfellow initially proposed Generative Adversarial Networks (GANs) in 2014. It's a two-network system with a Discriminator and a Generator. The arrangement of this network is trained in an adversarial manner, with the Generator model G attempting to replicate the real or original data distribution using a random noise vector sent off to the Discriminator model D. The authors' Proposed Model will be explored in-depth in the following sections [19].

![Figure 1. Training of Generative Adversarial Neural Networks](image)

However, Abadi and Andersen just started a research direction on using adversarial neural cryptography to safeguard communications. Its goal is to develop neural networks that can learn to encrypt data without being taught any specific encryption technique. This method is based on generative adversarial networks (GANs), which pit neural networks against each other in order to achieve a goal in the presence of an opponent (i.e., another neural network). GANs are based on the notion of two neural networks fighting to produce a new collection of data that may be used as actual data. GANs are remarkable in their capacity to imitate diverse forms of data and are thus widely employed in picture and speech synthesis to produce synthetic data that is indistinguishable from the original data distribution [20].

The intention was to present a system in which two NNs identified Alice and Bob attempted to communicate while restricting what a third NN named Eve might learn by listening in on the conversation. They did not specifically define cryptographic methods to any of these neural networks in their research. In fact, they demonstrated that Alice and Bob might effectively learn how to execute encryption and decryption, depriving Eve of access to the communication. They
did not, however, reveal any solution taught by the networks in their work. We'll go through their system in this part to give you a broad idea of how ANC works.

The Concept of Adversarial Neural Cryptography was introduced in [20] and it illustrated that it could be used as a reliable approach to implement Neural Cryptography in IoT [21]. As aforementioned, adversarial training adds Eve as a second neural network there in cryptosystem that is typically taught with less knowledge than Bob. Eve only gets the ciphertext in, whereas Bob gets both secret keys and even the ciphertext. In comparison including both Alice and Bob, Eve gains a small advantage by having more training iterations. Alice must make it increasingly challenging for Eve to read ciphertext, hence increasing the cryptosystems’ security level. Eve can be given further benefits to improve the security of cryptosystems; for example, two congruent neural cryptosystem technologies are based around adversarial neural networks were presented. The first method requires that the adversary neural network (Eve) acquires a portion of something like the secret keys and indeed the ciphertext in order to decrypt plaintext or output the entire set of secret keys. To forecast secret keys or decode the plaintext, the second method utilizes attacker neural networks. Traditional cryptographic attacks, on the other hand, can be used in neural cryptography to improve the security of neural cryptosystems. Following Abadi and Andersen's work, a flurry of research emerged to investigate the model's security (e.g. [22]), as well as to prolong this to a presumed considered secure protocol, and a slew of others [23].

Using neural cryptosystems and binary latent space, we offer a unique lightweight IoT device authentication, encryption, and key distribution technique. End-to-end encryption techniques used in the adaptive neural cryptosystems are symmetric, public-key, also without keys. The findings suggest that it might be a potential security and privacy solution for the next substantial IoT systems.

3. **Security Implementation and Sources of Security Threats in IoT devices**

The upcoming future will be filled with so many fascinating IoT devices to control any Individual’s daily life. High jacking or compromising of these IoT personal assets can result in Devasting Situations for any Individual or an Organization. The study about security concerns in the field of IoT Infrastructure is a crucial point in Research as the failure of any IoT Suite can result in severe consequences. The objective of IoT security is not only to maintain privacy and confidentiality but also to protect the devices and data of the ecosystem from getting affected by any external threat. The technological exponential development in IoT is leading the growth in IoT security through various techniques, algorithms of communications, and frameworks. Various devices involved in the process developed, designed, and implemented in different application conditions due to which this situation is in need of solutions that are dynamic in Nature. Several services and interfaces are present to implement IoT for making the everyday task simple and easy. IoT ecosystem compromises three levels namely hardware, communication links, and services/interfaces in accordance with the latest Research. [25] On a broader observation an IoT Application can be classified into four levels of categorization:

1. **Sensor layer,**

2. **Network Layer,**
3. Middleware Layer, and


Every level has its own distinguished plethora of technologies and respective implementation. In addition to that, each layer has its own set of threats and vulnerabilities which can cause a huge risk to the whole implemented suite. Different kinds of sensors are considered in the Sensor Layer which can detect the condition of the surrounding physical environment and help the user to have acquaintance with the physical surrounding. Several reasons can cause distress regarding security concerns in the sensing layer, like Low power node High jacking, Malicious Code Injection Attack, False Data Injection Attack, Side-Channel Attacks, Interference, eavesdropping, Sleep Deprivation Assaults, Booting attacks, etc [24].

Most kinds of attacks can be tackled and delayed while using the technique of Adversarial Neural Cryptography at the sensor layer with initial establishments of network for IoT. The data flowing in the IoT suite from various sensors can be made secure with this technique with major processing in the cloud. Disturbing data from the sensors by the attacker can result in very disastrous circumstances as will lead to prediction or detection manipulation of any situation.

In a similar manner, the Network layer can also be secured through this technique as with the increasing advancement trends in the technology can provide nodes of the IoT Infrastructure enough power to implement this technique more efficiently. This technique will evidently improve the performance and robustness in terms of security for the Middleware layer and the Application Layer too. The main concern of the implementation lies in the Application layer of any IoT Suite. Securing the data and communication at the Application level is the main focal point of any cryptographic algorithm.

For many years to come, it will be projected regarding security implementations over the Internet of Things that they will continue to be the target and attack vectors in the future. This may be because of the increasing quantity of IoT devices in different networks, the available heterogeneity among the IoT protocols, and device makers' inadequate or default security safeguards. As a security precaution, cyber security methods compromising authentication, encryption, as well as firewalls should be implemented into the IoT suites. This, however, is inadequate. The Internet of Things is distinguished from traditional networks through interactions and integrations of physical and cyber systems connected over the Internet.

New increasing vulnerabilities, including open communication channels, the existence of hostile actions, that can be due to any reason in the network, and unsecured physical items, can lead to new other forms of risks to IoT Infrastructure. Because of their occasional patching and upgrades, IoT devices are prone to surface attacks: many IoT devices come with little, if any, encryption, or authentication. Furthermore, as these devices are frequently positioned in danger-prone environments over the Internet and are continuously accessible, they may have minuscule or negligible protection available for unauthorized physical access.

For tackling IoT security problems, authentication and encryption may be acceptable choices. However, efficient authentication and encryption for less-powered, computationally and resource-wise constrained devices are still in their preliminary stages and don’t guarantee
the absence of nodes that can be hostile in the network, such as malfunctioning devices or PCs. Furthermore, for simplicity, manufacturers commonly employ manual hardcoded credentials like passwords, which almost always results in a significant failure for authentication.

Over Alternative situations, it costs a lot to implement trust and reputation-based detection of a malicious node as it can lead to the high cost of communication, delay in the end-to-end connection with elevated false-positive rate. For the Overall development of IoT security and proper implementation, it is required to work over the techniques of lightweight cryptographic techniques with multifactor authentication over the Application and Network Layer of the Infrastructure. Also, to study and develop alternatives to these using AI or Deep Learning for exponential enhancements in the domain. To overcome all the related issues, Adversarial Neural Cryptography is introduced for low-cost encryption at the Physical or Application Layer of IoT, as technological advancements can provide enough computational powers to the node that Adversarial Neural Cryptography can give significant results for the purpose. [26]

Furthermore, certain malware attacks might cause more severe infrastructure damage. The method provides the key applications for IoT security and can protect against threats such as eavesdropping, denial of service, and so on, thanks to encryption and authentication at several tiers. Adversarial Neural Cryptography comes into play when the complexity of various algorithms diminishes day by day as the electronic world evolves.

4. Security Implementation through Adversarial Neural Cryptography

The idea of Generative Adversarial Network was proposed by Ian Goodfellow along with his team date back in 2014. As illustrated in Figure 2, the technique works on the principle of two networks, namely the Generator and the Discriminator, working together and training each other in adversarial mode.

![Figure 2. Working of Generative Adversarial Neural Networks](image)

The Generator G is designed to intend to generate data from random noise of latent space. Prior to that, the Discriminator D is trained in the initial phase to discriminate between the real data and the generated data. In the next phase of implementation, the Discriminator
discriminated between the data to categories of Real and Fake Data for improvising the Outcome. After Discrimination is done by the Discriminator, it returns the probability of data about its belonging to the category of Real or Fake Data. Over this basis, the Result of the Generator is optimized to generate real-like synthetic data. The primary aim of this arrangement is to increase the probability of the generated data such that it is not easy for the discriminator to distinguish between the real dataset or the synthetic dataset. In contrast, the arrangement works to contend between the two networks in Zero-Sum Game, where the Generator G is trying to amplify the probability and Discriminator D is trying to reduce that. The Loss Function of the above-mentioned situation is presented numerically as:

\[
\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p(z)} [\log (1 - D(G(z)))]
\]

Here in the equation, \(E\) symbolizes Expectation, \(x\) represents the real data sample, \(z\) represents the random noise, \(G(z)\) represents the data generated by G Generator, \(D(G(z))\) represents the probability of the D Discriminator over the Generated Data, and \(D(x)\) represents the probability of the Discriminator over real data \(x\). As mentioned previously, the aim of the arrangement is to achieve the value of \(D(G(z))\) nearer to 1 by G and nearer to 0 by D.

GANs are proved very successful in the situation of Multimedia Processing of Records ranging from Synthesizing Images and Videos out of nowhere to enhancing the quality of the multimedia records. GANs can generate images of people or places which do not really exist. GANs have recently found applications in sectors like security. With the Great success of applicability, GANs got the way to enter in the domain of prediction of security threats and analyzing the places for loopholes in the working systems. The method of prediction of these loopholes can result in a more robust and efficient way to tackle the security-related assaults in advance.

The working idea of Generative Adversarial Networks was taken on the advanced level of implementing Cryptography through Adversarial Neural Cryptography to secure the communication channel between two parties by Martin Abadi and David G. Andersen in their work in 2016. This situation is quite similar to the conventional symmetric key encryption where two independent candidates try to communicate through the Channel and Eve tries to eavesdrop on the channel and get decipher the original plain text without any knowledge about the secret key possess only by the communicating parties. This kind of attack will be considered a COA (ciphertext only attack) which can be summed up as one of the substandard attacks through the attacker. For the present context, all three participants here are Neural Networks that are trained to reach a particular goal in the end after the process. The intention of this training over this set of Neural Networks is to minimize the distance of Plaintext \(P\) with Output Plaintext by Bob \(PBob\). Here, \(P\) is the Input of Alice and \(PBob\) is the Output of Bob. Eve is another Neural Network that will be trained Adversarial with the intention to Reduce the Distance of Plaintext \(P\) and the Output of Eve \(PEve\). On the Contrary of Eve’s objective, Alice and Bob wish to intensify the distance among the plaintexts. The generative adversarial network (GAN) is the inspiration for the competitive method shown here (Goodfellow et al., 2014a). Because no precise strategies are stated ahead of time, our protection aim is rather ad hoc and is dependent on the adversarial neural network.
The ANC method is based on the traditional typical cryptographic situation depicted in Figure 3 compromising of the three participating parties namely, Alice, Bob, and Eve. During the communication between Alice and Bob, both wish to transfer the data without any hurdle and encrypt data with a shared secret Key. Whereas Eve is an attacker who wishes to eavesdrop on the channel to extract the same plaintext from Ciphertext which was sent from Alice to Bob. Since all these mentioned entities are Neural Network for the case, all the Values (Key K, Plaintext P, PBob, PEve, and C) will be the tuples of floating-point numbers instead of a sequence of bits.

These three designated Neural Networks works with concerned parameters, for present instance suppose A, B and E for Alice, Bob and Eve respectively. Proceeding further with the idea contemplate EA (A, P, K) as the Output from Alice which is applied encryption on Plaintext P and respective secret Key K. Similarly contemplate DB (B, C, K) as the output from Bob which is applied decryption over the ciphertext C and respective secret Key K. Now, the distance can be evaluated to scale distance from the actual plaintext and the estimated plaintext from the ciphertext. This Distance can be calculated through:

\[ d(P, P') = \frac{1}{N} \sum_{i=1}^{N} |P_i - P'_i| \]

Here for the case N is the span of plaintexts.

Now, the intention of Eve is to reassemble and get the actual plaintext precisely. The Loss Function of Eve can be represented as:

\[ L_E(\theta_A, \theta_E, P, K) = d(P, D_E(\theta_E, E_A(\theta_A, P, K))) \]

In the context of Numerical Representation LE (A, E, P, K) represents the incorrectness in Eve’s evaluation of Plaintext. The predicted value for the Loss function of Eve over the dispersed plaintext and keys can be represented as:
\[ L_E(\theta_A, \theta_E) = E_{P,K}[L_E(\theta_A, \theta_E, P, K)] \]

With multiple iteration, a situation can be attained by optimizing the loss function of Eve to attain “ideal Eve”:

\[ O_E(\theta_A) = \arg\min_{\theta_E} (L_E(\theta_A, \theta_E)) \]

Correspondingly, the Loss function for Bob can also be evaluated over the dissemination of plaintext with keys as:

\[ L_B(\theta_A, \theta_B, P, K) = d(P, D_B(\theta_B, E_A(\theta_A, P, K), K)) \]

\[ L_B(\theta_A, \theta_B) = E_{P,K}[L_B(\theta_A, \theta_B, P, K)] \]

During the communication between the communicating parties to maintain authentication and Integrity the arrangement is needed to avoid Eve from getting anything about their actual communication. This could be achieved through summing up LB along with optimally evaluated LE as it can lead to the loss function for Communication between Alice and Bob:

\[ L_{AB}(\theta_A, \theta_B) = L_B(\theta_A, \theta_B) - L_E(\theta_A, O_E(\theta_A)) \]

Ultimately, by reducing \( L_{AB} \), one may derive the "optimal Alice and Bob":

\[ (O_AO_B) = \arg\min_{(\theta_A, \theta_B)} (L_{AB}(\theta_A, \theta_B)) \]

We may deduce that the loss of Eve is much more times more than Bob's, implying that Bob will be able to retrieve the message while training for Adversarial Neural Cryptography in IoT Devices, but Eve will not. End-to-end encryption techniques used in the proposed neural cryptosystems are symmetric, public-key, and without keys. The suggested neural cryptosystems can be taken through a number of tests to check how well they function and how secure they are, and the findings demonstrate that they have prospective potential in implementing security and maintaining privacy solution over the communication channel for upcoming smart generation of large-scale suite of IoT systems.

5. Conclusion

Diving in the method for creating a setup of a neural network-based automatic encryption system. Starting with the fundamental symmetric key model, we analyzed the system's security using numerous specialized statistical models, demonstrating that the original suggested method is insecure in terms of distinguishability. Though Adversarial Neural Cryptography is having a lot of scope in securing channels of communication, it can also be utilized to protect IoT devices and their respective suites from eavesdropping and high jacking them. Using the studied technique it can be concluded that ANC can provide an advanced form of cryptography with minimal human interference and maintain integrity, confidentiality, and authentication during Communication for IoT and IoE Infrastructures. The new encryption methods are more resistant to a variety of attacks and are more durable and adaptive. Future work will focus on how to enhance neural networks so that both parties can collaborate more effectively for communication with considerably lesser steps, and how to establish the appropriate hyper-
parameters to improve randomness in order to withstand various forms of attacks. It's also interesting looking at how to use neural networks to create additional security solutions with a lot of features.

Various other methodologies and techniques can be explored in the expansion of this dimension to indoctrinate neural networks for hiding plaintext in ciphertext or establishing a robust and smart system to provide a more secure channel for communication in the future study using the presented approach. Furthermore, future advancements may result in lower communication, calculation, and time overheads, as well as enhanced security.

References


