International Journal of Computer Applications Technology and Research
Volume 13–Issue 10, 90 – 98, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1310.1009

# The Impact of Adversarial Attacks on 5G Network Systems

Okonkwo Chinonso Joseph
Department of Computer Science
Chukwuemeka Odumegwu Ojukwu University
Nigeria

Prof Ogochukwu C. Okeke
Department of Computer Science
Chukwuemeka Odumegwu Ojukwu University
Nigeria

**Abstract**

This paper presents a qualitative investigation into the Quality of Service (QoS) and adversarial attack impacts on 5G Network technology. Recently 5G networks have contributed significantly to the advancement of telecommunication technology. This study identifies some of the negative impacts of adversarial attacks on 5G networks such as network configuration manipulation, exposure to malicious software, manipulation of hardware, leakage of information and authentication abuse. The research methodology adopted in the study is model-driven development. The 3 categories of adversarial attack such as gradient-based attack, score-based attack, and decision-based adversarial attack models are presented. Then, the detection techniques applied to countering the effects of adversarial attacks which include gradient masking/obfuscation, robust optimization and adversarial example detection techniques are discussed comprehensively. The work was concluded by recommending the implementation of a regularization method for the mitigation of adversarial attacks in future studies due to its flexible performance capacity and scalability.

**Keywords:** 5G; Adversarial Attack; Perturbation; Regularization; Artificial Intelligence

## 1. INTRODUCTION

A few decades ago saw the emergence of mobile wireless communication networks, which have facilitated information sharing across states, cities, nations, and even continents. Wireless communication is always being improved in terms of data capacity, speed, frequency, technology, and latency. These alterations have been divided into four generations of mobile wireless technology (Adebusola et al., 2020).

Over the past fifteen years, mobile and wireless networks have experienced exponential expansion. The smooth integration of cellular networks like GSM and 3G is the main goal of 4G. Multimode user terminals are considered essential for 4G, although varying QoS support and security protocols across various wireless technologies continue to be difficult to implement (Patel and Patel, 2017).

5G refers to the fifth generation of mobile technology. 5G technology has transformed how cell phones may be used with extremely high bandwidth. 5G is a high throughput; broad area coverage packet-switched wireless technology. A 20 Mbps data throughput and a frequency range of 2 to 8 GHz are made possible by 5G wireless utilization of millimeter wireless and orthogonal frequency division multiplexing (OFDM). 5G will be a network with a packed

architecture. The genuine wireless network, known as the 5G communication system, is anticipated to be able to enable wireless World Wide Web (www) services between 2010 and 2015 (Emma and Peng, 2020). Concerns regarding artificial intelligence's (AI) and machine learning's (ML) susceptibility to adversarial effects are growing as these technologies become more and more integrated into nearly every sector, including 5G mobile networks. Adversarial machine learning is the study of learning in the face of adversaries, and it has drawn increasing interest from researchers in a variety of fields, including computer vision and natural language processing (Goodfellow et al., 2015).

The goal of an adversarial machine learning attack is to manipulate the training process, either directly poisoning the training data or by injecting perturbations to the training samples such that the target model is trained with erroneous features and subsequently makes errors later in the inference time. An adversarial machine learning attack can occur during either the training or the inference stage(Steinhardt et al., 2017). This paper presents the challenges of adversarial network attacks in 5G network technology by hampering the network's quality of service. Then the various kinds of adversarial attacks are presented to establish a better understanding of the attack model. Furthermore, key techniques applied for defending a network from adversarial attacks are presented such as gradient masking, robust optimization and adversarial example detection are presented.

## 2. ADVERSARIAL ATTACK AND IMPACTS ON 5G NETWORK TECHNOLOGY

The cost of the models that employ this strategy has increased in tandem with the success that AI and ML have had recently, making them the most sought-after target for adversarial example assaults. Deep Neural Networks (DNNs) typically employ a gradient-based optimizer during training and have a differentiable loss function. This allows for the creation of adversarial examples based on gradients by altering an input sample in the direction of the gradient of the loss function relative to the input sample (Christopher, 2021). In white-box circumstances, this enables the creation of an adversarial perturbation to execute a non-targeted assault.

Because of the 5G network's increased complexity, speed, and new features, network security is more important than ever for both 5G providers and customers. Similar to other support systems, supporting a wider range of services calls for additional resources and may result in security issues being overlooked. It is important to note that attacks discovered here are also inherited because the Internet Protocols (IPv4/IPv6) handle a large portion of the communication inside the architecture. Below we explore some of the hostile security issues 5G faces (Farooqui et al., 2022; Angelo et al., 2023):

### a. Network Configuration Manipulation

Network configuration manipulation attacks encompass several techniques such as routing assaults, which are often referred to as DNS manipulation, routing table poisoning, or tampering with cryptographic keys and rules. These attack

methods are directed at the DNS server, the Policy Control Function (PCF), or the Access and Mobility Management Function (AMF). Attacks against the MME and PCRF would be directed from the EPC's point of view. Using a least-privilege permission architecture and requiring reviews of changes for all users are two possible ways to reduce attacks (Park et al., 2021). DNS Security (DNSSEC) extensions can be used as a countermeasure to stop DNS tampering. A public key is provided by DNSSEC to validate the outcome of a DNS query.

**b. Malicious Software**

Software assaults on Core Networks (CN) have the potential to destroy data or make services unavailable. One should routinely apply software updates to fix vulnerabilities to defend against these assaults. In addition, important data should be backed up in case of data corruption.

**c. Hardware Manipulation**

A side-channel attack is a popular technique that may be applied against actual hardware present in CN. In real terms, a side-channel attack manipulates or obtains data by using current measurements from a specific device. However, the side-channel attack has a high exploit complexity because it is a physical attack. Additionally, preventing side-channel assaults necessitates bespoke hardware, raising the deployment cost.

**d. Information Leakage**

Unauthorized access to leaked logs, cryptographic keys, and user data. Attacks of this kind would be aimed at the SMF. Implementing IPsec tunnel encryption as a countermeasure might guarantee IP packet integrity and privacy.

**e. Authentication Abuse**

The outcome of conducting privilege escalation violates integrity. The AMF and the Authentication and Key Agreement (AKA) protocol, a challenge-response system built on symmetric cryptography and a Sequence Number (SQN), are targets of these kinds of attacks. Research has shown that an Exclusive-OR (XOR) and a lack of randomization may be used to alter a replay attack, which AKA guards against.

## 3. RESEARCH METHODOLOGY

The methodology adopted for the development of this paper is Model Driven Development (MDD). The most crucial low-code development tenet is model-driven development. It's a software development process that allows teams to graphically design complicated systems using reduced abstractions of pre-built components. Model-driven development lowers human-process interference through automation and simplifies complexity through abstraction. In model-driven development projects, the model is not interpreted into code but rather is executable at runtime. This enables code-centric projects to avoid frequent operations and quality problems with model-driven development (Farshidi et al., 2020).

## 4. ADVERSARIAL ATTACK MODEL

The method of creating an adversarial example using a victim model and a natural sample is known as an adversarial attack. This approach to creating adversarial instances is shown in Figure 1. The natural input in this case is represented by x0, and the DNN can accurately predict its label y0. The goal of an adversarial assault is to identify a

minor perturbation δ that will cause the victim model to incorrectly classify the adversarial example x∗ = x0 + δ, which seems to be identical to x0 to humans (Li et al., 2021). The attack techniques may be classified into three categories: (1) gradient-based, (2) score-based, and (3) decision-based, depending on the information required. The majority of these techniques are capable of both targeted and untargeted assaults. Typically, an attack technique falls into one of the three groups; however, new research indicates that combining strikes from different categories may result in a more effective attack (Croce and Hein, 2020).

**4.1 Gradient-Based Attack**

Many of the assault techniques used today fit under this group. These techniques create adversarial instances by using the gradients of the loss of the input. For example, the Fast Gradient Sign Method (FGSM) (Goodfellow et al., 2015) uses a step size to regulate the `∞ norm of perturbation and creates adversarial instances depending on the sign of gradients.

**4.2 Score-Based Attack**

In practice, the attackers might not have access to certain model data, such as the gradient. The assault techniques based on scores don't need gradients to be accessible. Based on the victim classifier's output scores, f(x)i, they launch adversarial assaults. Chen et al. (2017), for instance, suggested a technique to create adversarial instances using the estimated gradient and estimate the gradient using score information.

**4.3 Decision-Based Attack**

In many real-world scenarios, the attacker simply has access to the model's projected labels—they are not privy to gradient or score data. Both gradient-based and score-based approaches fail when the only information given is the projected label c(x). A transfer attack technique was presented by Papernot et al. (2017), and it just needs observations of the labels that the model predicts. The primary concept is to train a replacement model that bears resemblance to the original model and then target the replacement model.

## 5. ADVERSARIAL ATTACK DETECTION TECHNIQUES

Improving the robustness of DNNs to defend against adversarial cases has been the subject of much study. Generally speaking, techniques to improve model resilience may be divided into four basic categories: adding hostile instances to the training set, using randomization to thwart adversarial attacks, using projection to eliminate adversarial perturbations, and identifying adversarial examples rather than accurately categorizing them are the four main strategies (Li et al., 2021). Different solutions have been suggested as countermeasures against adversarial instances to safeguard the security of deep learning models. These countermeasures may be divided into three primary types:1) Gradient masking/Obfuscation, 2) Robust optimization and 3) Adversarial examples detection (Xu et al., 2020).

**5.1 Gradient Masking/Obfuscation**

Gradient masking/obfuscation is a tactic where a defence purposefully conceals the model's gradient information to trick their opponents, as the majority of attack techniques

rely on this information to determine the classifier's gradient (Hinton et al., 2015).

**a. Shattered Gradients**

Pre-processing the input data is one way that certain researchers, including (Buckman et al., 2018; Guo et al., 2017), attempt to safeguard the model. They then train a DNN model f on g(X) after adding a non-smooth or non-differentiable pre-processing g(.). Adversarial assaults fail because the trained classifier f(g(.)) is not differentiable in terms of x.

**b. Stochastic/Randomized Gradients**

To confuse the opponent, some defence tactics attempt to randomize the DNN model. We train a collection of classifiers, for example, s= {Ft: t =1,2,3,...,k}. We pick a classifier at random from the list and forecast the label y while evaluating data x. Due to the adversary's ignorance about the classifier that the prediction model uses, the assault success rate will be lower.

**c. Exploding & Vanishing Gradients**

Before categorizing them, generative models are suggested to project a possible adversarial example onto the benign data manifold by both PixelDefend (Song et al., 2017) and Defense-GAN (Samangouei et al., 2018). Defense-GAN employs GAN architecture, whereas PixelDefend utilizes the PixelCNN generative model (Oord et al., 2016; Silver et al., 2016). It is possible to think of the generative models as a purifier that turns hostile samples into benign ones.

**5.2 Robust Optimization**

Robust optimisation techniques seek to alter the DNN model's learning process in order to increase the classifier's resilience. They research the process of acquiring model parameters that can yield accurate forecasts on prospective adversarial cases. The primary goals of the studies in this topic are learning model parameters in order to reduce the average adversarial loss.

A resilient optimisation algorithm should, in general, be aware of any possible threats or attacks beforehand. Next, the defences construct classifiers that are impervious to this particular attack.

**a. Regularization Methods**

Another class of strong defensive strategies makes use of randomization to fight off hostile examples. Adversarial perturbation may be thought of as noise, and by adding random elements to the model, many strategies have been put forth to increase the resilience of DNNs.

Xie et al. (2018) presented a straightforward pre-processing technique to randomise neural network input in an effort to exclude any possible adversary disruption. The input is randomly enlarged to multiple sizes throughout the testing phase, and then randomly padded zeros are inserted around each of the scaled inputs. The authors showed that big datasets like ImageNet might benefit from the application of this straightforward technique. Similarly, Zantedeschi et al. (2017) demonstrated that the learnt model would become somewhat more stable against adversarial cases by utilising a modified ReLU activation layer (called BReLU) and augmenting the training data with noise in the origin input (Carlini and Wagner, 2017)

**5.3 Adversarial (re)training**

**1) Adversarial training with Fast Gradient Sign Method (FGSM)**

Goodfellow et al. (2014) introduced the concept of adversarial training using the Fast Gradient Sign Method (FGSM), denoted by (x', y). This method involves incorporating adversarial examples generated during the training process. By introducing counterexamples with accurate labels (x', y) into the training set, the objective is to train the model to accurately predict the label of forthcoming adversarial instances. This inclusion in the training set helps inform the classifier that x' belongs to class y, enhancing the model's robustness against adversarial attacks.

**2) Adversarial Training with Projected Gradient Descent (PGD)**

Rather than utilising single-step assaults like FGSM, the PGD adversarial training proposes employing a projected gradient descent attack (Madry et al., 2017). One way to think about the PGD assaults is as a heuristic for identifying the "most adversarial" scenario.

**3) Ensemble Adversarial Training**

Ensemble adversarial training, according to Tramer et al. (2017), developed an adversarial training technique that can defend CNN models against single-step attacks and be used to big datasets like ImageNet. Their primary strategy is to add hostile instances made from other pre-trained classifiers to the classifier′s training set.

**b. Provable Defences**

It has been demonstrated that adversarial training works well at shielding models from aggressive instances. That being said, there is still no official assurance on the trained classifiers' safety. It would be hazardous to immediately deploy these adversarial training algorithms in safety-critical jobs since we never know if more aggressive attacks may breach such protections.

## 6. CONCLUSION AND RECOMMENDATION

This paper presents a qualitative investigation in the Quality of Service (QoS) and adversarial attack impacts on 5G Network technology. The study identifies some of the negative impacts of adversarial attacks on 5G networks such as network configuration manipulation, exposure to malicious software, manipulation of hardware, leakage of information and authentication abuse. The research methodology adopted in the study is model-driven development. The 3 categories of adversarial attack such as gradient-based attack, score-based attack, and decision-based adversarial attack models are presented. The methods that enhance the ML model robustness for model protection such as augmenting the training data with adversarial examples, leveraging randomness to defend against adversarial attacks, removing adversarial perturbations with projection, and detecting the adversarial examples instead of classifying them correctly are identified. Then, the detection techniques applied to countering the effects of adversarial attacks which include gradient masking/obfuscation, robust optimization and adversarial example detection techniques are discussed comprehensively. This paper recommends the application of the regularization method for the early detection of adversarial attacks due to its reduced model complexity, improved transferability detection, noise tolerance and scalability.

## 7. RESEARCH HIGHLIGHTS

1. This research identified the major threats to 5G network systems.
2. Detailed exploration of countermeasures such as regularization methods for early threat detection was discussed.
3. Advanced detection frameworks, to mitigate adversarial attacks and secure 5G infrastructure was discussed
4. To protect machine learning models within 5G networks, resilient strategies against adversarial threats were suggested.

## 8. REFERENCES

Adebusola J., Ariyo A., Elisha O., Oubunmi A., & Julius O., (2020) An Overview of 5G Technology. 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS) 978-1-7281-3126-9/20/$31.00 ©2020 IEEE 10.1109/ICMCECS47690.2020.24085

Angelo B., Wøidemann K., & Andersen B., (2023) 5G Attacks and Countermeasures. In Proceedings of 25th International Symposium on Wireless Personal Multimedia Communications IEEE. https://doi.org/10.1109/WPMC55625.2022.10014962

Buckman J., Roy A., Raffel C., & Goodfellow I.., (2018) Thermometer encoding: One hot way to resist adversarial examples. In Proceedings of the 6th International Conference on Learning Representations, Vancouver, Canada, 2018.

Carlini N., & Wagner D., (2017) Towards evaluating the robustness of neural networks. In 2017 IEEE Symposium on Security and Privacy (SP), 39–57

Carlini N., & Wagner D., Adversarial examples are not easily detected: Bypassing ten detection methods. In Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, ACM, Dallas, USA, pp.3–14, 2017. DOI: 10.1145/3128572.3140444.

Chen P., Zhang H., Sharma Y., Yi J., & Hsieh C. (2017) Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. In Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, 15–26.

Christopher C., (2021) An Introduction to 5g, The New Radio, 5G Network and Beyond. Vol. 1, 2021

Croce F., & Hein M. (2020) Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. In International Conference on Machine Learning, 2206–2216.

Emma X., & Peng L., (2020) 5G Network: An Overview of the Pros and Cons. ©IDOSR PUBLICATIONS International Digital Organization for Scientific Research ISSN: 2550-794X IDOSR JOURNAL

OF SCIENTIFIC RESEARCH 5(2) 40-46, 2020.

Farooqui M., Arshad J., Khan M., (2022) A Layered Approach to Threat Modelling for 5G-Based Systems. Electronics 2022, 11

Farshidi S., Jansen S., & Fortuin S., (2020) Model-driven development platform selection: four industry case studies. Software and Systems Modeling https://doi.org/10.1007/s10270-020-00855-w

Goodfellow I., Shlens J., & Szegedy C., (2014) Explaining and harnessing adversarial examples. ArXiv: 1412.6572, 2014

Goodfellow I., Shlens J., & Szegedy C., (2015) Explaining and harnessing adversarial examples. 2015, arXiv:1412.6572.

Guo C., Rana M., Cisse M., L., & van der Maaten (2017) Countering adversarial images using input transformations. ArXiv: 1711.00117, 2017.

Hinton G., Vinyals O., & Dean J.., (2015) Distilling the knowledge in a neural network. ArXiv: 1503.02531, 2015.

Li Y., Cheng M., Hsieh C., & Lee T., (2021) A Review of Adversarial Attack and Defense for Classification Methods. arXiv:2111.09961v1 [cs.CR] 18 Nov 2021

Madry A., Makelov A., Schmidt L., Tsipras D., & Vladu A., (2017) Towards deep learning models resistant to adversarial attacks. ArXiv: 1706.06083, 2017.

Papernot N., McDaniel P., Goodfellow I., Jha S., Celik Z., & Swami A. (2017) Practical black-box attacks against machine learning. In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, 506–519.

Park S., Kim D., Park Y., Cho H., Kim D., & Kwon S., (2021) 5G Security Threat Assessment in Real Networks. Sensors, 2021.

Patel B., & Patel M., (2017) Introduction About 5G Mobile Technology. International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV6IS060397 (This work is licensed under a Creative Commons Attribution 4.0 International License.) Published by : www.ijert.org Vol. 6 Issue 06, June – 2017

Samangouei P., Kabkab M., & Chellappa R., (2018) Defense-GAN: Protecting classifiers against adversarial attacks using generative models. ArXiv: 1805.06605, 2018.

Silver D., Huang A., Maddison C., Guez A., Sifre L., G. Driessche van den, J., Antonoglou I., Panneershelvam V., Lanctot M., Dieleman S., Grewe D., Nham J., Kalchbrenner N., Sutskever I., Lillicrap T., Leach M., Kavukcuoglu K., Graepel T., & Hassabis D., (2016) Mastering the game of go with deep neural networks and tree search. Nature, vol.529, no.7587, pp.484–489, 2016. DOI: 10.1038/nature16961.

Song Y., Kim T., Nowozin S., Ermon S., & Kushman N., (2017) Pixeldefend: Leveraging generative models to understand and defend against adversarial examples. ArXiv: 1710.10766, 2017.

Steinhardt J., Koh P., & Liang P., (2017) Certified defenses for data poisoning attacks. in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 1–13

Tramer F., Kurakin A., Papernot N., Goodfellow I., Boneh D., & McDaniel P., (2017) Ensemble adversarial training: Attacks and defenses. ArXiv: 1705.07204, 2017

Xie C., Wang J., Zhang Z., Ren Z., & Yuille A., (2018) Mitigating adversarial effects through randomization. In International Conference on Learning Representations.

Zantedeschi V., Nicolae M., & Rawat A., (2017) Efficient defenses against adversarial attacks. In Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, 39–49.

Zolotukhin M., Miraghaehi P., Zhang D., & Hamalainen T., (2022) On Assessing Vulnerabilities of the 5G Networks to Adversarial Examples. IEEE Access Digital Object Identifier 10.1109/ACCESS.2022.3225921