Using GAN for Measurement Report Generation to Overcome Limited Data for FBS Detection

Yeongshin Park, Bumjoon Kim, Bonam Kim, and Ilsun You*

Kookmin University, Seoul, Republic of Korea
{p17030508, bj1430}@kookmin.ac.kr, {kimbona9, ilsunu}@gmail.com

Abstract

This paper address the challenge of detecting False Base Stations (FBS) in situations with limited available data. Our approach involves the creation of synthetic Measurement Reports (MR) through Generative Adversarial Networks (GANs). The results of our experiments convincingly illustrate the practical application of GAN-generated data and its successful verification for similarity to genuine data using Convolutional Neural Networks (CNNs).

Keywords: 5G Security, False Base Station Detection, Measurement Report, GAN

1 Introduction

False Base Station (FBS) refers to malicious base stations emitting stronger signals than the surrounding ones, causing mobile devices to connect to them and enabling various attacks[1]. To address this issue, machine learning is occasionally used, using Measurement Report (MR) generated by User Equipment (UE) based on the signals from nearby base stations.

Machine learning requires a substantial amount of data. If the dataset is limited, the machine learning process may not perform adequately. However, obtaining a sufficient number of MRs for machine learning is not always possible. For instance, in cases where FBS exist for a short period and disappear, UE may not generate enough MRs. In such situations, machine learning may struggle to detect similar FBS attacks that could occur in the future.

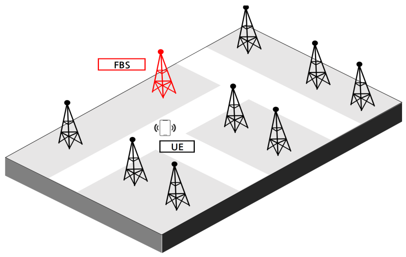
Therefore, in this paper, to address such situations, we generated MRs similar to actual ones using Generative Adversarial Network (GAN). GAN is a deep learning model in which two neural networks, a generator and a discriminator, compete with each other to generate realistic fake data[2]. The GAN-generated dataset includes MRs when FBS existed. We then used CNN to compare it with real MR data and evaluate its accuracy in detecting FBS.

2 Experiment

Dataset The dataset was created in the UERANSIM environment. One UE continuously generated MRs while moving at a constant speed among gNBs arranged in a 3x3 grid. The experiment lasted for one hour, resulting in 3602 MRs. For the actual input dataset to be used in the GAN, FBS had to be included. The base station in position 4 was designated as a FBS. Random values between 5 and 10 were selected and added to the RSRP values of the 4-th base station in all MRs.

The 7th International Conference on Mobile Internet Security (MobiSec'23), Dec. 19-21, 2023, Okinawa, Japan, Article No.P-15

*Corresponding author: Department of Financial Information Security, Kookmin University, Seoul, 02707, Republic of Korea, Tel: +82-02-910-5127



(a) Experiment Setup

65	69	59	66	68	59	57	58	55
65	70	60	69	68	60	57	58	55
65	70	60	65	68	60	57	58	55

(b) Real data

63	68	60	64	71	61	58	59	56
63	68	60	64	71	61	58	59	56
63	67	60	64	72	61	58	60	57

(c) GAN-generated data

Figure 1: Experiment Setup and Dataset

Experiment Procedure The GAN model in TensorFlow was used to generate and store data. Key steps including data loading and preprocessing, defining generator and discriminator models, GAN training, and evaluating the generated text. Initially, data was preprocessed in a 3x3 format. The model was trained by comparing generated text with real text, using activation functions like sigmoid and ReLU. Through the GAN, 10020 MRs were generated, and the GAN's discriminator mistakenly classified the generated data as 100% real. To assess if the GAN data was indistinguishable from real data, CNN was used for comparison.

Experiment Result In order to verify whether the data generated by GAN can actually be used for detecting FBS, the dataset was examined using Convolutional Neural Network (CNN). In both scenarios, the CNN utilized the same dataset, consisting of 36518 MRs for the normal data and 10020 for abnormal data. Consequently, when using CNN to detect FBSs, MRs containing actual FBSs achieved an accuracy of 92.4%, and the GAN-generated MRs achieved a 91.6% accuracy. By comparing the dataset generated by GAN with the real dataset, it can be observed that there is not a significant difference in accuracy.

3 Conclusion

In this paper, we investigated the use of GAN to generate a deficient dataset of FBSs for potential detection of real FBSs. Ultimately, we found that when using MRs generated by GAN and employing CNN for detection, the difference in accuracy compared to using actual MRs for detection was not significant. This suggests that GAN-generated data can indeed be effectively employed for this purpose.

Acknowledgments This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-00952)

References

[1] Prajwol Kumar Nakarmi, Jakob Sternby, and Ikram Ullah. Applying machine learning on rsrp-based features for false base station detection. https://doi.org/10.1145/3538969.3543787, 2022.

[2] Ian J. Goodfellow, Jean Pouget-Abadie, Bing Xu Mehdi Mirza, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. https://arxiv.org/abs/1406.2661, last viewed Jan 2014.