Mantik Journal, 8 (4) (2025), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Pattern recognition of 5G device serial number using K-Nearest Neighbors (K-NN) machine learning algorithm

Zulham Sitorus[1], Robin Antoni[2], Yohannes France Limbong[3], Jelly Rolley Sitompul[4], Sella Monika Br Tarigan[5]

[1,2,3,4,5]Master of Informatic Technology Department, University of Pembangunan Panca Budi, Medan, North Sumatera, Indonesia

A R T I C L E I N F O

Article history:

Received Jan 22, 2025
Revised Jan 28, 2025
Accepted Feb 04, 2025

Keywords:

5G Device Serial Number;
K-Nearest Neighbors (K-Nn) Algorithm;
Machine Learning;
Pattern Recognition.

ABSTRACT

5G networks are the latest generation of mobile communications technology that offer significant improvements in speed, capacity, and connectivity. However, along with the benefits it brings, it also brings a new set of challenges in the form of security breaches. Many 5G devices have been lost on-site. These devices are ABIA, AMIA, ASIB. Each of these devices has a serial number as identification data for each device. The rise of theft cases is due to the existence of collectors who are able to buy expensive stolen 5G devices for resale. So, the research will make the introduction of 5G device serial numbers using the Machine Learning (ML) with K-Nearest Neighbors (K-NN) algorithm. This pattern recognition is success to be done then become a guidance to recognizing stolen 5G devices. Next, this device cannot be used (deactivated) and be sold by system. This can break the demand and supply chain for stolen 5G devices. Based on the testing, there are 6 mismatches of 20 data testing or 70% data match.

Corresponding Author:

Jelly Rolley Sitompul,
Master of Informatic Technology Department
University of Pembangunan Panca Budi
Gatot Subroto Street, 4th, Medan, North Sumatera, 20122
Email: jelly.rolleys@gmail.com

1. INTRODUCTION

Fifth-generation (5G) networks are an evolution of wireless communication technology that promises significant improvements in speed, capacity, low latency, and spectrum efficiency (Dangi et al., 2022)(Kaur et al., 2021)(Storck & Duarte-Figueiredo, 2020). Every 10 years, mobile network technology migration takes place. By 2020, the use of 4th Generation (4G) technology will be almost ubiquitous in Indonesia (Mustar et al., 2019)(Nurcahya et al., 2020). Now it is time to enter the next stage, which is 5G-based communication connectivity.

Fifth-generation (5G) technology promises high speed, low latency, and large capacity to serve increasingly complex needs (Thantharate et al., 2022)(Lau et al., 2023)(Jazyah, 2023). However, security intrusions are still a priority that must be considered by stakeholders in 5G networks. Various security intrusions have occurred related to 5G networks such as intrusions by hackers/attackers (Mahmood et al., 2023)(Iavich et al., 2021), quality of service (QoS) intrusions (Abdellah et al., 2021), and

network intrusions connected to unmanned aerial vehicles (UAVs)(Shrestha et al., 2021). In addition, interference does not only occur in the system but can also occur in devices. Disruptions to devices can be in theft (Karnik et al., 2020), vandalism (Praveenchandar et al., 2023), dan sabotage (Kuruvatti et al., 2019).

At this time, there are frequent cases of theft of 5G devices, especially ABIA, AMIA, ASIB devices. This theft case still occurs because there is a high demand and supply for stolen 5G device. Security analysis of 5G networks is crucial to understand the threats that may arise and develop effective mitigation efforts (Zhang et al., 2021)(Kwon et al., 2021)(Wazid et al., 2021). The threat of 5G device theft can disrupt 5G communication network services. Network performance will drop and customers will leave providers who do not provide the best service for the community. The 5G device has each serial number. It indicates the device belongs to one of the providers and this is genuine. For this reason, something is needed to prevent the rampant theft cases. In this research, pattern recognition of serial number will be carried out as an identity for each 5G device. This recognition uses Machine Learning (ML) with the K-Nearest Neighbors (K-NN) algorithm.

The K-NN algorithm has often been used to solve pattern recognition such as Distributed Denial of Service (DDoS) attacks have disrupted the availability of 5G networks (Dong & Sarem, 2020)(Kilincer et al., 2021). Previous research related to pattern recognition of banknote serial number (Jang et al., 2022)(Choi et al., 2019). This research based on optical character recognition framework. The dataset used is 150,000 currency photos of 5 currencies, namely Won, US Dollar, Yen, Rupee, and Euro. There is a tradeoff between inference speed and serial number recognition accuracy. The serial number of industrial product recognition for baseline score and increases the efficiency of product tracing in industries (Hsu et al., 2022). Magnetic hard disk serial number recognition can improve the efficiency of magnetic information destruction (Xu et al., 2021).

ML with KNN algorithm has the potential to be developed as a promising system to overcome various problems that exist in society. It has the ability to learn from data, recognize patterns, and make predictions and decisions without human intervention (Bertolini et al., 2021)(Grierson et al., 2021)(Cuocolo et al., 2020).

2. RESEARCH METHOD

The research method used is a qualitative approach through literature studies. This approach allows researchers to collect and analyze data from various published sources, so as to provide a comprehensive overview of the topic to be studied. Furthermore, this method is developed with a qualitative descriptive design, where data collected from various literatures is statistically analyzed to identify trends and patterns in the 5G device serial number system. Qualitative descriptive research is suitable for providing an in-depth understanding of phenomena based on existing data. Data can be collected from various sources, namely from internal companies, goods suppliers, and the internet. This data collection method can be done based on extraction from existing data, questionnaires, direct surveys to the market, and searching for data on the internet. Data has a standardization that has been set by the manufacturer. This standard is used as a unique code that will be used as identification data for each 5G device.

Data Pre-Processing has been collected is then cleaned of duplicate data. The data should be consistent and have a variable scale that can be converted into a predefined format. Combining data from multiple data sources can provide a more comprehensive picture. However, this can be a challenge in terms of data consistency. Ensure data accuracy and consistency by crosschecking from various other data sources. Validating data by a team of experts, as well as using logic rules to detect and prevent data

inconsistencies. This validation requires a careful and systematic approach to ensure data quality and data reliability.

Machine Learning K-Nearest Neighbors (K-NN) algorithm has training data and testing data. The total amount of training data is 64 data which comes from various sources such as from extracting existing data, surveying the field, and searching the internet. This training data is the comparison data for each data tested. The amount of testing data is 20 data. They are shown in Figure 1, Figure 2, Figure 3 and Figure 4. Next, combination of these numbers are written into the excel file according to 84 combination numbers.

Figure 1. Photo of ASIB

Figure 2. Photo of AMIA

Figure 3. Photo of ABIA

unitName	productCode	Serial Number
ABIA	473096A.107	DH242103625
ABIA	473096A.107	DH223424433
ABIA	473096A.107	DH223424445
AMIA AirScale Indoor Subrack	473098A	1185332491
AMIA AirScale Indoor Subrack	473098A	1073575854
AMIA AirScale Indoor Subrack	473098A	1073451132
ASIB AirScale Common	473764A.102	DH242332709
ASIB AirScale Common	473764A.102	DH222444615
ASIB AirScale Common	473764A.102	EA224352630

Figure 4. Collection of serial number of ASIB, ABIA, AMIA

Pattern recognition experiments will be conducted with techniques such as feature extraction and classification performed on serial numbers that have been collected from 5G device serial number labels. Figure 4 is an example of a unique serial number code consisting of alphanumeric characters. The first step in the character extraction process is to detect the serial number region from prior knowledge of the serial number size and location. The region of interest (ROI) in the image is outlined using a bounding box. The ROI image with uneven illumination is then operated with a pre-processing method to minimize the grayscale variance and to remove noise in the image background.

Programming code is made in Matlab online version with link: https://matlab.mathworks.com/ Programming code contents are placed in the text box. Previously, the supporting files must be saved into the same folder as the matlab program code. In this research, supporting files consist of data_latih.xlsx, data_uji.xlsx, hasil_KNN.xlsx and matlab sourcecode is KNN.m. matlab GUI display form online version as shown in the figure 5.

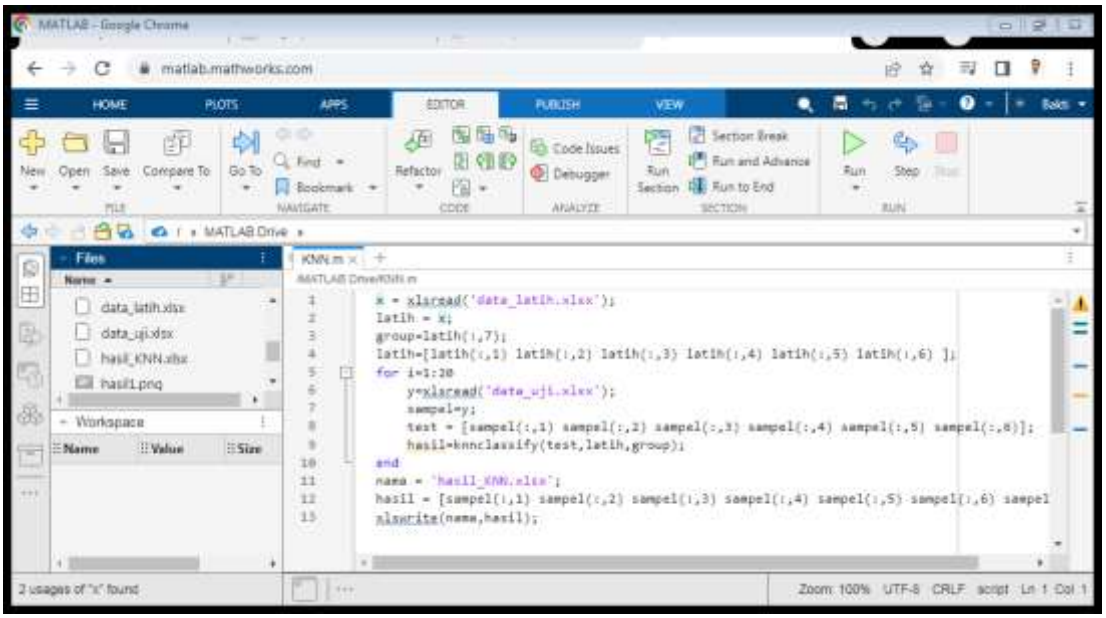

Figure 5. Matlab GUI display form

```
    x = xlsread('data_latih.xlsx');
    latih = x;
    group=latih(:,7);
    latih=[latih(:,1) latih(:,2) latih(:,3) latih(:,4) latih(:,5) latih(:,6) ];
    for i=1:20
       y=xlsread('data_uji.xlsx');
       sampel=y;
       test = [sampel(:,1) sampel(:,2) sampel(:,3) sampel(:,4) sampel(:,5)
sampel(:,6)];
       hasil=knnclassify(test,latih,group);
    end
    nama = 'hasil_KNN.xlsx';
    hasil = [sampel(:,1) sampel(:,2) sampel(:,3) sampel(:,4) sampel(:,5)
sampel(:,6) sampel(:,7)];
    xlswrite(nama,hasil);
```

Figure 6. Programming code

3. RESULTS AND DISCUSSIONS

As a qualitative approach, researchers have collected serial number data from various literature such as photos of devices from the field and warehouse, working contract files, brochures, and various other sources and internet as well. The researchers also collect and analyze data from various published sources to provide a comprehensive overview of 5G device serial number. Serial number has a standardization set by the manufacturer. This standard is used as a unique code that will be used as identification for each 5G device.

By qualitative descriptive approach is for providing an in-depth understanding of phenomena based on theft of devices. Theft is caused by economic factors and maybe a conspiracy with the company's own internal parties. For this reason, it is necessary to break the chain of theft through the pattern recognition of the device serial number.

Pre-processing data has been collected by making sequential combinations of numbers. The data should be consistent and have a variable number and next step into converted into a predefined format. For example, serial number of ABIA is DH242103625 suitable with list in Figure 4. Next step, this number is sorted as much as possible. Make sure that the data is consistent and accurate. The validation is needed to ensure data quality and realibility.

Classification used K-Nearest Neighbors (KNN) which according to 6 parameters. They are number of characters, number of letter and number combinations, DH character, EA character, 11 character, and 10 character (it shows the company code that makes the device). By clearance process, duplicate data and inconsistent data are removed, so that researchers get primary data as much as 64 training data and 20 testing data. Furthermore, the data normalization process is carried out by calculating the maximum and minimum values of training data and testing data. Data normalization is carried out based on existing parameters. Data normalization is the process of scaling the value so that the data has a close distance between one data and the other data. In this study, the results of data normalization are described in the value range 0 to 1 for each serial number.

The next step is to determine the k value and the Euclidean distance value for each data object. The value of k can be found using the square root calculation formula of the training data. The training data is 64, so the value of k = 8 and the distance is done by grouping into 4 groups. This distance value is used as the target value. Target value is to describe the highest matching value that can explain that the serial number is genuine. Target value is according to 0,1,2, and 3. Value of "3" is the highest value.

The next step is data clustering done by grouping data with 8 nearest neighbors based on Euclidean distance. The last step is to test the data with KNN programming using Matlab. Combination of these numbers are written into the excel file according to 84 combination numbers. The dataset is according to 84 combination numbers and target value as shown in Figure 7. The 84 number combinations (dataset) are then divided into 64 training data and 20 testing data. This numbers are saved into file "data_latih.xlsx" and "data_uji.xlsx". After program has running, the outputs are saved into file "hasil_KNN.xlsx". Based on the results, there are 6 target value mismatch of 20 data testing. The output is shown in Figure 8.

EA224352630	DH242332709	1073575854	1185332491	DH234224119	DH242103625	3
EA224352631	DH242332710	1073575855	1185332492	DH234224120	DH242103626	3
EA224352632	DH242332711	1073575856	1185332493	DH234224121	DH242103627	3
EA224352633	DH242332712	1073575857	1185332494	DH234224122	DH242103628	3
EA224352634	DH242332713	1073575858	1185332495	DH234224123	DH242103629	3
EA224352635	DH242332714	1073575859	1185332496	DH234224124	DH242103630	3
EA224352636	DH242332715	1073575860	1185332497	DH234224125	DH242103631	3
EA224352637	DH242332716	1073575861	1185332498	DH234224126	DH242103632	3
EA224352638	DH242332717	1073575862	1185332499	DH234224127	DH242103633	3
EA224352639	DH242332718	1073575863	1185332500	DH234224128	DH242103634	3
EA224352640	DH242332719	1073575864	1185332501	DH234224129	DH242103635	3
EA224352641	DH242332720	1073575865	1185332502	DH234224130	DH242103636	3
EA224352642	DH242332721	1073575866	1185332503	DH234224131	DH242103637	3
EA224352643	DH242332722	1073575867	1185332504	DH234224132	DH242103638	3
EA224352644	DH242332723	1073575868	1185332505	DH234224133	DH242103639	3

Figure 7. Dataset

EA224352630	DH242332709	1073575854	1185332491	DH234224119	DH242103625	3	3	
EA224352631	DH242332710	1073575855	1185332492	DH234224120	DH242103626	3	3	
EA224352632	DH242332711	1073575856	1185332493	DH234224121	DH242103627	3	3	
EA224352633	DH242332712	1073575857	1185332494	DH234224122	DH242103628	3	3	
EA224352634	DH242332713	1073575858	1185332495	DH234224123	DH242103629	3	2	Mismatch
EA224352650	DH242332729	1073575874	1185332511	DH234224139	DH242103645	2	2	
EA224352651	DH242332730	1073575875	1185332512	DH234224140	DH242103646	2	2	
EA224352652	DH242332731	1073575876	1185332513	DH234224141	DH242103647	2	2	
EA224352653	DH242332732	1073575877	1185332514	DH234224142	DH242103648	2	2	
EA224352654	DH242332733	1073575878	1185332515	DH234224143	DH242103649	2	0	Mismatch
EA224352670	DH242332749	1073575894	1185332531	DH234224159	DH242103665	1	1	
EA224352671	DH242332750	1073575895	1185332532	DH234224160	DH242103666	1	1	
EA224352672	DH242332751	1073575896	1185332533	DH234224161	DH242103667	1	1	
EA224352673	DH242332752	1073575897	1185332534	DH234224162	DH242103668	1	3	Mismatch
EA224352674	DH242332753	1073575898	1185332535	DH234224163	DH242103669	1	0	Mismatch
EA224352690	DH242332769	1073575914	1185332551	DH234224179	DH242103685	0	0	
EA224352691	DH242332770	1073575915	1185332552	DH234224180	DH242103686	0	0	
EA224352692	DH242332771	1073575916	1185332553	DH234224181	DH242103687	0	0	
EA224352693	DH242332772	1073575917	1185332554	DH234224182	DH242103688	0	2	Mismatch
EA224352694	DH242332773	1073575918	1185332555	DH234224183	DH242103689	0	3	Mismatch

Figure 8. Data output

4. CONCLUSSION

Pattern recognition using K-Nearest Neighbors (K-NN) algorithm has been done for 5G device serial numbers. The testing use 84 combination of serial number as primary data. This data are divided into 64 training data and 20 testing data. The testing was done using matlab program. Based on the result, there are 6 target value mismatches into 20 testing data. It means 6/20 equal 30% data is mismatch. The success of this program is 70% data matching. However, this research still has limitations, namely the amount of training data and testing data is still very small so that the percentage of success of the program still has to be tested again. Further research is reproduce primary data of 5G device serial numbers so that the program output can be more accurate. Serial number has been successfully recognized then become a guidance to recognizing stolen 5G devices. Next, this devices can not be used (deactivated) and be sold by system. It will be useful for breaking the chain of stolen goods.

REFERENCES

Abdellah, A. R., Mahmood, O. A., Kirichek, R., Paramonov, A., & Koucheryavy, A. (2021). Machine learning algorithm for delay prediction in IoT and tactile internet. *Future Internet*, *13*(12). https://doi.org/10.3390/fi13120304

Bertolini, M., Mezzogori, D., Neroni, M., & Zammori, F. (2021). Machine Learning for industrial applications: A comprehensive literature review. In *Expert Systems with Applications* (Vol. 175). https://doi.org/10.1016/j.eswa.2021.114820

Choi, E., Chae, S., & Kim, J. (2019). Machine learning-based fast banknote serial number recognition using knowledge distillation and bayesian optimization. *Sensors (Switzerland)*, *19*(19). https://doi.org/10.3390/s19194218

Cuocolo, R., Caruso, M., Perillo, T., Ugga, L., & Petretta, M. (2020). Machine Learning in oncology: A clinical appraisal. In *Cancer Letters* (Vol. 481). https://doi.org/10.1016/j.canlet.2020.03.032

Dangi, R., Lalwani, P., Choudhary, G., You, I., & Pau, G. (2022). Study and investigation on 5g technology: A systematic review. In *Sensors* (Vol. 22, Issue 1). https://doi.org/10.3390/s22010026

Dong, S., & Sarem, M. (2020). DDoS Attack Detection Method Based on Improved KNN with the Degree of DDoS Attack in Software-Defined Networks. *IEEE Access*, *8*.

https://doi.org/10.1109/ACCESS.2019.2963077
Grierson, D., Rennie, A. E. W., & Quayle, S. D. (2021). Machine Learning for Additive Manufacturing. *Encyclopedia*, *1*(3). https://doi.org/10.3390/encyclopedia1030048
Hsu, M. M., Wu, M. H., Cheng, Y. C., & Lin, C. Y. (2022). An Efficient Industrial Product Serial Number Recognition Framework. *Proceedings - 2022 IEEE International Conference on Consumer Electronics - Taiwan, ICCE-Taiwan 2022*. https://doi.org/10.1109/ICCE-Taiwan55306.2022.9869266
Iavich, M., Iashvili, G., Avkurova, Z., Dorozhynskyi, S., & Fesenko, A. (2021). Machine Learning Algorithms for 5G Networks Security and the Corresponding Testing Environment. *CEUR Workshop Proceedings*, *3187*.
Jang, W., Lee, C., Jeong, D. S., Lee, K., & Lee, E. C. (2022). Multi-Currency Integrated Serial Number Recognition Model of Images Acquired by Banknote Counters. *Sensors*, *22*(22). https://doi.org/10.3390/s22228612
Jazyah, Y. H. (2023). 5G Security, Challenges, Solutions, and Authentication. *International Journal of Advances in Soft Computing and Its Applications*, *15*(3). https://doi.org/10.15849/IJASCA.231130.04
Karnik, A., Adke, D., & Sathe, P. (2020). Low-Cost Compact Theft-Detection System using MPU-6050 and Blynk IoT Platform. *2020 IEEE Bombay Section Signature Conference, IBSSC 2020*. https://doi.org/10.1109/IBSSC51096.2020.9332214
Kaur, J., Khan, M. A., Iftikhar, M., Imran, M., & Emad Ul Haq, Q. (2021). Machine Learning Techniques for 5G and beyond. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3051557
Kilincer, I. F., Ertam, F., & Sengur, A. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. *Computer Networks*, *188*. https://doi.org/10.1016/j.comnet.2021.107840
Kuruvatti, N. P., Hernandez, R., & Schotten, H. D. (2019). Interference Aware Power Management in D2D Underlay Cellular Networks. *IEEE AFRICON Conference*, *2019-September*. https://doi.org/10.1109/AFRICON46755.2019.9133851
Kwon, S., Park, S., Cho, H. J., Park, Y., Kim, D., & Yim, K. (2021). Towards 5G-based IoT security analysis against Vo5G eavesdropping. *Computing*, *103*(3). https://doi.org/10.1007/s00607-020-00855-0
Lau, I., Ekpo, S., Zafar, M., Ijaz, M., & Gibson, A. (2023). Hybrid mmWave-Li-Fi 5G Architecture for Reconfigurable Variable Latency and Data Rate Communications. *IEEE Access*, *11*. https://doi.org/10.1109/ACCESS.2023.3270777
Mahmood, I., Alyas, T., Abbas, S., Shahzad, T., Abbas, Q., & Ouahada, K. (2023). Intrusion Detection in 5G Cellular Network Using Machine Learning. *Computer Systems Science and Engineering*, *47*(2). https://doi.org/10.32604/csse.2023.033842
Mustar, M. Y., Chamim, A. N. N., Putra, K. T., Nugraha, V. D. H., & Jusman, Y. (2019). Analysis and Optimization of 4G LTE Network in Jombang City East Java. *Journal of Electrical Technology UMY*, *3*(1). https://doi.org/10.18196/jet.3148
Nurcahya, Y. A., Ismawati, E., & … (2020). Peran Religiusitas Dan Batasan Waktu Audit Terhadap Efektivitas Fungsi Internal Audit Di Bmt Karisma Magelang. *Nominal Barometer Riset …*.
Praveenchandar, J., Kumar, S. V., Paul, A. C., Mukunthan, M. A., & Maharajan, K. (2023). Deep Learning Algorithms in Mobile Edge with Real-Time Abnormal Event Detection for 5G-IoT Devices. *International Journal of Interactive Mobile Technologies*, *17*(17). https://doi.org/10.3991/ijim.v17i17.42805
Shrestha, R., Omidkar, A., Roudi, S. A., Abbas, R., & Kim, S. (2021). Machine-learning-enabled intrusion detection system for cellular connected uav networks. *Electronics (Switzerland)*, *10*(13). https://doi.org/10.3390/electronics10131549
Storck, C. R., & Duarte-Figueiredo, F. (2020). A Survey of 5G Technology Evolution, Standards, and Infrastructure Associated with Vehicle-to-Everything Communications by Internet of Vehicles. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3004779
Thantharate, A., Tondwalkar, A. V., Beard, C., & Kwasinski, A. (2022). ECO6G: Energy and Cost Analysis for Network Slicing Deployment in Beyond 5G Networks. *Sensors*, *22*(22). https://doi.org/10.3390/s22228614
Wazid, M., Das, A. K., Shetty, S., Gope, P., & Rodrigues, J. J. P. C. (2021). Security in 5G-Enabled Internet of Things Communication: Issues, Challenges, and Future Research Roadmap. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2020.3047895
Xu, Z., Liu, X., Tang, J., Li, P., & Zhang, Z. (2021). Magnetic Hard Disk Serial Number Recognition

Method Based on Machine Vision. *Proceedings of the 33rd Chinese Control and Decision Conference, CCDC 2021*. https://doi.org/10.1109/CCDC52312.2021.9601536

Zhang, R., Zhou, W., & Hu, H. (2021). Towards 5G Security Analysis against Null Security Algorithms Used in Normal Communication. *Security and Communication Networks*, *2021*. https://doi.org/10.1155/2021/4498324