**Journal of Telematics and Informatics (JTI)**
Vol.7, No.3, September 2019, pp. 158~165
ISSN:2088-8708


# SIMBOX IDENTIFICATION USINGK-NEAREST NEIGHBOR BASED ON SPECTRUM ANALYZER

***Agung Suryowibowo, **Imam Much Ibnu Subroto,* Eka Nuryanto Budi Susila**
* Electrical Engineering Departement, Universitas Islam Sultan Agung,Indonesia
Kaligawe Raya Street Km. 4, Terboyo Kulon, Genuk, Semarang, Central Java - Indonesia
Email: suryowibowoagung@yahoo.co.id

***ABSTRACT***

Telecommunication Service Provider should deal with illegal players (grey operators) who do not have permission to conduct international voice service. These illegal players perform their activities by passing international incoming traffic using Simbox devices. To identify visual simbox usage is very difficult and less reliable, therefore by using spectrum analyzer and K-Nearest Neighbor (K-NN) method is one way to identify simbox usage. The attributes used in the identification process are Location / Document, Strong Frequency Signal, and by applying K-NN algorithm based on proximity of training data with data testing. The determination of this attribute is based on GSM DCS 1800 MHz uplink frequency measurement in Cilacap and Banyumas area. The identification process was conducted on six frequencies points on 18 data with the largest signal strength as training data. Moreover, the signal strength data testing by using 32 data gives result 81.25% accuracy. The results of K-NN algorithm calculations can be implemented to identify the use of simbox, hence it can be used as a reference for mobile operators to identify simbox usage in other areas.

***Keywords:*** *International Termination of Traffic Refilling (RTTI), K-Nearest Neighbor (K-NN), Identification of Simbox Usage.*

## 1. Introduction

In running an Incoming International business a Telecommunication Service Provider has many challenges to face. In addition to competing with other international service providers, Telecommunications Service Provider must deal with gray players who do not have permission to operate international voice service. These illegal players perform by passing the international incoming traffic that should be through the International Gate of the International (SGI) owner of the international voice service license, through the illegal VOIP line. The practice of bypassing international incoming traffic is done by Fraudsters (for frauders) to benefit by exploiting different international incoming rates with retail tariffs, especially on-net retail tariffs (calls between customers within the same carrier). The practice is carried out by channeling international incoming traffic from abroad via the VOIP line and then into a device that has the ability to switching at once redial by using Subscriber Identification Module (SIM) Card operators tailored to the purpose of the call [1][2]. The device is known as SIMBOX which contains several SIM cards and functions as a traffic receiver, performs a switch function, and performs redial[3][4]. With the practice of International Termination of Traffic Refilling (RTTI) which is an illegal business, Telecommunication Service Providers face business risks in the form of Lost Opportunity Revenue from international incoming services through VOIP and will threaten the growth of other Clear Channel services[5][6].

## 2. Research Methods

The research method applied in this research can be seen in figure 1.

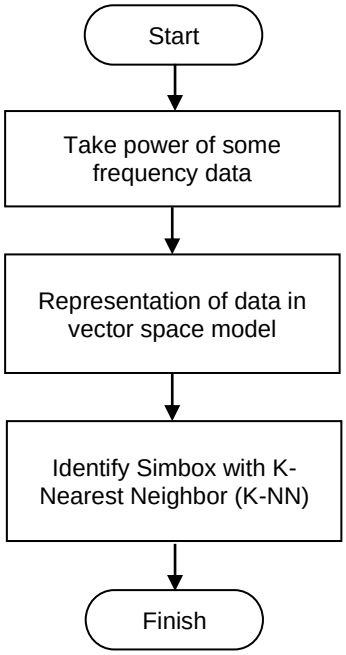


Figure 1. Research Methodology

Based on the measurement of GSM 1800 MHz uplink frequency using spectrum analyzer, six frequencies with the largest signal strength are obtained as data representation in the form of vector space model (data training and data testing)[7][8][9]. Then calculated using the algorithm K- Nearest Neighbor to identify the use of simbox in the area that has been determined.

**K-Nearest Neighbor (K-NN)**

The K-Nearest Neighbor (K-NN) algorithm is a method that uses a supervised algorithm. K-NN includes an instance-based learning group. This algorithm is also one of the lazy learning techniques. K-NN is done by searching for group K objects in the closest training data (similar) to objects in new data or data testing [3].

In general to define the distance between two objects A and B, the Euclidean Distance formula is used in equation 1.

$$D_{AB} = \sqrt{\sum_{i=1}^{n} (X_{A,i} - X_{B,i})^2} \quad \text{..... (1)}$$

Where the matrix $D_{AB}$ is the scalar distance of the two vectors $A$ and $B$ of the matrix with dimension n dimensions[12]. The K-NN algorithm can generally be illustrated with Flowchart in Figure 2.

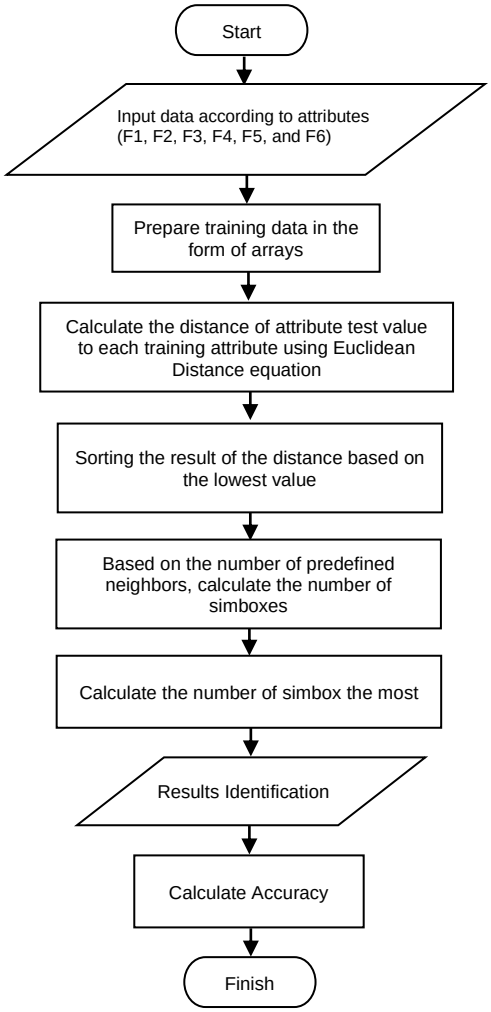


Figure 2. Flowchart K-NN

Enter data according to attribute of data of six frequency with biggest signal strength (F1, F2, F3, F4, F5, and F6), then prepare training data in the form of array which then calculate the distance of data attribute attribute value to each attribute of training data using equation Euclidean Distance. The next step classifies the result of the distance based on the lowest value (rank). After that calculate the number of simbox based on the number of neighbors that have been determined, then calculate the most simbox so as to generate simbox user identification by knowing the accuracy of the test.

**Testing Accuracy**

Confusion matrix is a tool used to evaluate the classification model to estimate the correct or false object. A matrix of prediction that will be compared with the original class of input or in other words contains actual and predicted value information on the classification.

As for precision level calculation, recall, and accuracy on confusion matrix:

$$Precision = \frac{Corect\ result}{(Corect\ result + Unexpected\ result)} \times 100\% \quad \text{..... (2)}$$

$$Recall = \frac{Corect\ result}{(Corect\ result + Missing\ result)} \times 100\% \quad \text{..... (3)}$$

$$Accuracy = \frac{Corect\ result + Corect\ absence\ of\ result}{(sum\ of\ all\ result)} \times 100\% \quad \text{..... (4)}$$

## 3. Results and Discussion

In accordance with the research methodology in the previous chapter, several important things to be done in completing this study consist of measurement, data collection, K-NN calculations, analysis, and conclusions.

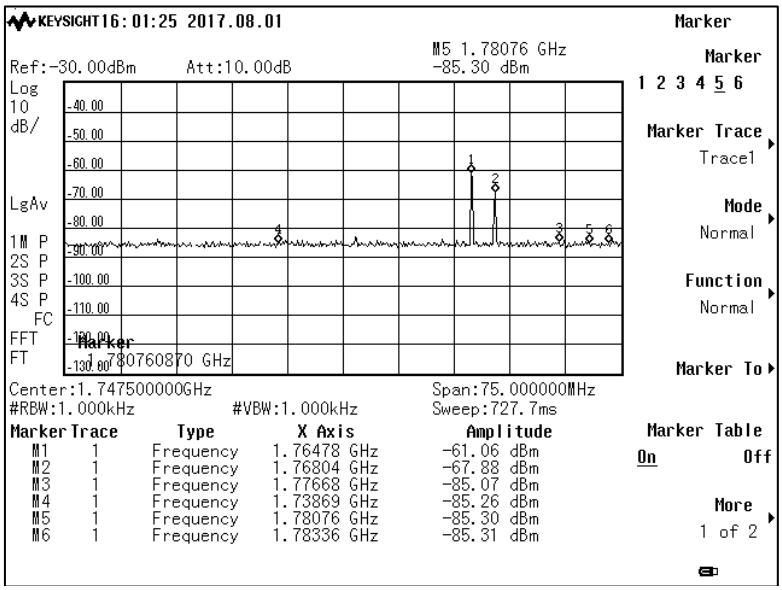

Figure 3. Example display capture of spectrum analyzer

Figure 3 is an Example display capture of spectrum analyzer, the result of measuring radio frequency uplink taken as many as 6 radio frequencies (Freq 1, Freq 2, Freq 3, Freq 4, Freq 5, and Freq 6) by measuring the signal strength at each measurement location[10][11].

**Technical Parameter Measurement**

The measurement of this technical parameter is carried out by measuring the radio frequency of the GSM uplink at the 1800 MHz DCS frequency used by the mobile user. The measurements were conducted in 50 different locations located in some areas of Cilacap and Banyumas districts[7][9][13]. The results of the measurements are in Table 1.

Table 1. Measurement Results

| No. | Location | | Measurement results | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | F1 | F2 | F3 | F4 | F5 | F6 |
| 1. | Location 1 | Freq. (MHz) | 1738.67 | 1764,78 | 1768,04 | 1776,68 | 1780,76 | 1783,36 |
| | | Level (-dBm) | 85,26 | 61,06 | 67,88 | 85,07 | 85,3 | 85,31 |
| 2. | Location 2 | Freq. (MHz) | 1752,39 | 1764,61 | 1766,08 | 1769,51 | 1772,28 | 1777,98 |
| | | Level (-dBm) | 77,14 | 54,72 | 85,79 | 59,08 | 77,02 | 80,95 |
| 3. | Location 3 | Freq. (MHz) | 1755,16 | 1762,98 | 1767,06 | 1769,51 | 1777,66 | 1784,18 |
| | | Level (-dBm) | 81,40 | 76,22 | 56,90 | 70,92 | 85,61 | 66,08 |
| .... | .... | .... | .... | .... | .... | .... | .... | .... |
| 50. | Location 50 | Freq. (MHz) | 1761,52 | 1763,31 | 1764,78 | 1772,11 | 1774,72 | 1783,69 |
| | | Level (-dBm) | 64,67 | 67,33 | 61,75 | 71,94 | 68,27 | 59,36 |

**Data Collection**

As a method of robust data collection of DCS 1800 MHz cellular frequency signals which is the main basis in this identification process is to use the measurement results. Measurements were carried out at 50 sites in Cilacap and Banyumas districts of Central Java Province.

Data cleaning is done to reduce noise effects during the calculation process and eliminate unused attributes. Furthermore the process of normalizing the data. The result of normalization of data can be seen in Table 2.

Table 2. Normalization of Data

| No. | Document | Frequency with the largest signal strength **(-dBm)** | | | | | |
|---|---|---|---|---|---|---|---|
| | | F1 | F2 | F3 | F4 | F5 | F6 |
| D1 | Location 1 | 85,26 | 61,06 | 67,88 | 85,07 | 85,30 | 85,31 |
| D2 | Location 2 | 77,14 | 54,72 | 85,79 | 59,08 | 77,02 | 80,95 |
| D3 | Location 3 | 81,40 | 76,22 | 56,90 | 70,92 | 85,61 | 66,08 |
| .... | .... | .... | .... | .... | .... | .... | .... |
| D50 | Location 50 | 64,67 | 67,33 | 61,75 | 71,94 | 68,27 | 59,36 |

**Calculation of K-NN**

Before doing the calculation process K-NN need to be prepared data training, and data testing and determine the parameter K as the number of nearest neighbors, where the value of K is 3, 5 and 7.

Table 3. Data *Training*

| No. | Document | Frequency with the largest signal strength (-dBm) | | | | | | Category |
|---|---|---|---|---|---|---|---|---|
| | | F1 | F2 | F3 | F4 | F5 | F6 | |
| D1 | Location 1 | 85,26 | 61,06 | 67,88 | 85,07 | 85,3 | 85,31 | Normal |
| D2 | Location 2 | 77,14 | 54,72 | 85,79 | 59,08 | 77,02 | 80,95 | Normal |
| D3 | Location 3 | 81,40 | 76,22 | 56,90 | 70,92 | 85,61 | 66,08 | Normal |
| .... | .... | .... | .... | .... | .... | .... | .... | .... |
| D18 | Location 18 | 74,55 | 58,63 | 57,55 | 72,95 | 62,43 | 70,00 | Simbox |

A total of 18 data from 50 overall data were used as training data on simulation on K-NN method on measurement result. The data has been normalized to eliminate frequencies, and uses the largest signal strength at 6 measuring frequencies.

Table 4. Data *Testing*

| No. | Document | Frequency with the largest signal strength (-dBm) | | | | | | Prediction |
|---|---|---|---|---|---|---|---|---|
| | | F1 | F2 | F3 | F4 | F5 | F6 | |
| D19 | Location 19 | 84,65 | 85,56 | 84,68 | 85,13 | 74,55 | 73,46 | Normal |
| D20 | Location 20 | 85,56 | 74,55 | 84,65 | 85,13 | 84,68 | 73,46 | Normal |
| D21 | Location 21 | 78,00 | 75,04 | 59,34 | 84,27 | 75,24 | 79,21 | Simbox |
| .... | .... | .... | .... | .... | .... | .... | .... | .... |
| D50 | Location 50 | 64,67 | 67,33 | 61,75 | 71,94 | 68,27 | 59,36 | Simbox |

Table 4 shows 32 data testing as K-NN testing. From this data will be generated the conclusion of a data will enter into simbox, normal, or based on the value of adjacent that have been processed.

Furthermore, the calculation of K-NN is done by using Euclidean Distance equation. For example the calculation of new data to be calculated is the first data with Location 19. The process of calculation is done kesetiap data testing so that later will result in distance value in accordance with the amount of data testing. From the calculation results obtained distance as in Table 5.

Table 5. Distance Calculation Result

| Data Testing | Data Training | | | | | | | | | | | | | | | | | | Prediction | Actual Value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | D11 | D12 | D13 | D14 | D15 | D16 | D17 | D18 | | |
| D19 | 33,75 | 41,83 | 35.33 | 46,53 | 34,77 | 52,77 | 31,75 | 32,34 | 23,35 | 64,66 | 34,74 | 46,54 | 21,07 | 32,1 | 29,33 | 52,62 | 23,16 | 43,25 | Normal | Normal |
| D20 | 24,58 | 35,48 | 32,36 | 42,92 | 33,27 | 52,7 | 26,07 | 28,36 | 23,15 | 64,14 | 27,57 | 46,66 | 30,01 | 34,0 | 28,45 | 53,23 | 22,46 | 42,01 | Normal | Normal |
| D21 | 21,45 | 41,88 | 21,84 | 33,06 | 21,31 | 41,95 | 19,55 | 22,25 | 15,74 | 47,36 | 25,99 | 35,78 | 34,78 | 24,44 | 37,10 | 40,50 | 13,06 | 25,72 | Simbox | Normal |
| .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D50 | 40,56 | 40,01 | 27,01 | 12,36 | 28,09 | 19,12 | 31,76 | 21,71 | 25,39 | 26,2 | 34,12 | 10,03 | 45,62 | 27,92 | 37,17 | 22,39 | 34,91 | 18,42 | Simbox | Simbox |

Based on a predetermined K value of 3, 5, and 7, the distance value used is if K = 3, then taken is 3 smallest. So also if the value of K = 5, and if the value of K = 7.

Table 6. Comparison of K-NN Identification results

| Document | Prediction | Identify K-NN | | |
|---|---|---|---|---|
| | | K = 3 | K = 5 | K = 7 |
| D19 | Normal | Normal | Normal | Normal |
| D20 | Normal | Normal | Normal | Normal |
| D21 | Simbox | Normal | Normal | Simbox |
| .... | .... | .... | .... | .... |
| D50 | Simbox | Simbox | Simbox | Simbox |

Based on the comparison of the results of the identification in Table 6, the calculated simbox-simbox and normal-normal results are then divided by the amount of data testing and multiplied by 100% to obtain the accuracy of the K-NN identification results in Table 7.

Table 7. Comparison of K-NN Identification Accuracy

| Percentage | Identify K-NN | | |
|---|---|---|---|
| | K=3 | K=5 | K=7 |
| Precision | 60 % | 54,55 % | 70 % |
| Recall | 66,67 % | 66,67 % | 70 % |
| Accuracy | 78,13 % | 75 % | 81,25 % |

The comparison of K-NN identification accuracy in Table 7 on the three parameters of K was found that the highest accuracy rate was 81.25% at K = 7, while the lowest accuracy level at K = 5 was 75%.

## 4. Conclusions

Identify the use of simbox against training data and data testing using parameter K with value K = 3, K = 5, and K = 7 there are documents identified as simbox. The calculation of K-NN algorithm applied to dedicate the accuracy level of identification of simbox usage on indigo K = 7 with highest yield of 81,25%. According to the results of identification of the use of simbox-based spectrum analyzer using K-NN identified as many as 10 documents using simbox. Suggestions that can be given for further research that is related to the process of analysis used in the process of identifying the use of simbox-based spectrum analyzer can be done with several other classification methods to find out the best accuracy of several algorithms in the same case.

## Reference

[1] Hiyam Ali El Tawashi. *Detecting Fraud in Cellular Telephone Network.* Thesis. Gaza: Faculty of Commer, Gaza Islamic University. 2010.

[2] Ilona Murynets, Michael Zabarankin, Roger Piqueras Jover, dan Adam Panagia. *Analysis and Detection of SIMbox Fraud in Mobility Networks*.

International Journal online. http://www.research.att.com/techdocs/TD_101254.pdf

[3] A. H. Elmi, S. Ibrahim, and R. Sallehuddin. *Detecting sim box fraud using Support Vector Machine abd Artifical Neural Network*. International Journal online in IT Convergence and Security 2012. Springer, 2013, pp. 575–582. http://link.springer.com. 2012.

[4] Mustakim, Giantika Oktaviani F. K-Nearest Neighbor Classification Algorithm As Prediction Prediction System of Student Achievement. Journal of Science, Technology and Industry, Vol 13, pp 195-202. 2016.

[5] Firdaus Fadzil. Illegal By Pass For International Calls : Industry Position. July 2103.

[6] Widhiatmoko, Hesti Susilawati, Rahmat K. Analysis of VoIP (Voice over Internet Protocol) Performance in WiMAX (Worldwide Interoperability for Microwave Access) Network In DKI Jakarta. Journal : Electrical Engineering Study Program, Faculty of Science and Engineering, University of General Soedirman.

[7] Denny Setiawan. The Development of Wireless Telecommunication Technology and Challenges For Indonesia Ahead. Ministry of Communications and Information Technology DG of SDPPI. 2014.

[8] Widi Amanasto. Cloud Computing Business Opportunities and Regulatory Challenges.ICT Regulatory Management Telkom Indonesia. 2012.

[9] GSMA. 2G/2.5G/3G Roaming. Official Document: IR.50 GSM Association. 2006.

[10] Yi-Bing Lin, Imrich Chlamtac. *Wireless and Mobile Network Architectures*. John Wiley & Sons, INC. Replika Press Pvt. Ltd, Kundli. 2005.

[11] Jorg Eberspacher, Hans-Jorg Vogel. *GSM Switching, Services and Protocols*. John Wiley & Sons. Biddles Ltd, Guildford, UK.1999.

[12] Imam Much Ibnu Subroto. *K-Nearest Neighbor*. Department of Informatics Engineering, Faculty of Industrial Technology, Unissula.

[13] http://disnakertrans.jatengprov.go.id/databidang/penempatan