

Suspicious Call Detection Using Bayesian Network Approach

SYED MUHAMMAD AQIL BURNEY, QAMAR UL ARIFEEN, NADEEM MAHMOOD, SYED
ABDUL KHALIQ BARI

Department of Computer Science, University of Karachi, Main University Road, PAKISTAN
burney@uok.edu.pk, arifeen@uok.edu.pk, nmahmood@uok.edu.pk, akbari@uok.edu.pk
<http://www.uok.edu.pk>

Abstract: - The rapid advancements in the fast and rapidly growing field of information and communication technologies (ICT's) has lead us to new thinking paradigm and it is being implemented in all walks of life including business, finance, health, management, engineering, basic sciences, sports, social sciences and many other domains. There are many advantages of speedy growth of internet and mobile phones in the society and people are taking full advantage of them. However this technology is widely used by the criminals for the execution of criminal or terrorist activities. The state of Pakistan is going through a period where they are crushing criminal and terrorist networks throughout the country. There are number of terrorist and criminal activities in the last few years. This study focuses on the use of related equipment like mobile phone, SIM's etc. in criminal or terrorist activities. We have analyzed call detail records (CDR) collected from tower data of five mobile companies by using geo-fencing approach. The classification of suspicious call detection and identification is done by using Bayesian classifier approach. In the research document, the researcher exhibits approaches for establishing Bayesian systems from previous information and review Bayesian techniques for improvisation of these models. We encapsulate approaches for structuring and learning parameters in Bayesian system, including various methods to learn from Bayesian database.

Key-Words: - Information and communication technology (ICT), Bayesian Network, call data record (CDR), criminal network, geo-fencing, suspicious call detection, geo-fence.

1. Introduction¹

The use of mobile telecommunication network and related services for any sort of criminal or terrorist activities is illegal under the cyber law of Pakistan. With the passage of time and due to the advancements in mobile phone technology the criminals and their networks [1] are very effectively using modern mobile phone technology in their illegal and criminal activities. Therefore, investigators require more knowledge of the latest tools and techniques to identify criminal by tracing his mobile phone record and to identify and apprehend the culprit.

Modern technology is being used by criminals very effectively in their suspicious activities. Therefore need of comprehensive knowledge and prompt access to the data/ information and latest tools and techniques is utmost required for the investigators to cope with criminals' ulterior motives. Computer/mobile technologies are being improved by every dawn introducing new prospects and models in a very compactly designed with high amenities of transferring data through cell phones from various modes.

¹ Data Courtesy: CPLC

Table 1: Snapshot of call data record (CDR) [1]

Call Type	Calling Party	Called Party	Date and Time	Duration (in Seconds)	IMSI	IMEI No
Incoming	923021111111	923020000003	2/1/2014 9:10	186	#####	1010101010101
Incoming	923021111111	923020000000	2/1/2014 10:19	55	#####	1010101010101
Incoming	923021111111	923132000000	2/1/2014 10:41	49	#####	1010101010101
Incoming	923021111111	923472020202	2/1/2014 10:56	56	#####	1010101010101
Outgoing	923021111111	923020000000	2/1/2014 11:02	26	#####	1010101010101

Use of mobile phones emerged as an inevitable necessity in every walk of life round the globe. Fast computational capability of devices and a variety of applications in cellular technology encourages criminals to use it in variety of ways, as well as a shield against Law Enforcement Agencies (LEAs), and appeals them to perform covert activities. The increase in digital crimes based on cell phones [2] and internet the amendments are being made in the existing laws of the country and new cyber and internet laws are implemented for controlling such crimes. A new field has emerged as digital forensics [3] and lot of such evidence is being collected and also admissible in the court of law [4] [5] [6] [7] [8].

The Law Enforcement Agencies (LEA) of Pakistan has learned that illegal cell phone SIMs from neighbor country are being smuggled and used for most of the criminal activities in Pakistan Martin. The government of Pakistan has instructed Pakistan Telecommunication authority (PTA) to stop this illegal traffic [9] [10], however this is difficult due to various dilemmas. SIMs are illegally transferred to Pakistan and are easily available in connecting province and other parts of the country. Criminals extensively make use of these SIMs in kidnapping for ransom, terrorism, extortion and other related crimes. The Law Enforcement Agencies are using location based tracking [11] [12].

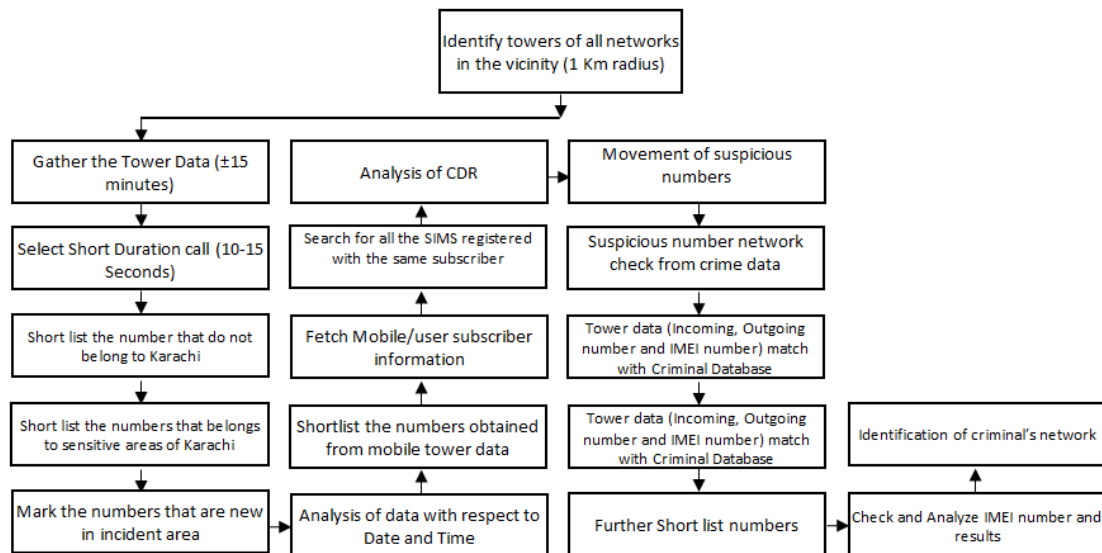


Fig. 1: Conceptual Model for Analysis of Cell data

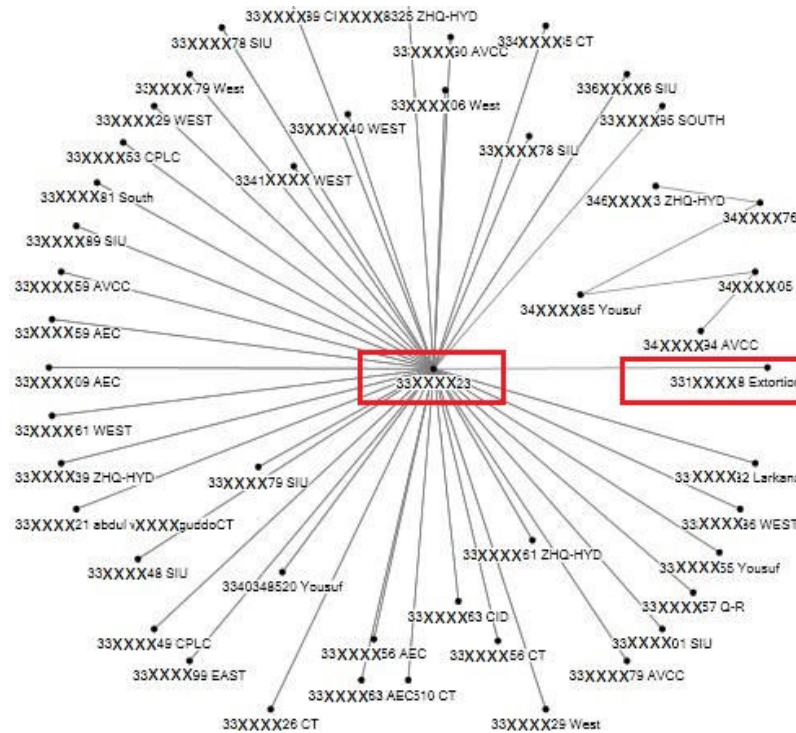


Fig. 2: Result of CDR analysis [1]

Geo-fencing help the LEAs to better understand the network of calls both ends (receiving and calling) from the venue where a crime incident or terrorist activity has taken place. Usually the tower data is collected from the radius of 1 Kilometer from the point of incident (crime) with approximately 15 minutes from the incident time. It can help track down the origin of suspicious voice calls specially the offenders that are responsible for these crimes. Although the geo-location databases have improved in recent past but the location accuracy isn't perfect [13].

We applied location based approach using geo-fencing technique for mobile phone crimes through the data available from service providers [14] [15] [16] [17] [18], which carried out using tower data analysis [19] which helps a lot in detecting the terrorists and their systems in Pakistan. We have

made use of soft computing methods/techniques such as probabilistic reasoning, Bayesian inference and rough set theory. Naïve Bayesian approach is used for classification of crime [20] [21]. We studied these approaches using available dataset on Pakistan.

This paper is organized in 6 sections. Section 2 outlines analyzing tower data used in identifying a suspicious criminal. Section 3 discusses the conceptual model for the identification suspects through tower data. In Section 4 we mentioned the process of constructing typical Bayesian structural data followed by the section 5 and 6 which are results and conclusion.

2. Analysis of Cell (Tower) Data

While investigating an incident, analysis of cell data is the key which can lead to explore the constituents

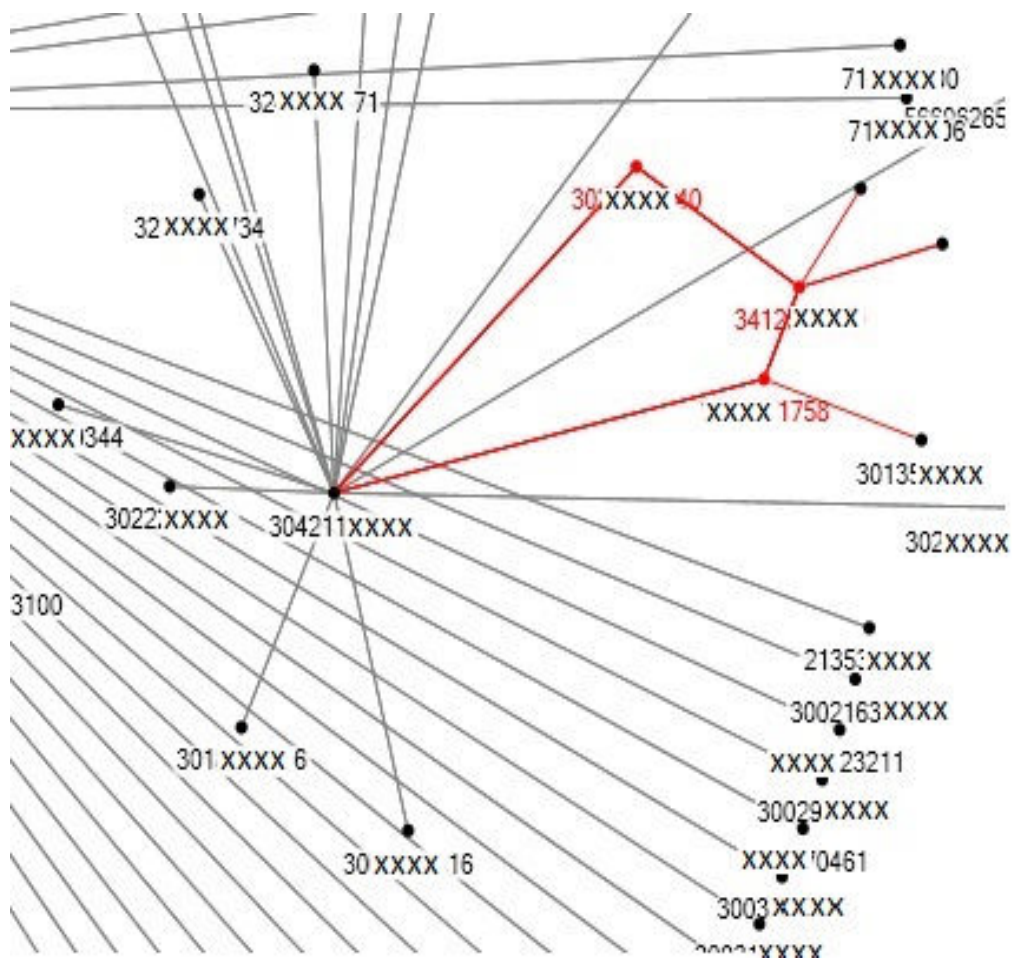


Figure 3: Analysis of CDR [1]

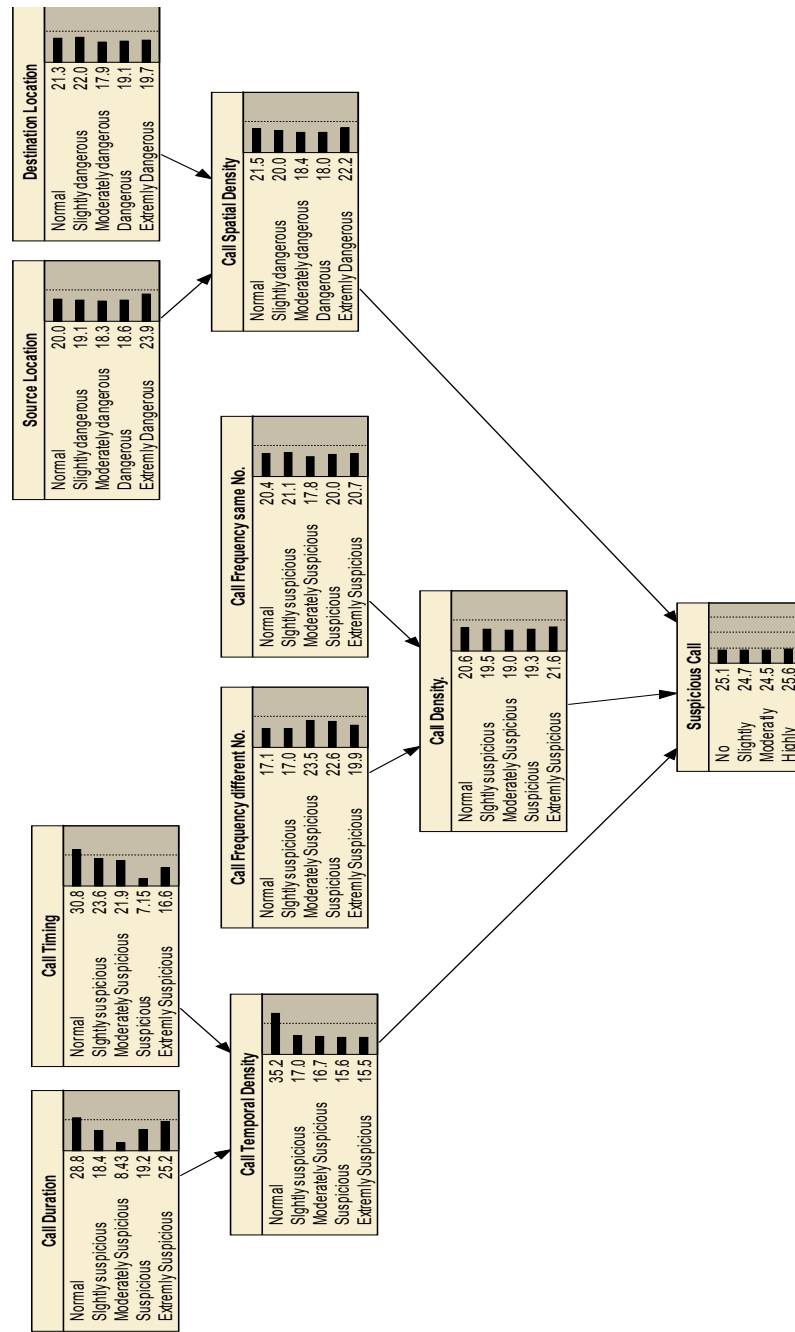


Figure 4: BN Classifier for Suspicious call identification [1]

used in incident and their operations. Given below is the proposed diagram for the analysis of tower data. It shows the steps involved in geo-fencing of the area where a crime or terrorist activity took place to the analysis of CDR and to identify, monitor and track the call data of the suspects along with the area where they are located.

3. Identifying a suspicious criminal

It is observed that the features and call patterning of a suspicious criminal distinct from regular call receivers. For example, a criminal generally works with a network that is prepaid. Terrorists are usually present in the locations where these elements are found. The LEAs contain a complete record of these areas and call data record of suspects for scanning data in any criminal or terrorist act. Suspects often make short duration calls which may last less than 60 seconds and in many cases it lasts around 15 seconds.

Its reason is that during the period of carrying out a terrorist attack or a criminal activity, they synchronize their plan without long time consuming chat. Further, if a suspected person calls or receives a call from the suspicious places during a criminal activity or a terrorist attack (i.e., 60 minutes before or after); this marks him a suspicious person in a criminal activity. The match of call by this person with the criminals' database strengthens the belief about this criminal to be useful in the investigation of that specific criminal activity.

If a suspicious criminal does not come under the umbrella of such things is disqualified as a suspicious person and hence, not observed for more findings but if a suspected individual is seen as a terrorist, the call data records are scrutinized for further information. This support to identify most frequent callers and receivers. The following table shows the records of calls made and received.

Table 2: Detailed record of a suspect

Called Party	Incoming	Incoming SMS	Outgoing	Outgoing SMS	Grand Total
923020000000	157	69	59	11	296
923020000001	28	17	4	4	53
923020000003	37	21	9	12	79
923063131313	10	4	1	-	15
923121111111	15	8	9	3	35
923132000000	16	6	3	1	26
923315151515	13	13	3	1	30
923472020202	14	1	2	1	18

Table 2 depicts that “923020000003” & “923020000000” are the figures which show the people with frequent interactions and investigations simultaneously.

CDR. LEAs in Karachi map the figures of these suspects on its database [22]. After applying data storing technique, results obtained in the form of graphs to better portray the scenario as illustrated in

Table 3: Portion of CDR show IMEI numbers used by the suspect

B Party cell no	Date-time	IMEI
455454494E4753	12/02/2014 19:44	351933047609050
455454494E4753	13/02/2014 10:37	352035056080930
455454494E4753	13/02/2014 10:47	352464055524820
455454494E4753	14/02/2014 09:27	352464050524820
455454494E4753	14/02/2014 16:55	356236044849350
455454494E4753	01/03/2014 12:03	357582050164320
455454494E4753	13/03/2014 21:12	355932045208850

The detailed study of CDR leads to various SIMs or sets used by the suspects which can easily be done by the IMEI and IMSI numbers listed in CDR. For example, if CDR of a specific IMSI is found that identifies all the IMEI numbers utilized with IMSI. Further we can query for the IMEI numbers already acquired in prior step acquire IMSI numbers involved in the activity. This workout is helps investigation to discover large network of mobile phones used by the criminal/suspects in covert activities.

Software is used to uncover connections among numerous mobile / IMEI / IMSI numbers present in a

Fig. 2.

The following figure shows the analysis of geo-fencing data of several criminal incidents vs interconnections among culprits, involved in a variety of crimes. This highlights the cell numbers in-command for operation and reporting back to the main player(s) in the upper hierarchy of the criminal network.

4. Building Casual Bayesian Structure from Data Set

Various methods are used to carry out suspected calls. In the call tracking procedure, the interest variables have variation in dimension of position

are described, involving learning technique for partial information. We have used real-world case study in this research.

Table 4: States in the BN classifier

S. No	Node	States				
		Normal	Slightly Suspicious	Moderately Suspicious	Suspicious	Extremely Suspicious
1.	Call Duration	✓	✓	✓	✓	✓
2.	Call Timing	✓	✓	✓	✓	✓
3.	Call Temporal Density	✓	✓	✓	✓	✓
4.	Call frequency same number	✓	✓	✓	✓	✓
5.	Call Frequency Different number	✓	✓	✓	✓	✓
6.	Call Density	✓	✓	✓	✓	✓
7.	Call Source Location	✓	✓	✓	✓	✓
8.	Call Destination Location	✓	✓	✓	✓	✓
9.	Call Spatial Density	✓	✓	✓	✓	✓
10.	Suspicious call	X	✓	✓	✓	✓

(spatial) and time (temporal) [23] [24]. Caller ID, source and destination of call timing, call duration all are used to track suspected calls. The Bayesian models used encodes dependencies among all variables. Moreover, it promptly deals circumstances with missing data entries [25] [26] [27].

The Bayesian network is used to find out typical relations and achieve insights in suspicious call identification problem, as well as to foreshadow the results of intervening the calls.

In this paper, methods are presented for building Bayesian systems from previous information and summarized Bayesian statistical techniques for using databases for the improvement of such models. In this context, supervised and unsupervised learning methods for both i.e., parameter and structures of a Bayesian system

4.1 States

Bayesian structure using data set is used in learning. Different statistical tools are available to process such data. We have used SPSS package for statistical processing. In this study we have applied Bayesian Net classifier for leaning structure. The variables identified are shown in table 4.

4.2 Belief Updating

In probabilistic reasoning, Belief updating is vital component. Network probabilities (or, decisions) are altered; as a node is updated. An efficient algorithm is inevitable for belief update. In this paper Junction Tree algorithm (JTP) is utilized as basic approach. We have adopted junction tree algorithm in Bayesian networks for probabilistic inference. The Bayesian system for suspicious call figures out a technique for suspicious call; we utilize this system to view any probable interest. For example, from the Bayesian system in the figure, the probability of suspicious

call is shown. Findings of the other variations can be used as follows:

$$\rho(Cs) = \rho(Dt / Cd, Tc) * \rho(Dcs / Ls, Ld) * \rho(Dc / Fcsn, Fcon) \dots\dots\dots (1)$$

where,

Cs = Suspicious call
Dt = Temporal density

5000 in size record for training and 1000 record for testing. The results show accuracy around 89%.

4.3.1 Matrix of Confusion

Using Priori (data), we utilize the probability of a given Bayesian network. Hence, bypass to learn the probability of different parameter; we update its posterior distribution. Same approach is followed for

Table 5: Confusion Matrix for the BN classifier [1]

S. No.	Actual	Predicted	False Positive	False Negative	% Accuracy
Not Suspicious	320	330	10		96.5
Slightly Suspicious	72	66	—	6	91.5
Moderately Suspicious	62	64	—	2	97.8
Suspicious	28	30	2	—	92.5
Highly Suspicious	12	10	—	2	84.4
	500	500	15	9	?

Cd = Call duration
Tc = Call timing
Dcs = call spatial density
Ls = source location
Ld = destination location
Dc = Call density
Fcsn = call frequency same number
Fcon = call frequency other number

4.3 The Model

The model encompasses the structural and local probable distribution of a Bayesian system using the already known databases [28]. The results are set of techniques for analyzing data which uses previous information combining it with databases to produce improvised post information. Keeping in view the previous probability of suspicious call duration, we have made the model as shown in the figure. For the training of model we have made two data sets of

probability in a Bayesian system. Particularly the physical joint probability distribution is encoded from causal knowledge about the problem, in network.

Once a Bayesian network is constructed (from previous data/ information/ knowledge), we determine numerous interest probabilities from the network. In our problem which concerned to suspicious call detection; we are interesting about the probability of suspicious call, given, observations of the call duration.

5. Results

We have computed the probability estimation of different variables identified in the table and are given below

Table 6: Probability Estimation for “Call Density” [1]

Call Density	
Normal	0.20582
Slightly suspicious	0.19533
Moderately Suspicious	0.19016
Suspicious	0.19275
Extremely Suspicious	0.21594

The results suggest that the probability estimation for the call density of normal calls is 0.20582, while the slightly suspicious calls possess the call density of 0.19533, see the following table. The probability estimation for the call density of moderately suspicious calls is 0.19016 and that for the suspicious calls is 0.19275. Lastly, according to the results; the probability estimation for the call density of extremely suspicious calls is 0.21594.

Table 7: Probability Estimation for “Call Duration” [1]

Call Duration	
Normal	0.28751
Slightly suspicious	0.18362
Moderately Suspicious	0.084316
Suspicious	0.19221
Extremely Suspicious	0.25235

The results suggest that the probability estimation for the call duration of normal calls is 0.28751 while the slightly suspicious calls possess the call duration of 0.18362, see the following table. The probability estimation for the call duration of moderately suspicious calls is 0.084316 and that for the suspicious calls is 0.19221. Lastly, according to the results; the probability estimation for the call duration of extremely suspicious calls is 0.25235.

Table 8: Probability Estimation for “Call Frequency different Number” [1]

Call Frequency different Number	
Normal	0.17063
Slightly suspicious	0.16983
Moderately Suspicious	0.23516
Suspicious	0.22557
Extremely Suspicious	0.1988

The results suggest that the probability estimation for the call frequencies various numbers of typical calls is 0.17063 while the slightly suspicious calls possess the call frequency different number of slightly suspicious calls is 0.16983, see the following table. The probability estimation for the call frequency different number of the moderately suspicious calls is 0.23516 and that for the suspicious calls is 0.22557. Lastly, according to the results; the probability estimation for the call frequency different number of extremely suspicious calls is 0.1988.

Table 9: Probability Estimation for “Call Frequency Same Number” [1]

Call Frequency Same Number	
Normal	0.2042
Slightly suspicious	0.21119
Moderately Suspicious	0.17802
Suspicious	0.2
Extremely Suspicious	0.20659

The results suggest that the probable estimation for the call frequency same number of typical calls is 0.2042 while the slightly suspicious calls possess the call frequency same number of 0.21119, see the following table. The probability estimation for the call frequency same number of the moderately suspicious calls is 0.17802 and that for the suspicious calls is 0.2. Lastly, according to the results; the probability estimation for the call frequency same number of extremely suspicious calls is 0.20659.

Table 10: Probability Estimation for “Call Spatial Density” [1]

Call Spatial Density	
Normal	0.21461
Slightly suspicious	0.19962
Moderately Suspicious	0.18389
Suspicious	0.17951
Extremely Suspicious	0.22237

The results suggest that the probability estimation for the call spatial density of the normal calls is 0.21461 while the slightly suspicious calls possess the call spatial density of 0.19962, see the table. The probability estimation for the call spatial density of moderately suspicious calls is 0.18389 and that for the suspicious calls is 0.17951. Lastly, according to the results; the probability estimation for the call spatial density of extremely suspicious calls is 0.22237.

Table 11: Probability Estimation for “Call Temporal Density” [1]

Call Temporal Density	
Normal	0.35198
Slightly suspicious	0.17009
Moderately Suspicious	0.18389
Suspicious	0.15596
Extremely Suspicious	0.15536

The results suggest that the probability estimation for the call temporal density of normal calls is 0.35198 while the slightly suspicious calls possess the call temporal density of 0.17009 as shown in the table. The probability estimation for the call temporal density of moderately suspicious calls is 0.18389 and that for the suspicious calls is 0.15596. Lastly, according to the results; the probability estimation for the call temporal density of extremely suspicious calls is 0.15536.

Table 12: Probability Estimation for “Call Timing” [1]

Call Timing	
Normal	0.30789
Slightly suspicious	0.23576
Moderately Suspicious	0.21878
Suspicious	0.071528
Extremely Suspicious	0.16603

The results suggest that the probability estimation for the call timing of normal calls is 0.30789 while the slightly suspicious calls possess the call timing of 0.23576 as shown in the table. The probability estimation for the call timing of moderately suspicious calls is 0.21878 and that for the suspicious calls is 0.071528. Lastly, according to the results; the probability estimation for the call timing of extremely suspicious calls is 0.16603.

Table 13: Probability Estimation for “Destination Location” [1]

Destination Location	
Normal	0.21259
Slightly_suspicious	0.21978
Moderately_Suspicious	0.17902
Suspicious	0.19141
Extremely_Suspicious	0.1972

The results suggest that the probability estimation for the destination location of normal calls is 0.21259 while the slightly suspicious calls possess the destination location of 0.21978, see the table. The probability estimation for the destination location of moderately suspicious calls is 0.17902 and that for the suspicious calls is 0.19141. Lastly, according to the results; the probability estimation for the destination location of extremely suspicious calls is 0.1972.

Table 14: Probability Estimation for “Source Location” [1]

Source Location	
Normal	0.2002
Slightly suspicious	0.19121
Moderately Suspicious	0.18322
Suspicious	0.18621
Extremely Suspicious	0.23916

The results suggest that the probability estimation for source location of normal calls is 0.2002 while the slightly suspicious calls possess the source location of 0.19121, see the table. The probability estimation for the source location of moderately suspicious calls is 0.18322 and that for the suspicious calls is 0.18621. Lastly, according to the results; the probability estimation for the source location of extremely suspicious calls is 0.23916.

Table 15: Probability Estimation for “Suspicious Call” [1]

Suspicious Call	
No	0.25145
Slightly	0.24686
Moderately	0.24539
Highly	0.2563

The results suggest that the probability estimation for the nil suspicious calls is 0.25145 while the slightly suspicious calls possess the probability estimation of 0.24686, see the table. The probability estimation for the moderately suspicious calls is 0.24539 and that for the highly suspicious calls is 0.2563.

6 Conclusion

In this research we focus on the method of mobile phones data used by the customers. We also study the call data records of different users in order to track the crimes carried out by certain criminal elements across the globe. This research also throws light on the various methods and techniques of using a Bayesian

network. All these methods have been shown in various tables and figures in detail.

References

- [1] Mohamed Ridza Wahiddin Abdul Waheed Mahesar, Ahmad Waqas, Nadeem Mehmood, Asadullah Shah, 2015, "Analyzing the Weighted Dark Networks using Scale-Free Network Approach", WSEAS transactions on computers, Vol. 14, pp: 748-759.
- [2] Qamar ul Arifeen, S M Aqil Burney, Nadeem Mahmood, Kashif Rizwan, Syed Abdul Khaliq Bari, “Economic Review and Street crime Analysis Using Cell Phone Snatching and Theft”, Current Economics and Management Research, ISSN 2356-8887, www.textroad.com. 5(5)266-275, 2015.
- [3] Burde,R. and Khan,T., (2012) Traceability in Digital forensic Investigation. International Journal Advanced Research in Computer Science & Software Engineering vol 2, issue 10, Oct 2012. ISSN: 2277128X
- [4] Qamar ul Arifeen, Soft computing approaches in forensics: an intelligent system approach, a Ph. D. thesis submitted to the University of Karachi, Pakistan, 2014.
- [5] Ashley Brinson, Abigail Robinson, and Marcus Rogers (2006), A Cyber forensics ontology: creating a new approach to studying cyber forensics, Digital Investigation, Elsevier pp 37-43.
- [6] Baggili, I. (2011). Digital Forensics and Cyber Crime: Second International ICST Conference, ICDF2C 2010, Abu Dhabi, United Arab Emirates, October 4-6, 2010, Revised Selected Papers. Springer Science & Business Media.
- [7] Goel, S. (2010). Digital Forensics and Cyber Crime: First International ICST Conference, ICDF2C 2009, Albany, Ny, USA, September 30 - October 2, 2009, Revised Selected Papers. Springer Science & Business Media.

- [8] Jones, A. (2008) Keynote speech. In: First International Conference on Forensic Applications and Techniques in Telecommunications, Information and Multimedia, Adelaide, Australia, January 21–23, 2008.
- [9] Khan, Z. (2014). A New War Zone for Pakistan's Islamists. Pakistan's Counterterrorism Challenge, 169.
- [10] Pakistan Telecommunication Authority, Online: <http://www.pta.gov.pk/>. Accessed September, 2014
- [11] Ahmed M. N, Muhammad Elmogy and A. M. Raid (2013) Fuzzy crimes investigation framework for tracking data. International Journal of Computer Applications, vol. 84-1, pp 34-43.
- [12] Alin C popesu and HanyFarid (2013), Statistical tools for digital forensic, Department of Computer Science at Dartmouth College, Hanvor, USA.
- [13] MacDonald, A. D., Sather, M. W. and Snapp, J. W. (2014), "Methods and apparatus for mobile station location estimation." U.S. Patent No. 8,676,231.
- [14] Wang, S., Min, J., and Yi, B. K. (2008). "Location based services for mobiles: Technologies and standards." IEEE international conference on communication (ICC), 2008.
- [15] Dmitry, N. and Sneps-Sneppe, M., Geofence and network proximity. *Internet of Things, Smart Spaces, and Next Generation Networking*. Springer Berlin Heidelberg, 2013. pp. 117-127.
- [16] Humphries, Laymon Scott, and Huey-Jiun Ngo (2007). "Method and system for tracked device location and route adherence via geofencing." U.S. Patent No. 7,164,986.
- [17] Baumgartner, K., S. Ferrari, and G. Palermo (2008). "Constructing Bayesian networks for criminal profiling from limited data." *Knowledge-Based Systems* 21.7: pp 563-572.
- [18] Oatley, Giles C., and Brian W. Ewart. (2003) "Crimes analysis software: 'pins in maps', clustering and Bayes net prediction." *Expert Systems with Applications* 25.4: 569-588.
- [19] Eagle, N., Quinn, J. A., and Clauset, A. (2010). Methodologies for continuous cellular Tower Data Analysis.
- [20] Martin, G. (2013). Terrorism and Transnational Organized Crime. Transnational Organized Crime: An Overview from Six Continents, 163.
- [21] Gelman, Andrew, et al. *Bayesian data analysis*. Vol. 2. Chapman & Hall/CRC, 2014.
- [22] Citizens-Police Liaison Committee (CPLC). Online: <http://www.cplc.org.pk/>. Accessed July, 2014.
- [23] Burney, A., Mahmood, N., and Abbas, Z. 2012. Advances in Fuzzy Rough Temporal Databases. In Proc 11th WSEAS International Conference on AIKED, University of Cambridge UK, pp 237-242.
- [24] Burney, A., Mahmood, N., and Ahsan, K. 2010. TempR-PDM: A Conceptual Temporal Relational Model for Managing, Patient Data. In Proc. Int. WSEAS conference AIKED, University of Cambridge UK, pp. 237-243.
- [25] Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. Machine learning, Vol. 29, No. 2-3, pp. 131-163.
- [26] Gonzalez, M.C. Hidalgo, C.A. Barabasi, A.L. (2008) understanding individual human mobility patterns. A Reality mining: sensing complex social system. Personal and Ubiquitous computing 10, 255-268.
- [27] Imam, K. (2011). Police and the Rule of Law in Pakistan: A Historical Analysis, Berkeley Journal of Social Sciences, Vol. 1, No. 8.
- [28] Burney, A., Mahmood, N., & Abbas, Z. (2012a). Advances in Fuzzy Rough Temporal Databases, In Proceeding: 11th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Base at University of Cambridge, UK, Feb. 22-24, 2012.