Investigation Study on Traffic Prediction in Cellular Networks

M. Govindarajan[1], Dr. V. Chandrasekaran[2], M. Balakrishnan[3]

[1]Associate Professor, Department of Computer Applications

[2]Professor, Department of Electronics and Communication Engineering,

[1, 2] Velalar College of Engineering and Technology, Thindal, Erode-12

[3]Assistant Professor, Department of MCA, Excel Business School, Komarapalayam. Emailid:balkrish2k@gmail.com

Abstract - Big data is the large volume of data stored on the daily basis. Big data in cellular network is a challenging one for performing traffic prediction. With increasing number of mobile devices accessing Internet, cellular traffic gets increased in the past few years. Traffic prediction is a demanding one because of the spatial and temporal dynamics depending on different user behavior. An accurate cellular network traffic prediction guarantees the good quality of service during the data access. Many works have been introduced for cellular network traffic prediction with different user Internet behaviors. However, the prediction time and accuracy performance was not improved. For addressing these problems, the survival study is carried out for different cellular network traffic prediction.

Keywords - Big data, traffic prediction, spatial, temporal dynamics, cellular network, user behavior.

I. Introduction

Cellular network is a communication network distributed over the land areas termed as cells. Every network server was served by one fixed-location transceiver. The base station presented cell with network coverage for voice transmission, data and additional content types. A cell uses different frequencies from neighbouring cells to avoid interference and to provide better service quality. Cellular traffic prediction is an essential one in network planning and management for real-time decision making and short-term prediction. The traffic prediction in cellular network comprises time series analysis, regression technique, gray systematic method, neural network etc. Traffic prediction is the process of forecasting the traffic volumes with historical speed and volume data. This paper is organized as follows: Section 2 portrays the review on different traffic prediction techniques in cellular networks, Section 3 describes the study and analysis of existing traffic prediction techniques, Section 4 explains the possible comparison between them. In section 5, the limitations and discussion of existing traffic prediction techniques are discussed. Section 6 concludes the paper.

II. Literature Review

The call detail records data were employed to identify the anomalies in network. K-means clustering was introduced in [1] for verification of anomalies. An anomaly free data was employed through removing the anomalous activities. But, the prediction accuracy was not improved using k-means clustering. A novel deep learning architecture termed Spatial–Temporal Cross-domain neural Network (STCNet) was introduced in [2] to collect the complex patterns hidden in cellular data. However, the prediction time consumption was not reduced using STCNet architecture.

A device-level lightweight malware identification and classification framework was introduced in [3] for Android malware identification. Network traffic was mirrored from wireless access point to server for data analysis. However, lightweight framework was not suitable for malware samples to enhance the detection performance. Traffic management system was introduced in [4] depending on prediction information to minimize the problems in metropolitan area. The emergent intelligence method employed the historical and spatio-temporal data for predicting the expected patterns of traffic density and travel time in every zone and region. But, the prediction accuracy was not improved using the traffic management system.

A new deep learning architecture was constructed in [5] to address the supervised regression problems. The architecture was based on additive network model with gradient boosting techniques and learning blocks stacked iteratively. However, gaNet was not employed to address the classification issues. The decomposition of in-cell and inter-cell data traffic was carried out with graph-based deep learning approach in [6] to perform the cellular traffic prediction to classify spatio-temporal dependency of urban cellular traffic. But, error rate was not minimized by graph-based deep learning approach

A new fuzzy deep-learning approach (FDCN) was introduced in [7] for forecasting the citywide traffic flow. A fuzzy deep convolutional network model was introduced to enhance the traffic flow prediction depending on spatial and temporal traffic flow correlation. However, the reinforcement learning was not considered with the external factors in traffic flow prediction. Extending Labeled Data (ELD) was employed in [8] to identify the label of unknown mobile traffic to enhance labelled mobile traffic data. ELD identified mobile traffic with encrypted payload. But, ELD failed to identify the streaming and web through limiting ServerTag and robust flow features.

An extensive analysis was performed in [9] on daily internet traffic data created from January to December 2017 in smart university at Nigeria. However, the prediction time consumption was not minimized through extensive analysis. A quadri-dimensional approach was introduced in [10] to construct the service quality management (SQM) tree in Big Data platform. But, the feature selection was not carried out by quadri-dimensional approach. A deep learning approach was introduced in [11] to model nonlinear dynamics of wireless traffic. The spatial and temporal dependence of cell traffic were collected through densely connected convolutional neural networks. But, prediction accuracy was not improved using deep learning approach

III. Traffic Prediction in Cellular Networks

Wireless cellular network comprised the group of cells covering large geographical area. Each cell has base station for bandwidth management in cell. A new user enters network in random cell when sufficient bandwidth resources were available in particular cell. The user moves from one cell to another cell while spending some time in every cell depending on their mobility model. Traffic development in cellular network is driven by enhanced smart phone subscription and increase in average data volume per subscription. The traffic prediction has attracted large interest for increasing the reliability and network efficiency.

A. Deep Transfer Learning for Intelligent Cellular Traffic Prediction Based on Cross-Domain Big Data

Deep learning based traffic prediction architecture combined cross-domain datasets into unified representation. The spatial, temporal and external factors influenced the traffic generation collected through ConvLSTM and CNN. The dense connectivity pattern improved the feature propagation for traffic prediction. A new deep learning architecture termed Spatial-Temporal Cross-domain neural Network (STCNet) was introduced to gather the complex patterns hidden in cellular data. Through adopting convolutional long short-term memory network as subcomponent, STCNet has strong ability in modeling the spatial-temporal dependencies. The cross-domain datasets were gathered and modeled through STCNet to capture external factors. A clustering algorithm partitioned the city areas into groups. A consecutive inter-cluster transfer learning plan was introduced to gather the regional differences and similarities from spatial and temporal domain. A successive inter-cluster transfer learning strategy was introduced to improve the knowledge reuse performance. The model-based deep transfer learning was carried out to use the spatiotemporal similarities of cellular traffic types for prediction performance enhancement.

B. A mobile malware detection method using behavior features in network traffic

A device-level lightweight malware identification and classification framework was introduced to identify the network traffic for finding the mobile malware as all malicious behaviors of malware were accomplished by network interface. The network traffic was promising one for disclosing the malicious traces of malware. The designed method used traffic mirroring technology to gather the network traffic generated through mobile apps. The generated network traffic was sent to the server for data analysis. The detection method was not based on user surfing habits, device resources or additional device specific factors. The designed method extracted the traffic features and employed detection model depending on machine learning to identify whether app is malicious or not. The meaningful functionality displayed final detection results and examined the reason behind the malicious observations. It allowed users to observe each feature contribution to attain the final result and informed users regarding the rationale of final decision.

C. Call Detail Records Driven Anomaly Detection and Traffic Prediction in Mobile Cellular Networks

Big data analytics examined the user and network information with machine learning tools. The call detail records data were employed to identify the anomalies in network. An unsupervised machine learning algorithm termed k-means clustering for authentication and verification of anomalies. A design was made for resource distribution, fault detection and avoidance through anomaly detection. An anomaly free data removed anomalous activities and instructed the neural network model. Through anomaly free data, anomalous activity effects were observed in the training model. An autoregressive integrated moving average model was employed to forecast the future traffic for user

IV. Performance Analysis of Traffic Prediction in Cellular Networks

The traffic prediction techniques in cellular networks are compared with epochs number of data points. The traffic prediction performance is enhanced by using various parameters. An experimental evaluation is implemented using java language with CDR Dataset. CDR Dataset named Community Resource for Archiving Wireless Data Dartmouth (CRAWDAD) is a wireless network data resource used for research community. CRAWDAD dataset comprised the mobile phone records of 142 days from September 2010 to February 2011. The dataset includes the five fields, namely date, user activity, time, type, direction and duration. The date of user activity is denoted as YYYYMMDD. User activity is in form of incoming call, outgoing call, incoming SMS and outgoing SMS. Time of user activity is denoted in hhmmss. Type represents the user activity type. Direction denotes user activity direction (incoming or outgoing). Duration denotes the voice call duration. The traffic prediction in cellular network is enhanced by utilizing the various parameters, namely,

- Prediction accuracy,
- Prediction time and
- Activity Level

A. Prediction Accuracy

Prediction accuracy (PA) is defined as the ratio of number of data points that correctly predicts the cellular network traffic to the total number of data points. It is measured in terms of percentage (%). It is given by

$$PA = \frac{Number\ of\ data\ points\ that\ correctly\ predicts\ traffic}{Total\ number\ of\ data\ points} * 100 \quad (1)$$

From (1), prediction accuracy is calculated. When the prediction accuracy is higher, the method is said to be more efficient

TABLE 1. **TABULATION FOR PREDICTION ACCURACY**

Number of data points (Number)	**Prediction Accuracy (%)**		
	STCNet Architecture	**Device-level lightweight malware identification and classification framework**	**K-means clustering**
10	85	10	85
20	88	20	88
30	91	30	91
40	89	40	89
50	86	50	86
60	84	75	69
70	87	77	73
80	90	80	76
90	92	83	78
100	95	86	81

Table 1 describes the prediction accuracy of three different methods, namely STCNet Architecture, Device-level lightweight malware identification and classification framework and K-means clustering for varying number of data points ranging from 10 to100. The graphical representation of prediction accuracy is given in figure 1.

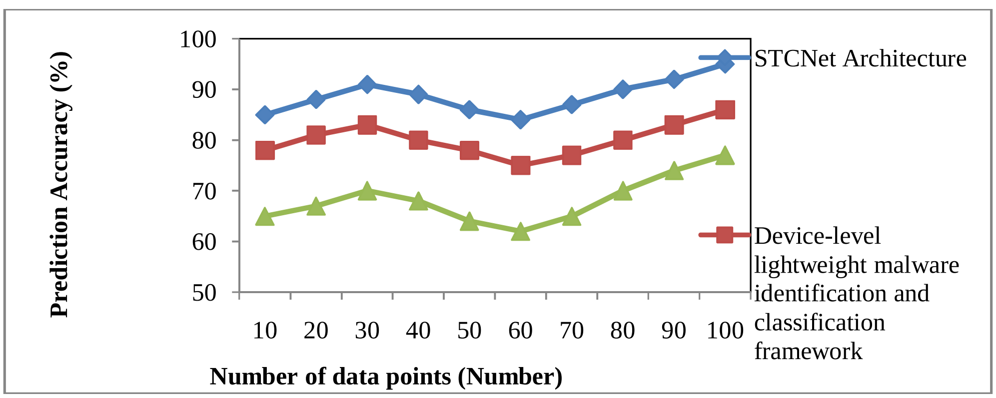

Fig. 1 Measurement of Prediction

B. Accuracy

Figure 1 explains the analysis of prediction accuracy versus number of data points taken from input dataset for three different methods. From the figure, it is clear that prediction accuracy using STCNet Architecture is higher than Device-level lightweight malware identification and classification framework and K-means clustering. This is because of introducing the consecutive inter-cluster transfer learning plan concepts to collect the regional differences and similarities from spatial and temporal domain. A successive inter-cluster transfer learning strategy enhanced the knowledge reuse performance. The model-based deep transfer learning employed the spatiotemporal similarities of cellular traffic types for prediction performance enhancement. Therefore, prediction accuracy of STCNet Architecture gets increased by 11% and 30% when compared to Device-level lightweight malware identification and classification framework andK-means clustering respectively.

C. Prediction time

Prediction time (PT) is defined as amount of time consumed for predicting the traffic in cellular networks. It is the difference of starting time and ending time for traffic prediction. It is measured in terms of milliseconds (ms). It is formulated as,

$$PT = Ending\ time - Starting\ time\ for\ traffic\ prediction \quad (2)$$

From (2), prediction time is calculated. When the prediction time is lesser, the method is said to be more efficient

TABLE 2 TABULATION FOR PREDICTION TIME

Number of data points (Number)	Prediction Time (ms)		
	STCNet Architecture	Device-level lightweight malware identification and classification framework	K-means clustering
10	35	25	30
20	37	27	32
30	40	29	34
40	42	32	37
50	45	35	39
60	48	38	41
70	51	41	44
80	54	45	48
90	57	49	52
100	60	52	55

Table 2 explains the prediction time of three different methods, namely STCNet Architecture,Device-level lightweight malware identification and classification framework and K-means clustering for varying number of data points ranging from 10 to100. Prediction time is calculated with respect to number of data points in input dataset. The diagrammatic representation of prediction time is shown in figure 2.

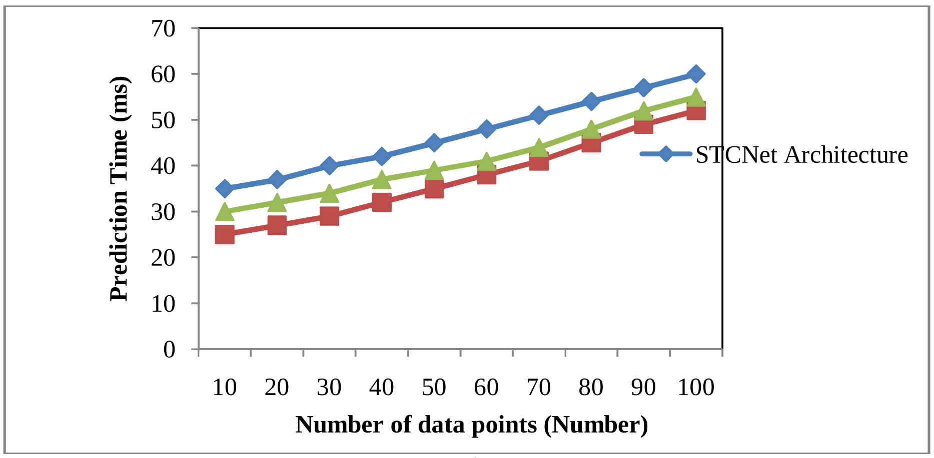

Fig. 2 Measurement of Prediction Time

Figure 2 illustrates the performance analysis of prediction time versus number of data points taken from input dataset. From the above figure, it is clear that prediction time using Device-level lightweight malware identification and classification framework is lesser than STCNet Architecture and K-means clustering. This is due to the application of traffic mirroring technology for identifying the network traffic generated through the mobile apps. After that, the generated network traffic was transmitted to the server for data analysis. It allowed the users to examine every feature contribution for attaining the final result and informed the users regarding rationale of final decision. As a result, prediction time of Device-level lightweight malware identification and classification framework gets reduced by 21% and 10% when compared to STCNet Architecture and K-means clustering.

D. Activity Level

Activity level is defined as the ratio of number of incoming call, outgoing call, incoming SMS and outgoing SMS to the time interval. It is measured in terms of activities per hour. It is formulated as,

$$AL = \frac{Number\ of\ incoming\ call, outing\ call, incoming\ SMS\ and\ outgoing\ SMS}{Time} \quad (3)$$

From (3), the activity level is determined. When the activity level is higher, the method is said to be more efficient.

Table 3 portrays the activity level of three different methods, namely STCNet Architecture, Device-level lightweight malware identification and classification framework and K-means clustering for varying time period from 3.00 hours to 24.00 hours. The diagrammatic representation of activity level is given in figure 3.

TABLE 3 TABULATION FOR ACTIVITY LEVEL

Number of data points (Number)	Activity Level (activities per hour)		
	STCNet Architecture	Device-level lightweight malware identification and classification framework	K-means clustering
3.00	0.2	0.2	0.3
6.00	0.3	0.3	0.5
9.00	0.8	1	1.5
12.00	1.5	1.9	2.2
15.00	0.8	1	1.5
18.00	2.5	3	4
21.00	0.3	0.3	0.5
24.00	0.3	0.3	0.4

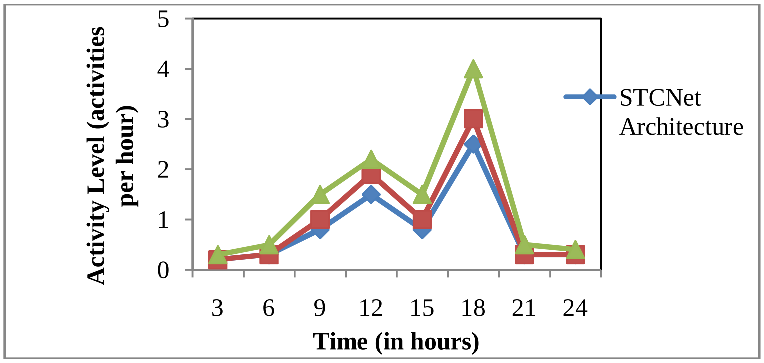

Fig. 3 Measurement of Activity Level

Figure 3 illustrates the analysis of activity level versus time interval for three different methods. From above figure, it is observed that activity level using K-means clustering is higher than STCNet Architecture and Device-level lightweight malware identification and classification framework. This is because, anomaly free data removed the anomalous activities and used in neural network model. An autoregressive integrated moving average model forecasted the future traffic for the user. Consequently, activity level of K-means clustering gets increased by 62% and 46% when compared to STCNet Architecture and Device-level lightweight malware identification and classification framework respectively

V. DISCUSSION AND LIMITATION OF TRAFFIC PREDICTION IN CELLULAR NETWORKS

Spatial–Temporal Cross-domain neural Network (STCNet) grouped the complex patterns hidden in cellular data. STCNet included strong ability in modeling the spatial–temporal dependency through convolutional short-term memory network. STCNet collected the external factors that affect the traffic generation. But, the prediction time consumption was not minimized using STCNet architecture. Device-level lightweight malware identification and classification framework was introduced for Android malware identification process. Network traffic created through mobile app was mirrored from the wireless access point to server for data analysis. The data analysis and malware detection were carried out on server side with lesser resource consumption on mobile devices. However, lightweight framework was not appropriate for many malware samples to enhance the detection performance. K-means clustering was introduced for authentication and verification of anomalies. The call detail records data identified the anomalies in networks. An anomaly free data was prepared through removing the anomalous activities with neural network model. The hidden information of user was employed for resource allocation and distribution. But, the prediction accuracy was not improved using k-means clustering.

A. Future Direction

The future direction of the work is to perform efficient traffic prediction in cellular networks by using machine learning and deep learning techniques with higher accuracy and lesser time consumption.

VI. CONCLUSION

A survival study of different traffic prediction techniques in cellular networks is carried out. From the study, lightweight framework was not suitable for various malware samples to increase the detection performance. In addition, the prediction accuracy was not enhanced by k-means clustering and prediction time consumption was not minimized using STCNet architecture. The wide range of experiments on existing traffic prediction techniques compares the results and discusses their limitations. Finally from result, the research work can be carried out by machine learning and deep learning techniques for traffic prediction techniques in cellular network with higher accuracy and lesser time consumption.

REFERENCES

[1] KashifSulta, Hazrat Ali and Zhongshan Zhang, “Call Detail Records Driven Anomaly Detection and Traffic Prediction in Mobile Cellular Networks”, IEEE Access, Volume 6,July 2018, Pages 41728 – 41737
[2] Chuanting Zhang, Haixia Zhang, JingpingQiao, Dongfeng Yuan and Minggao Zhang, “Deep Transfer Learning for Intelligent Cellular Traffic Prediction Based on Cross-Domain Big Data”, IEEE Journal on Selected Areas in Communications, Volume 37, Issue 6, June 2019, Pages 1389-1401
[3] Shanshan Wang, Zhenxiang Chen, Qiben Yan, Bo Yang, Lizhi Peng and ZhongtianJia, “A mobile malware detection method using behavior features in network traffic”, Journal of Network and Computer Applications, Elsevier, Volume 133, 1 May 2019, Pages 15-25
[4] Suresh Chavhan and PallapaVenkataram, “Prediction based traffic management in a metropolitan area”, Journal of Traffic and Transportation Engineering, Elsevier, July 2019, Pages 1-20
[5] Manuel Lopez-Martin, Belen Carro and Antonio Sanchez-Esguevillas, “Neural network architecture based on gradient boosting for IoT traffic Prediction”,Future Generation Computer Systems, Elsevier, Volume 100, 2019, Pages 656–673
[6] Xu Wang, Zimu Zhou, Fu Xiao, Kai Xing, Zheng Yang, Yunhao Liu and Chunyi Peng, “Spatio-Temporal Analysis and Prediction of Cellular Traffic in Metropolis”, IEEE Transactions on Mobile Computing, Volume 18, Issue 9, 2019, Pages 1-14
[7] Weihong Chen, Jiyao An, Renfa Li, Li Fu, GuoqiXie, ZakirulAlamBhuiyan and Keqin Li, “A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial–temporal data features”, Future Generation Computer Systems, Elsevier, Volume 89, 2018, Pages 78–88
[8] Zhen Liu, Ruoyu Wang and Deyu Tang, “Extending labeled mobile network traffic data by three levels traffic identification fusion”, Future Generation Computer Systems, Elsevier, Volume 88, 2018, Pages 1–13
[9] Aderibigbe Israel Adekitan, Jeremiah Abolade and OlamilekanShobayo, “Data mining approach for predicting the daily Internet data traffic of a smart university”, Journal of Big Data, Springer, Volume 6, Issue 11, 2019, Pages 1-23
[10] MaluambanzilaMinerveMampaka and MbuyuSumbwanyambe, “A quadri-dimensional approach for poor performance prioritization in mobile networks using Big Data”, Journal of Big Data, Springer, Volume 6, Issue 10, 2019, Pages 1-15
[11] Chuanting Zhang, Haixia Zhang, Dongfeng Yuan and Minggao Zhang, “Citywide Cellular Traffic Prediction Based on Densely Connected Convolutional Neural Networks”, IEEE Communications Letters, Volume 22, Issue 8, 2018, Pages 1-7