IJRISE International Journal of Research In Science & Engineering
Volume: 3 Issue: 4 July-August 2017
e-ISSN: 2394-8299
p-ISSN: 2394-8280

MOBILE MESSAGING APPLICATION FRAMEWORK FOR SERVICE USAGE CATEGORIZATION

Sumanth Sriramoju[1],Prof. Mangesh Ingle[2],Prof. Ashish Mahalle[3]
[1]Senior Pennsylvania State University Student -Information Sciences and Technology (BS)
[2,3]JCOET, Yavatmal

sumanthramoju@gmail.com

Abstract
Due to rapid growth of mobile technologies and the innovations in the field of mobile computing, it is possible to have mobile messaging applications like WhatsApp. These applications are capable of supporting messaging with different features. For instance WhatsApp can support different media for messaging including text, image, audio, video, video streaming, news feed and location sharing. Classification of such data into different content types can have very important utility in the real world. The rationale behind this is that by categorizing service usage in mobile applications can monitor traffic and know the bandwidth details. The network traiffic may be of many kinds like elastic and inelastic flows. The categorization can optimize network traffic and it can provide insights to allocate bandwidth and other resources to network flows. Some flows need continuous allocation of resources while others do not need. In this paper we considered mobile messaging application like WhatsApp. The traffic is captured and classifier is built in order to have better categorization of service usage. This can help in allocating resources like bandwidth with well informed decision making. A framework is proposed to achieve this. We built a prototype application to show the effectiveness of the proposed framework.

Keywords: Mobile computing, service usage classification, mobile messaging applications

INTRODUCTION

Usage of mobile applications became a common phenomenon due to the usage of mobile devices. There is rapid growth in mobile technologies and mobile computing. People of all walks of life are using mobile devices for various purposes. They perform operations like chatting, messaging, exchange of views, video, audio and commercial activities like bookings, shopping etc. The usage of different services in mobile applications are witnessed. The general usage types include news fed, short video, location sharing, stream video call, audio note, picture, text and other usages. The protocols used for communication are User Datagram Protocol (UDP) and Transmission Control Protocol (TCP). However, these are traditional protocols while new ones are used in various applications.

Figure 1: Mobile Apps for different services

As shown in Figure 1, it is evident that there are many applications running in mobile devices that can provide plethora of services. These services are captured and classified in this paper for effective decision making on resource allocations. In this paper, the focus is on the mobile applications that are used for messaging. These applications support different usage characteristics as mentioned above. The usage types such as text, video and audio have different requirements in mobile computing environment. For instance more bandwidth is consumed by the applications when video is transmitted. This is the important observation that led to the research on the estimation of resources needed based on the usage types. Usage of different services can lead to diverse needs of resource allocation. This is especially important in resource intensive applications running in resource constrained networks like wireless networks. As the usage types are more and there is bulk of information to be processed in order to estimate the resources needed for each type of service. In this paper, we use data mining technique for classifying different usage types in order to have better provision of resources. Different service types pertaining to WhatsApp are used to have this investigation. The rationale behind this is that WhatsApp is well known and its dataset provides ample scope for classification purposes. We proposed an algorithm that is used to have classification of usage types of services rendered by messaging applications. The remainder of the paper is structured as follows. Section II provides review of literature. Section III presents the proposed system in detail. Section IV presents implementation detils. Section V shows experimental results while section VI concludes the paper.

RELATED WORKS

Traffic classification is in the literature is reviewed and presented in this section. As explored in [1] traffic classification results in understanding different types of traffic being flown over Internet. Most of the Internet applications make use of well known ports pertaining to UDP and TCP. This is an important fact that simplifies the approach of traffic classification. There are some payload based methods as presented in [2]. The payload based methods are complex and they need to analyze based on the payload [3]. Encrypted traffic classification over Internet has two categories in classification. They include approaches based on host behaviour and approhes based on flow features [4]. Many data mining approaches are used in the literature for traffic classification. They include Support Vector Machine (SVM) [5], Naive Bayes [6], Bayesian Neural Neteorks [7] and decision trees [8]. There are some clustering algorithms also used in the literature. They include DBSCAN [9], K-Means [10], and AutoClass [11]. With respect to usage analsis of moble applications, digital footprints of mobile users is considered [12]. In-App usage analysis is given importance in this paper. Many researchers contributed towards this end. Custom logging tool [13], trace based analysis [14], cross layer interaction [15], and usage pattern analysis [16]. The concept of Hashtable is used in [17] for storing traffic and analyzing it. piece wise representation of time series data is explored in [18], [19], [20], [21] and [22]. In this paper we proposed a framework and an algorithm for classification of service usage in mobile messaging applications like WhatsApp. In-app usage of traffic flows are considered for investigating the dynamics of usage of services by mobile users.

Dataset Collection

Dataset collection is done by 20 users specifically trained to capture different services realted data. WeChat and WhatsApp data is collected for around 3 months. For both applications Internet traffic is captured for experimental study.

#	Usage Type	Packets	Bytes	Avg. Traffic /min
1	text	105K	20M	24K
2	picture	294K	215K	1,175K
3	audio note	82K	37M	29K
4	stream video call	2,222K	734M	2,850K
5	location sharing	15K	5,944K	149K
6	short video	70K	50M	823K
7	news feed	595K	381M	755K
8	outlier	1,517K	59.7M	194K

Table 1: Details of Internet traffic of WeChat application

#	Usage Type	Packets	Bytes	Avg. Traffic /min
1	text	33.724K	387.373K	1.222K
2	picture	79.626K	3.967M	22.709K
3	audio note	85.761K	6.432M	52.120K
4	stream video call	209.229K	4.127M	33.308K
5	location sharing	11.313K	289.964K	3.274K
6	short video	385.122K	43.295M	328.179K

Table 2: Details of Internet traffic of WhatsApp application

Table 1 and Table 2 show the details of the data collected. The usage type, packets, byte sof data and average traffic per minute are the details catpuredd. These datasets are used for analyzing and classification of service usage.

Proposed Traffic Classification Approach

We proposed a framework and an algorithm for service usage classification in mobile applications like WeChat and WhatsApp. The framework is implemented using Java programming language. The traffic data collected from the two applications are used for analysis and making different usage patterns.

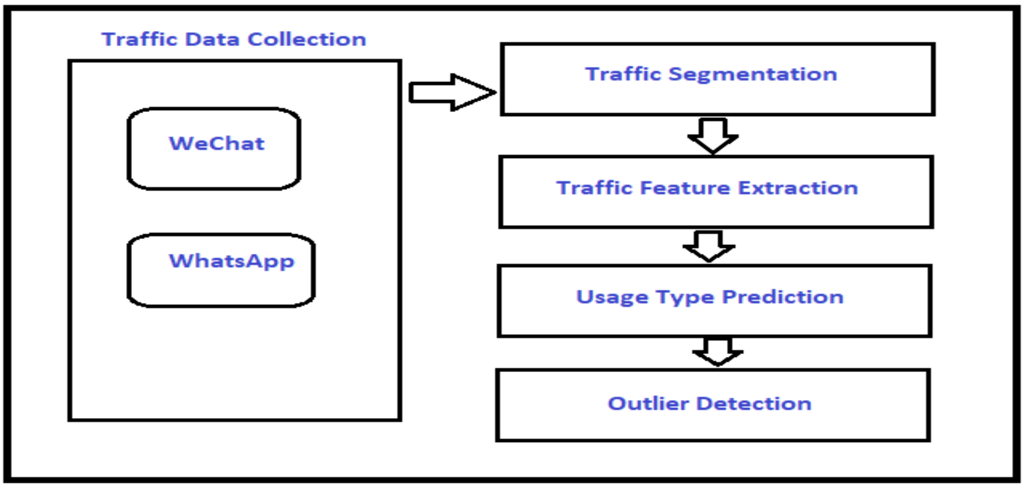

Figure 2: Proposed framework for service usage classification

As shown in Figure 2, it is evident that traffic data is collected from two applications. Then the traffic is subjected to segmentation. In this phase traffic is segmented based on sessions and then sessions to dialogs. After the segmentation, the traffic features are extracted. The traffic features considered include packet length and time delay. Based on the features of traffic usage types are predicted. Then outliers are detected. Outliers are the traffic patterns that are strange and do not appear like regular service patterns. This also can help in providing security to applications. The service usage classification is done using the following algorithm.

Algorithm: Mobile Traffic Based Service Usage Classification Inputs: WeChat and WhatsApp datasets 01 segment traffic flows into session and dialog based 02 in each segment find features 03 use features such as packet length and time delay 04 build a classifier 05 use classifier to classify service usage 06 update classifier for more accurate classification 07 continue step 1 and 6

Algorithm 1: Mobile Traffic Based Service Usage Classification

As shown in Algorithm 1, the traffic of two apps are analyzed and the usage types are predicted. The results revealed that the applications have diversified traffic flows and the usage analysis can help in making well informed decisions.

EXPERIMENTAL RESULTS

Experiments are made with different usage types considered. They are denoted as U1 to U7. They represent text, image, audio, video, location sharing, short video and news. Different evaluation metrics like overall accuracy, precision, recall and F-measure are used to evaluate the work carried out.

Length	Fluctuation	Descriptive Stat	Frequent Pattern	Time Delay	Combination
0.71	0.75	0.78	0.8	0.8	0.98

Table 3: Overall accuracy of WeChat data classification

International Journal of Research In Science & Engineering
Volume: 3 Issue: 4 July-August 2017
e-ISSN: 2394-8299
p-ISSN: 2394-8280

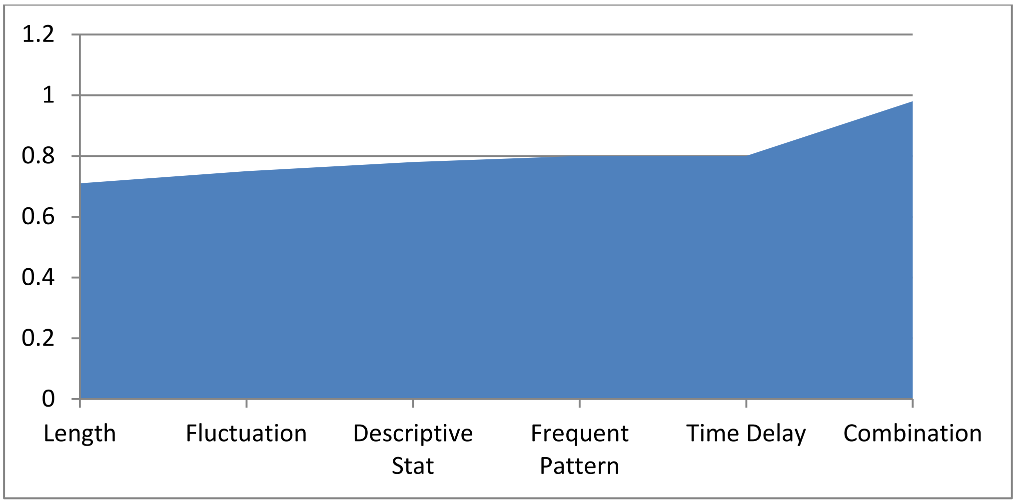

Figure 3: Overall accuracy of WeChat data classification

Length	0.85	0.68	0.67	0.71	0.7	0.85
Fluctuation	0.88	0.92	0.55	0.84	0.65	0.92
Descriptive Stat	1	0.58	0.9	0.86	0.8	0.62
Frequent Pattern	0.55	0.7	0.6	1	0.89	0.87
Time Delay	0.72	0.66	0.8	1	0.82	0.8
Combination	1	0.92	0.95	1	0.95	0.94

Table 4: Precision of different usage of WeChat

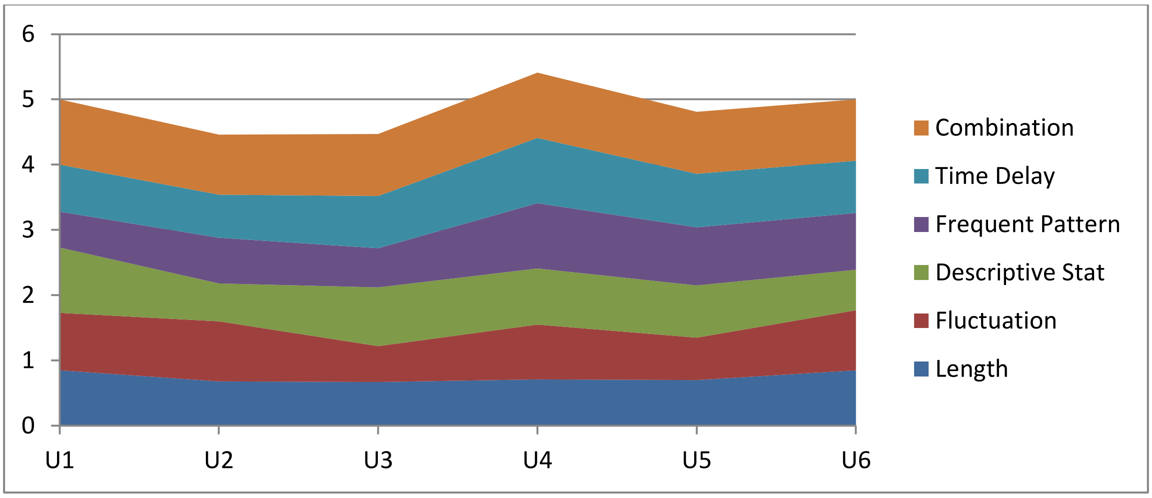

Figure 4: Precision of different usage of WeChat

Length	1	0.68	0.65	0.75	0.72	0.84
Fluctuation	0.58	0.88	0.84	0.7	0.68	0.86
Descriptive Stat	0.75	0.66	0.86	0.84	0.8	0.62
Frequent Pattern	0.6	0.82	0.75	1	0.82	0.78
Time Delay	0.88	0.75	0.8	1	0.85	0.55
Combination	1	0.95	1	1	0.91	0.94

Table 5: Recall of different usages of WeChat

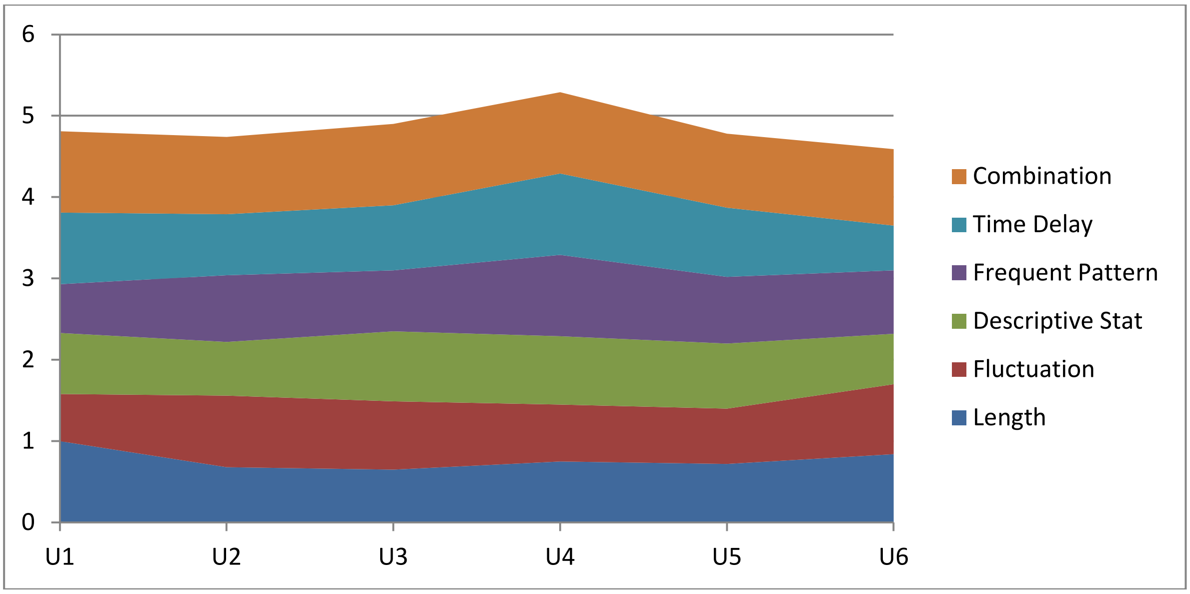

Figure 5: Recall of different usages of WeChat

Length	0.92	0.68	0.66	0.74	0.72	0.86
Fluctuation	0.7	0.9	0.65	0.75	0.65	0.9
Descriptive Stat	0.85	0.61	0.88	0.84	0.8	0.64
Frequent Pattern	0.58	0.75	0.68	1	0.85	0.83
Time Delay	0.8	0.7	0.8	1	0.82	0.68
Combination	1	0.92	0.98	1	0.92	0.94

Table 6: F-measure of different usages of WeChat

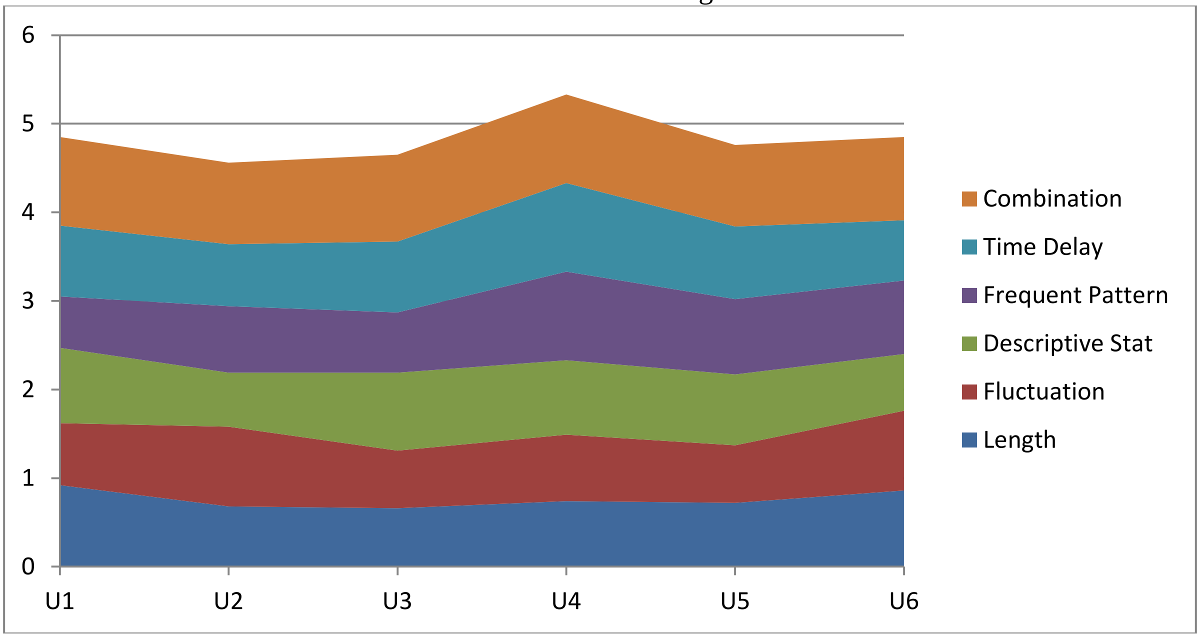

Figure 6: F-measure of different usages of WeChat

Length	Fluctuation	Descriptive Stat	Frequent Pattern	Time Delay	Combination
0.88	0.82	0.83	0.79	0.81	0.98

Table 7: Overall accuracy of classification of WhatsApp services

e-ISSN: 2394-8299
p-ISSN: 2394-8280

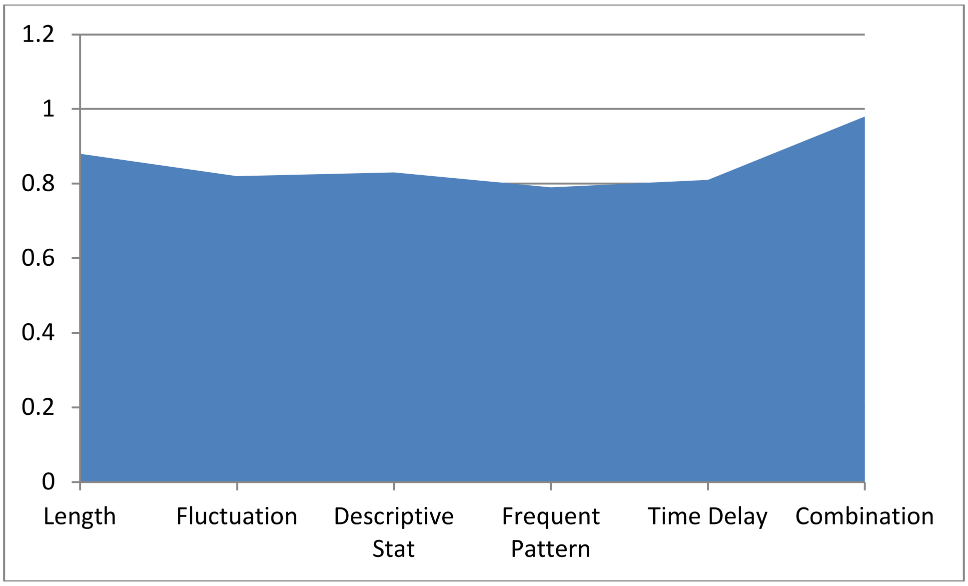

Figure 7: Overall accuracy of classification of WhatsApp services

Length	0.65	0.93	0.92	1	0.85	0.97
Fluctuation	0.84	0.78	0.8	0.97	0.75	0.82
Descriptive Stat	0.98	0.67	0.74	0.98	0.74	0.92
Frequent Pattern	0.76	0.74	0.56	0.94	0.92	0.9
Time Delay	0.84	0.68	0.79	0.98	0.85	0.74
Combination	1	1	0.92	1	0.97	0.97

Table 8: Precision of different usages of WhatsApp services

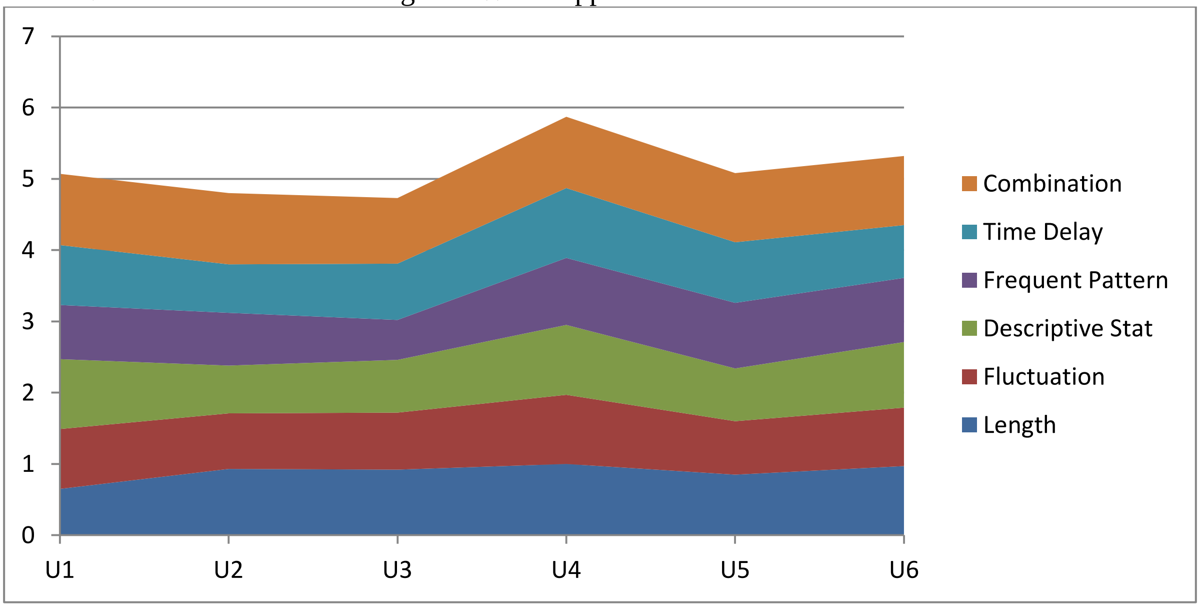

Figure 8: Precision of different usages of WhatsApp services

Length	0.86	1	0.96	0.94	0.74	0.68
Fluctuation	0.87	0.8	0.91	0.95	0.68	0.75
Descriptive Stat	0.88	0.82	0.71	0.96	0.78	0.85
Frequent Pattern	0.86	0.62	0.73	1	0.95	0.62
Time Delay	0.98	0.8	0.74	0.95	0.58	0.75
Combination	1	1	0.98	1	0.9	0.97

Table 9: Recall of different usages of WhatsApp

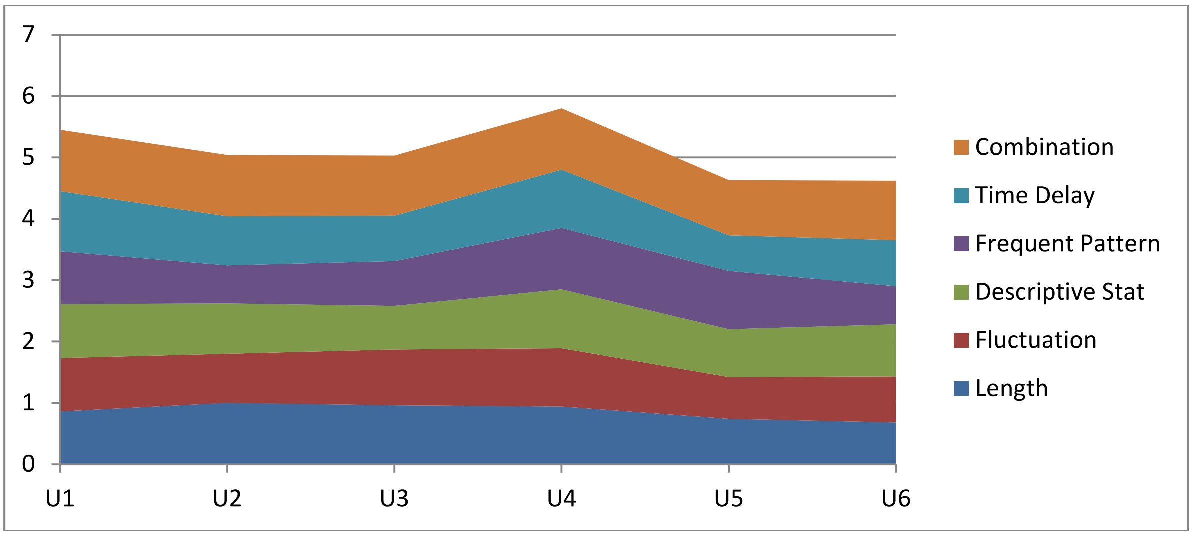

Figure 9: Recall of different usages of WhatsApp

Length	0.74	0.95	0.94	0.96	0.79	0.81
Fluctuation	0.85	0.79	0.84	0.96	0.72	0.79
Descriptive Stat	0.92	0.74	0.72	0.97	0.75	0.89
Frequent Pattern	0.8	0.68	0.64	0.97	0.94	0.73
Time Delay	0.9	0.74	0.75	0.97	0.68	0.74
Combination	1	1	0.95	1	0.93	0.97

Table 10: F-measure of different usages of WhatsApp

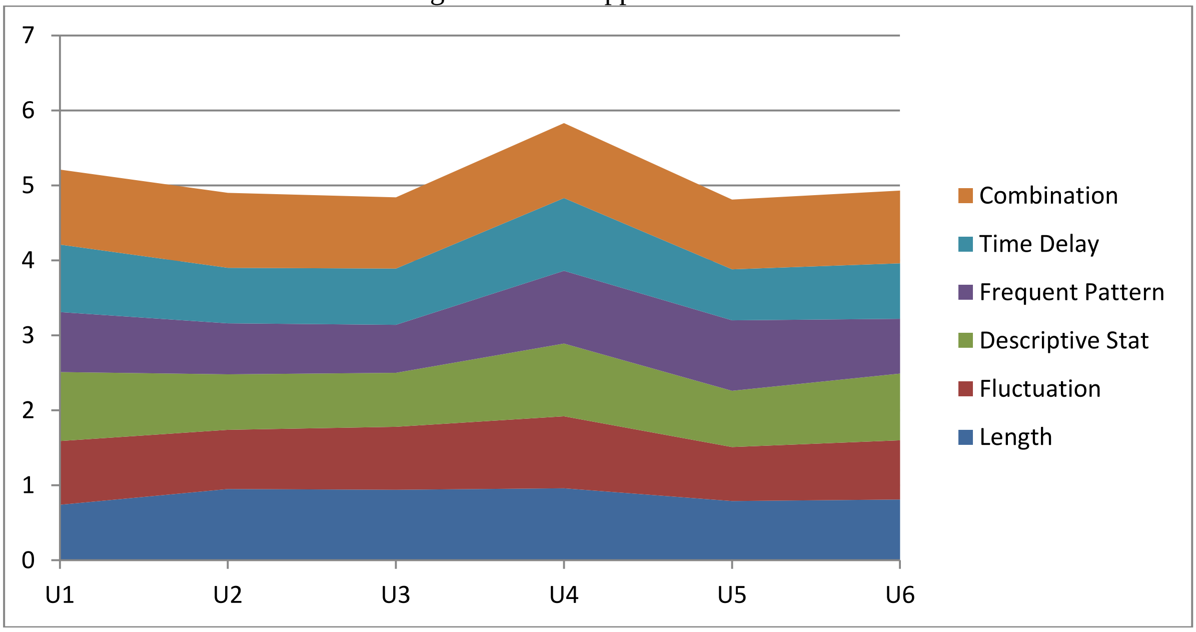

Figure 10: F-measure of different usages of WhatsApp

Conclusion and Future Work

In this paper, we studied the usage patterns of various services rendered by mobile applications with respect to messaging. Two applications such as WeChat and WhatsApp are used to investigate the service usage patterns. These two mobile applications are used to capture service usages. The datasets used for the service classification are considered due to the diversified messaging services. In this paper, we built a framework that is used to have classification mechanism that can predict different services based on the features extracted. First of all, the datasets are subjected to segmentation and then feature extraction is made. Then based on the packet length and delay time, the service usages are classified. We built a prototype application that is used to demostrte the proof of the concept. The work is evaluatedby using metrics like

e-ISSN: 2394-8299
p-ISSN: 2394-8280

precision, recall, F-measure and overall accuracy. The results revealed the usefulness of the proposed framework. This research is extened further to have more case studies and generalie the findings.

REFERENCES

[1] Thomas Karagiannis, Andre Broido, Michalis Faloutsos, et al. Transport layer identification of p2p traffic. In Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, 2004.

[2] Subhabrata Sen, Oliver Spatscheck, and Dongmei Wang. Accurate, scalable in-network identification of p2p traffic using application signatures. In Proceedings of the 13th international conference on World Wide Web, 2004.

[3] Riyad Alshammari and A Nur Zincir-Heywood. Machine learning based encrypted traffic classification: identifying ssh and skype. In Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on, 2009.

[4] Hyunchul Kim, Kimberly C Claffy, Marina Fomenkov, Dhiman Barman, Michalis Faloutsos, and KiYoung Lee. Internet traffic classification demystified: myths, caveats, and the best practices. In Proceedings of the 2008 ACM CoNEXT conference, 2008.

[5] Alice Este, Francesco Gringoli, and Luca Salgarelli. Support vector machines for tcp traffic classification. Computer Networks, 2009.

[6] Andrew W Moore and Denis Zuev. Internet traffic classification using bayesian analysis techniques. In ACM SIGMETRICS Performance Evaluation Review, 2005.

[7] Tom Auld, Andrew W Moore, and Stephen F Gull. Bayesian neural networks for internet traffic classification. Neural Networks, IEEE Transactions on, 2007.

[8] Nigel Williams, Sebastian Zander, and Grenville Armitage. A preliminary performance comparison of five machine learning algorithms for practical ip traffic flow classification. ACM SIG

[9] Jeffrey Erman, Martin Arlitt, and Anirban Mahanti. Traffic classification using clustering algorithms. In Proceedings of the 2006 SIGCOMM workshop on Mining network data, 2006.

[10] Laurent Bernaille, Renata Teixeira, Ismael Akodkenou, Augustin Soule, and Kave Salamatian. Traffic classification on the fly. ACM SIGCOMM Computer Communication Review, 2006.

[11] Sebastian Zander, Thuy Nguyen, and Grenville Armitage. Automated traffic classification and application identification using machine learning. In The IEEE Conference on Local Computer Networks, 2005.

[12] Anindya Ghose and Sang Pil Han. An empirical analysis of user content generation and usage behavior on the mobile internet. Management Science, 2011.

[13] Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, and Deborah Estrin. Diversity in smartphone usage. In Proceedings of the 8th international conference on Mobile systems, applications, and services, 2010.

[14] Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, and Shobha Venkataraman. Identifying diverse usage behaviors of smartphone apps. In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, 2011.

[15] Sriramoju Ajay, B. (2017). INTELLIGENT MOBILE APP FOR FINDING PATH AND TRACKING POST PACKETS USING ANDROID PLATFORM. International Journal Of Research In Science & Engineering, 3(2), 9.

[16] Sriramoju Ajay, B. (2017). Investigation of Feasible Tourist Destinations using Android Mobile App. International Journal Of Research In Science & Engineering, 3(2), 9.

[17] Babu, Sriramoju Ajay, and Namavaram Vijay. "Image Tag Ranking for Efficient Matching and Retrieval." (2016).

[18] Babu, Sriramoju Ajay, and Namavaram Vijay. "Design and Implementation of a Framework for Image Search Reranking." (2016).

[19] Babu, Sriramoju Ajay and Babu, S Shoban. "International Journal of Research and Applications Jan-Mar© 2016 Transactions 3 (9): 422-426 eISSN: 2349–0020."

[20] Babu, Sriramoju Ajay. "PARTICLE SWARM OPTIMIZATION ALGORITHM FOR ROUTING NETWORK" (2017).
[21] Babu, Sriramoju Ajay. "MODIFICATION AFFINE CIPHERS ALGORITHM FOR CRYPTOGRAPHY PASSWORD" (2017).
[22] Babu, Sriramoju Ajay. "Perceptual-Based Quality Metrics For Image and Video Services" (2015).
[23] Sriramoju, Ajay Babu. "Analysis on Image Compression Using Bit-Plane Separation Method" (2014).
[24] Babu, Sriramoju Ajay. "Objective Quality Metric Design For Wireless Image and Video Communication" (2014).
[25] Ajay Babu Sriramoju, Dr. S. Shoban Babu. "Study of Multiplexing Space and Focal Surfaces and Automultiscopic Displays For Image Processing" (2013).
[26] Sriramoju, Ajay Babu. "Image Processing: Lossy Compression by Color Quantization and Gct Modeling" (2012).
[27] Sriramoju, Ajay Babu. "Analysis on Lossless Image Compression in Image Processing" (2011).
[28] Feng Qian, Zhaoguang Wang, Alexandre Gerber, Zhuoqing Mao, Subhabrata Sen, and Oliver Spatscheck. Profiling resource usage for mobile applications: a cross-layer approach. In Proceedings of the 9th international conference on Mobile systems, applications, and services, 2011.
[29] Alok Tongaonkar, Shuaifu Dai, Antonio Nucci, and Dawn Song. Understanding mobile app usage patterns using in-app advertisements. In Passive and Active Measurement, 2013.
[30] Scott E Coull and Kevin P Dyer. Traffic analysis of encrypted messaging services: Apple imessage and beyond. ACM SIGCOMM Computer Communication Review, 2014.
[31] Janos Abonyi, Balazs Feil, Sandor Nemeth, and Peter Arva. Fuzzy clustering based segmentation of time-series. In Advances in Intelligent Data Analysis V. Springer, 2003.
[32] BR Bakshi and G Stephanopoulos. Representation of process trends-iv. induction of real-time patterns from operating data for diagnosis and supervisory control. Computers & Chemical Engineering, 1994.
[33] Eamonn J Keogh and Michael J Pazzani. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In KDD, 1998.
[34] Themis Palpanas, Michail Vlachos, Eamonn Keogh, and Dimitrios Gunopulos. Streaming time series summarization using userdefined amnesic functions. Knowledge and Data Engineering, IEEE Transactions on, 2008.

[35] Hagit Shatkay and Stanley B Zdonik. Approximate queries and representations for large data sequences. In Proceedings of the Twelfth International Conference on Data Engineering, 1996.