BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 10, No. 3, July 2022

Proactive Metering of Mobile Internet User Experience

Mete Uzun and Osman Abul

Abstract—**Having 67% worldwide share, mobile internet is very important for Internet Service Providers (ISPs). Since mobile Internet access is a collective service, Key Performance Indicators (KPIs) measuring quality of data traffic on select network segments/servers may not correctly indicate true user experience. For this reason, mobile ISPs are investing in sophisticated high-end commercial speed analysis systems which typically collect and analyze network traffic data from key network segments/servers. Unfortunately, their utility is quite limited as long as the proactive network intervention is considered. In this work, we develop a MapReduce based network speed analysis system which measures end-to-end network speed to quantify true user experience across multiple geographic regions and service categories. Also functioning as an online decision support system, it enables network administrators with timely ISP network intervention right before potential arrival of mass number of user complaints. The system has been tested with a leading mobile ISP in Turkey. The results confirm its effectiveness.**

Index Terms— **Big data, Hadoop, Internet service providers, MapReduce, Mobile internet, Network speed analysis, User experience.**

I. Introduction

PACKET-SWITCHING mobile networks are increasingly carrying Internet traffic for multimedia, social media, web browsing, games, VoIP, IPTV and many other applications. As a result, the mobile Internet traffic is growing very fast. To this end, monthly global mobile data traffic is estimated to surpass 77 Exabyte with a share of 20% of total Internet traffic by the end of 2022, according to the Visual Networking Index of the Global Mobile Data Traffic Forecast by Cisco [1]. This is indeed why the versatile data traffic analysis and mobile network management with optimizations are crucial business issues for mobile Internet Service Providers (ISPs) [2].

Internet speed is very important for mobile users. ISPs employ several Key Performance Indicators (KPIs) measuring various aspects of data traffic flow [3]. KPIs are typically measured on specific core network nodes/segments by probing/tapping approach [4]. Unfortunately, due to multiple hops with distinct nodes/segment features, measuring network performance on a specific probe gives little information as far as the users' overall network speed experience is concerned. Clearly, KPIs depending on the tapping approach may fail in correctly characterizing the network quality and the true user experience. This suggests that for a proactive solution we need to measure the network quality and the true user experience with a holistic end-to-end analysis [5].

We live in the big data era in which many advanced technology systems produce high volume data. The big data harbors big opportunities to improve Internet services, to meet user needs and to create business value. Performance measurements on very large number of connection streams uncover unknown correlations of network problems, network capacity on busy hours, user experiences of QoS, hidden patterns of usage, customer preferences, market trends and other useful business and network information [14]. The analytical findings from big data can lead to more effective network operations, improved QoS efficiency, prescribed competitive intelligence, new revenue opportunities, better customer service, and other benefits.

Characterization of the user experience and detection of network problems before customer complaints are very important feedbacks for ISPs and the telecommunication sector. Network capacity on busy hours, real user experience, Quality of Service (QoS) and correlations of network problems can be gathered from analyzing connection streams and performance indicators of network centric measurement on the content servers. Another example is that, customer Internet usage trend is a very fundamental information for determining sales strategy. It is very important that the market share, the market value and the market intelligence can be easily estimated for businesses striving to increase operational efficiency and gain competitive advantage. Moreover, competitive intelligence and competitor analysis can be compiled by analyzing patterns of customer usage and mobile network data segmentation on the content servers. In this way, we can translate network capabilities to user experience, QoS and competitive intelligence using big data analysis on the Content Delivery Network [6].

METE UZUN, is with Department of Computer Engineering TOBB University of Economics and Technology, Ankara, Turkey, (e-mail: meteuzn@gmail.com).
https://orcid.org/0000-0002-6329-2835

OSMAN ABUL, is with Department of Computer Engineering TOBB University of Economics and Technology, Ankara, Turkey, (e-mail: osmanabul@etu.edu.tr).
https://orcid.org/0000-0002-9284-6112

Manuscript received March 17, 2022; accepted June 22, 2022.
DOI: 10.17694/bajece.1089321

BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 10, No. 3, July 2022

Our system, called Proactive ISP Speed Meter, is able to processes large speed test results in online parallel/distributed manner with MapReduce. Then using the big data analytics, the network speed maps and the service usage rates are used to characterize the true user experience: (i) for each geographic region served, and (ii) for each content delivery service category, separately. Using the relevant extracted information from end-to-end roundtrip speed test results, we can proactively translate them into Quality of Experience (QoE) per geographic region/service category. We show that our system can proactively detect potential massive user complaints before they emerge. Moreover, it provides mobile ISPs with competitive intelligence that may be exploited as a part of the decision support in future infrastructure investments.

The paper is organized as follows. We present the background and related work in the next section. Section III introduces our methodology and two main deployments (Geographic Region Based and Content Delivery Service Based) of Proactive ISP Speed Meter. An experimental evaluation of the system on live mobile network of a leading ISP provider of Turkey is provided in Section IV. Finally, Section V concludes. A preliminary version of our work, which covers only the Geographic Region Based Deployment, appears in [5].

II. Background and Related Work

Throughput, Round Trip Time (RTT) and Latency are the main metrics that define the performance of data transmission speed. Throughput is defined as the amount of successfully transferred data in a unit time. We consider downlink and uplink speed separately. RTT measures the elapsed time between a request and its response. The smaller RTT, the better broadband speed the clients will experience [8]. Latency is the time delay between the release and the arrival of data transmission. Latency especially degrades the user experience on cloud-based applications due to much longer response times [9]. Big latency causes big RTTs. RTT, throughput and latency are dependable KPIs for user experience metering of network speed [10]. Hence, our system mainly depends on those KPIs for user experience analysis.

We can roughly categorize the Internet experience KPIs into three groups: (i) Network centric, (ii) User centric and (iii) Competitive intelligence centric (Table 1) [26]. Network centric KPIs include throughput, RTT and latency and serve as the basic QoS metrics. User centric KPIs, including page load time and TCP connection time, are mainly functions of network centric KPIs and serve as QoE. Competitive intelligence centric KPIs like number of visits/downloads, on the other hand, are useful to identify highly demanded services. Hence, they provide mobile ISPs with strategic decision making in future infrastructure investments.

Some highly consulted KPIs, which are commonly employed by most network quality evaluation systems, may not correctly surface speed bottlenecks or true user experience. Being a reliable data delivery protocol, Transmission Control Protocol (TCP) can detect errors in packets and recover from damaged, lost, duplicated or out-of-order delivered packets [7]. Therefore, the errors will affect Data Traffic Availability and Bad Session KPIs, but may not affect customer perception of network speed. As a result, very high service up percentages reported by Data Traffic Availability and Bad Session KPIs, measured at the nodes of packet core network (Gateway GPRS Support Node-GGSN, Packet Data Network Gateway-PGW) or at data optimization servers, does not necessarily mean that high speed Internet service is provided to mobile users. For such reasons, our system avoids the use of such KPIs.

TABLE I
KEY PERFORMANCE INDICATORS

KPI	Measurement
Throughput	Network Centric
Latency	Network Centric
RTT	Network Centric
Dropped Frames	Network Centric
Page Load Time	User Centric
TCP Connection Time	User Centric
Visits, Plays, Viewers Rate	Competitive Intelligence Centric
Content Category Rate	Competitive Intelligence Centric

There are web services and standalone tools that provide analysis of Internet access/speed metrics for individual users [11]. For instance, Speedtest [12] is one of the most widely used bandwidth testing analysis tool. However, such tools are user-centric and cannot automatically characterize service quality of the ISP network. As a result, such tools are not appropriate for assisting in large-scale bottleneck detection and network management, as a part of decision support. Our big data analytics based distributed system, on the other hand, automatically measures high volume of throughput, RTT and latency KPIs periodically and proactively characterizes network speed quality of the whole network.

Because of the explosive growth of Internet connections, Content Delivery Network (CDN) produce gigantic number of server-side connection streams [13]. Internet content delivery servers are harnessing wealth and volume of information of the Internet activities, which may include web server logs, click stream data, social media, online shopping, video watching, network activities, mobile-phone call details and sensors connected to Internet of Things. After all of the Internet activities of users, highly accessed data are accumulated on the content servers for fast re-accesses. Our system measures competitive intelligence centric KPIs per service category to identify highly accessed content servers.

Utilizing a highly effective and efficient MapReduce programming model [15, 16], Hadoop [17] is typically used to process big data [14]. Hadoop provides opportunities for big data analytics [18] for mobile network service providers [19].

III. Methodology

We developed a network speed analysis system, called Proactive ISP Speed Meter, for ISP wide Internet speed performance analysis. Our focus is on mobile ISPs which typically have radio access and core networks for providing Internet access to mobile users (Fig. 1).

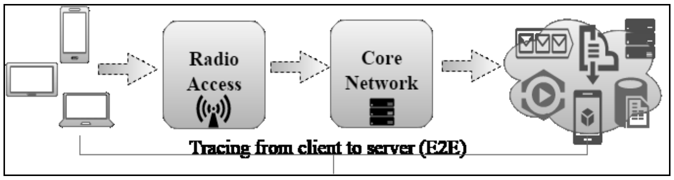

Fig. 1. Typical Mobile ISP infrastructure for Internet connection. Mobile devices access to ISP servers thorugh radio access and core network facilities.

Proactive ISP Speed Meter has two deployments: (i) geographic region based, and (ii) content delivery service based. The former deployment (Fig. 2) periodically samples the Internet access speed from each geographic region while the latter deployment (Fig. 4) analyzes the Internet access speed per content delivery service category. This way the ISPs are informed with how to gauge user experience per geographic region and per content delivery service category, and also to be able to react differently for various occasions of bottlenecks.

Proactive ISP Speed Meter's end-to-end speed analysis involves speed tests on the communication channel between mobile devices and servers on the network through the radio access and the core network interfaces [20]. Hence, the analysis results correctly indicate quality of network between clients and the server as the analysis does not depend on any specific segment of the radio access or the core network. Since the field tests can be done automatically without any technician interference, the network capacity planning can be efficiently conducted at any time.

A. Geographic Region Based Deployment

The system consists of three main components as shown in Fig. 2: (i) a client-side distributed speed test tool, called Speed Meter, running on smart devices as a daemon service using the ISP's address range, (ii) a speed test data analysis system, called "Speed Analyzer", running on Hadoop, and (iii) a dashboard application presenting analysis results for decision making.

1) Speed Meter with End-to-End Packet Analysis

The Speed Meter, implemented in Python and deployed as a daemon service on Linux, runs on geographically distributed mobile clients. It captures live data packets from the network and dumps them in pcap format using Dumpcap [21]. As a part of the test process it establishes a communication link with a server on the Internet at certain intervals for downloading and uploading various test files.

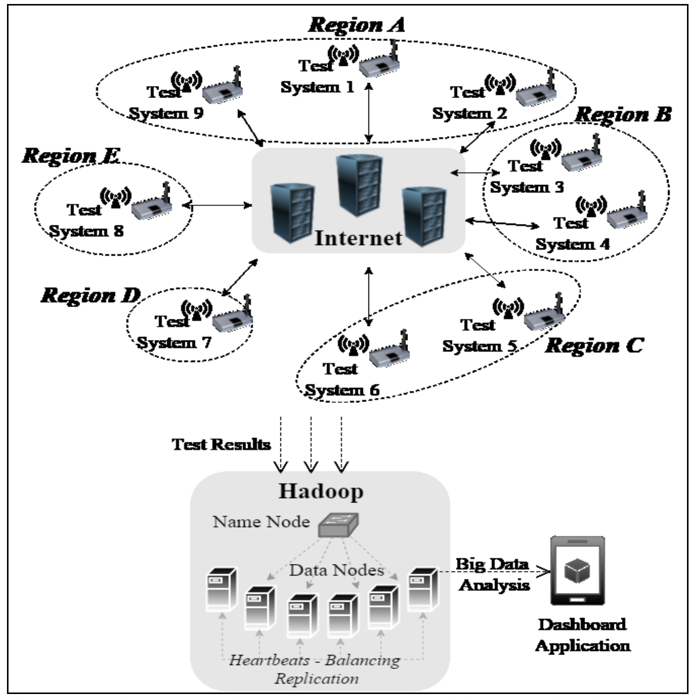

Fig. 2. Proactive ISP Speed Meter System (Geographic Region based deployment). It has three components of (i) the Speed Meter (running on clients), (ii) A Hadoop based speed analysis tool, and (iii) a dashboard application.

Tcptrace [22, 23] is used to analyze data captured by Dumpcap. The tool can produce several different types of output containing information on each connection seen, such as elapsed time, segments sent and received, retransmissions, round trip times, window advertisements, throughput, and more. The Speed Meter computes the download and upload speeds, RTT, average number of retransmissions grouped for each server name. Finally, the Speed Meter test results are pooled on a common area for further aggregate analysis. A sample output of which is shown in Table 2.

TABLE II
A SAMPLE SPEED METER RESULTS DATA

Sample Test Result					
System ID	Download (Mbit/s)	Upload (Mbit/s)	RTT (ms) (a2b ‖ b2a)	Retrans mit (a2b ‖ b2a)	Server
AnYen1	04.74	2.92	230.38 | 0.86	14 | 0	Vodafone
AnYen1	09.20	2.94	223.11 | 0.11	09 | 3	Avea
AnYen1	08.70	3.03	265.24 | 0.07	08 | 1	Doruknet
AnCan79	10.09	3.01	169.51 | 0.05	10 | 0	Vodafone
AnGol35	07.20	3.10	228.84 | 0.09	12 | 0	Vodafone
AnCan94	11.27	2.92	175.04 | 5.74	01 | 0	Avea

Not only the aggregate test results but individual test results collected by the Speed Meter is informative too. For instance, the hypothetical sample results given in Table 2 show horrendous user experience, e.g., file downloads taking far longer to complete. We outline the speed test results for each geographic region separately.

The Speed Meter algorithm is shown in Algorithm 1. Every 10 seconds (a configurable value) it uploads and downloads a packet and these communications are recorded by Dumpcap. The recorded communications are then analyzed by Tcptrace to give download/upload speeds and RTT for unlost communications. The speed test results are then sent to Hadoop for ISP wide analysis.

Algorithm 1 The Speed Meter Algorithm

```
timer = 10sn. start
while timer do
    while call("dumpcap") do
        runSpeedTest()
    end while
    Determine: results ⟸ call("tcptrace)
    if results =!NULL then
        download ⟸ calculateDownloadSpeed()
        upload ⟸ calculateUploadSpeed()
        rtt ⟸ calculateRtt()
    end if
    print download, upload, rtt
end while
```

Algorithm 1. The Speed Meter Algorithm. Each client runs this algorithm and emits the performance results to network speed logs, which is analyzed afterwards by speed data analysis component on Hadoop.

2) *Performance Analytics with MapReduce*

Coming from various regional Speed Meters, the speed test results as large data sets are processed and analyzed in parallel and distributed fashion. The Map and Reduce jobs are implemented in Java.

MapReduce works in the map and the reduce phases. Key-value pairs act as input and output at each phase. The input to the map function are raw test results as shown in Table 2. The map function (Algorithm 2) extracts system id with respective KPIs and emits them as its output. Then in the shuffle operation, the key-value pairs are grouped and sorted by key. After shuffling, the reduce function (Algorithm 3) averages the performance metrics to obtain the network speed. The speed labeling, indicated as "Good", "Average" and "Bad" tickets, qualifies the user experience for each test region. We use two thresholds, the value for which are experimentally selected, for ticket labeling. The data flow is illustrated in Fig. 3, with an example download speed obtained from Table 2.

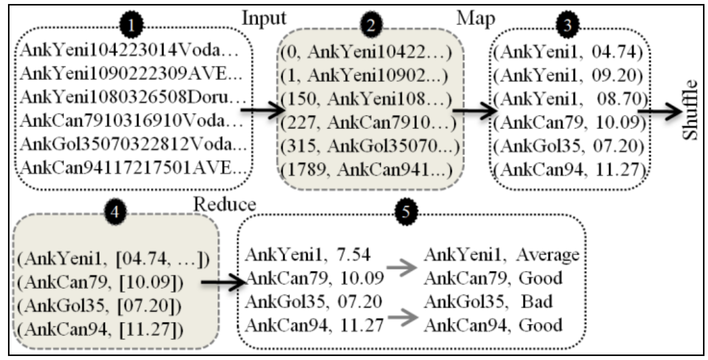

Fig. 3. Map and Reduce Data Flow. It shows an example run for download speed given in Table 2.

Algorithm 2 Map(docId a, doc d)

```
for term : docd do
    systemId ⟸ getSystemId(term)
    download ⟸ getDownload(term)
    upload ⟸ getUpload(term)
    rtt ⟸ getRtt(term)
    Emit(systemId, AssociativeArray[download, upload, rtt])
end for
```

Algorithm 2. The Mapper algorithm.

Algorithm 3 Reduce(systemId s, values[d1,d2...] d)

```
for value : d do
    downloadTotal ← downloadTotal + value[0]
    uploadTotal ← uploadTotal + value[1]
    rttTotal ← rttTotal + value[2]
    counter ++
end for
ticketDownload ⟸ label(totalDownload/counter)
ticketUpload ⟸ label(totalUpload/counter)
ticketRtt ⟸ label(totalRtt/counter)
Emit(s, AssociativeArray[ticketDownload, ticketUpload, ticketRtt])
```

Algorithm 3. The Reducer algorithm.

3) *Dashboard Application*

The Dashboard application contains a geographic region map for visualization of the respective tickets. The tickets are as follows: green colored "Good" ticket, red colored "Bad" ticket and yellow colored "Average" ticket. We would like to note that the visualization is online to allow enough time for the decision makers to react in a timely manner.

B. Content Delivery Service Based Deployment

With the ever-increasing mass access to certain content (e.g., popular videos, recent albums, flash news etc.) from the Internet sources, most ISPs either mirror such contents or cache them for fast accesses for their subscribers. To do so, they run specialized content delivery servers and check their performance frequently. For the sake of the overall QoE, any bottleneck due to content delivery servers needs to be identified and resolved instantly. Unfortunately, Geographic Region Based Deployment cannot directly provide content delivery server based user experience. For this reason, our system has another deployment named Content Delivery Service Based Deployment.

The Content Delivery Service Based Deployment (Fig. 4) consists of three main components. First component is the logs collector interface running as a daemon service. The logs collector warehouses performance indicators of network centric measurement and connection stream data (e.g., media & web QoE metrics) on the content delivery servers. The Hadoop-based logs analyzer is the second component. The analyzer, implemented in Java, derives high-quality structured data from unstructured text using entity extraction. The third component, implemented in C#, is a GUI dashboard application presenting analysis results for proactive decisions. The dashboard application constantly monitors the network status and provides users with decision making for operational excellence through competitive intelligence reports.

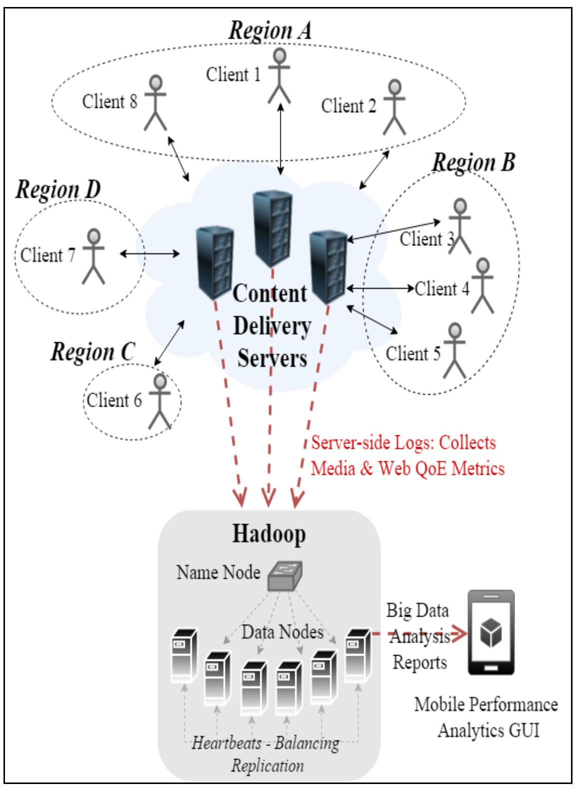

Fig. 4. Proactive ISP Speed Meter System (Content Delivery Category based deployment). It has three components; (i) server side logs collector (running on the server), (ii) a Hadoop based performance analytics tool, and (iii) a GUI dashboard application.

The user experience can be measured from server-side as it is possible to measure frequency of successful packet transmissions to any client device within the ISP's address range. Moreover, separate frequencies can be measured for each content delivery service, hence giving us the utilization of the respective service. The deployment (Fig. 4) collects performance indicators of network centric measurements and connection streams on the server-side. The connection logs are processed and analyzed in distributed fashion, which enabled us to easily and quickly process an extremely large amount of speed tests.

Fig. 5 shows the steps of the MapReduce operation which follows the same basic steps as given in Section 3.1.2.

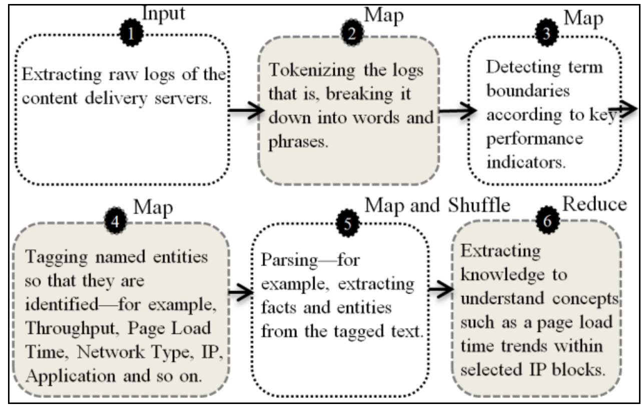

Fig. 5. Flow of MapReduce based Analysis.

Geographic Region Based Deployment requires client-side Speed Meters distributed across every geographic region and each of them needs to execute the speed tests constantly. That deployment has the advantage that user experience can be measured directly and accurately. However, it has the overhead of client-side installations and maintenance. Content Delivery Service Based Deployment, on the other hand, does not need client-side components as the objective with which is to measure performance of servers. It measures server-side initiated end-to-end network speed by listening actual user (subscriber) connections to content delivery servers.

The client-side analysis with Geographic Region Based Deployment provides only network centric KPIs of download/upload speeds and RTT. However, server-side analysis with Content Delivery Service Based Deployment provides us with richer set of KPIs, including congestion, page load time, TCP connection time, dropped frames, content category rate, time to playback, errors, DNS lookup time end, in addition to download/upload speeds and RTT. As a result, KPIs of the Geographic Region Based Deployment are purely network centric, while that of the Content Delivery Service Based Deployment are network centric plus user and competitive intelligence centric. Additionally, server-side analytics is designed to use real user connections. Monitoring web access performances combined with video analytics (video startup time, re-buffering, avg. bit rate) allows mobile ISPs to fully understand how websites load and videos play on the end-user device.

For some specific situations of the network problems, user and network centric KPIs may be analyzed, combined and correlated together. Because of that, server-side analysis system can enable us with analysis on different domains such as network centric, user centric and competitive intelligence centric. For instance, some problems may be solely caused by network centric, while some others may be caused by combinations of the network centric and user centric. Therefore, both network and user centric analyses offer advantages for the detection of network problems [24]. Like Geographic Region Based Deployment, Content Delivery Service Based Deployment analyzes big data to improve network services, user needs and businesses value [25].

IV. Experimental Evaluation

The proposed Internet Speed Analysis System is tested on the network of a leading mobile Internet ISP of Turkey. Our test network contains 20,500 base stations offering 3G and 4.5G services. This means that test results arrive in every minute from 20,500 stations and are pooled on the Hadoop platform. It is very difficult for a classical computing system to deal with such a high speed massive volume data. This is why we employed big data processing solutions including advanced analytics. This way we are able to transform the complex big data into actionable intelligence.

We experimented with both of the deployments and highlight a few interesting cases found during the live operation of the system. For the Geographic Region Based Deployment, we implemented the map with province granularity of Turkey which has 81 provinces. However, the implementation can be easily configured at cell, base station, neighborhood, county or specific region levels. In this way, the right granularity level will give more informative results.

A. Geographic Region Based Deployment

The most important feature of Geographic Region Based Deployment is scaling user experience and sentiments for each region separately.Provincial map of Turkey is shown in Fig. 6, where five interesting provinces (regions) are enumerated. The reason will be explained later.

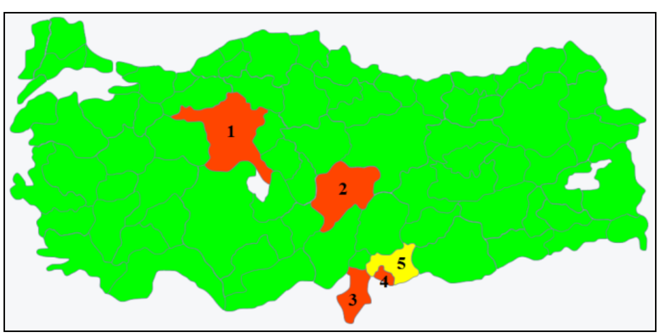

Fig. 6. Provincial map of Turkey, where five interesting provinces have been enumerated.

1) User Experience

The speed characteristics map as shown in Fig. 6 was obtained with plot analysis of test results, an example of which is given in Fig. 7. As shown in Fig. 7, download rates dropped sharply starting from July 29 in Region 1 (the plot for the Region 2 was similar). This is why Regions 1 and 2 in Fig. 6 are colored with red. The cause of the speed drop is being identified as an operational work (i.e., a maintenance of the base stations) carried out in these regions. These days, we experienced similar plots for Regions 3, 4 and 5. Since the speed drop is sharp in Regions 3 and 4, they are colored with red. However, Region 5 is colored with yellow as the speed drop in this region was moderate. It has been later understood that Syrian refugees caused an excess traffic in Regions 3, 4 and 5. This excess demand exceeded maximum capacity of the base stations in these regions.

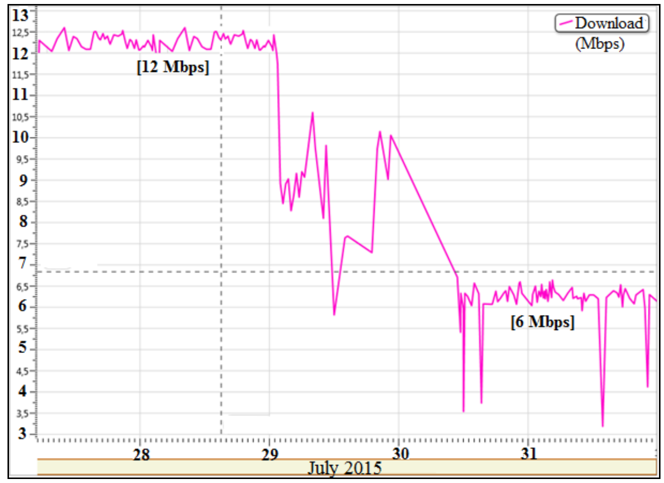

Fig. 7. Download Speed Rate Plot for Region 1.

Fig. 7-8 presents the correlation of our test results with the user feedbacks (complaints) from the Regions 1 and 2 of Fig. 6. Mobile users from the Regions 1 and 2 reported network speed problems on 29-31 July and 1-3 August. In these days, the fall of user satisfaction (Fig. 8) coincides with the decreasing download speed (Fig. 7). The agreement between Fig. 7 and 8 shows that our speed analysis system can proactively scale true user experience and network speed quality.

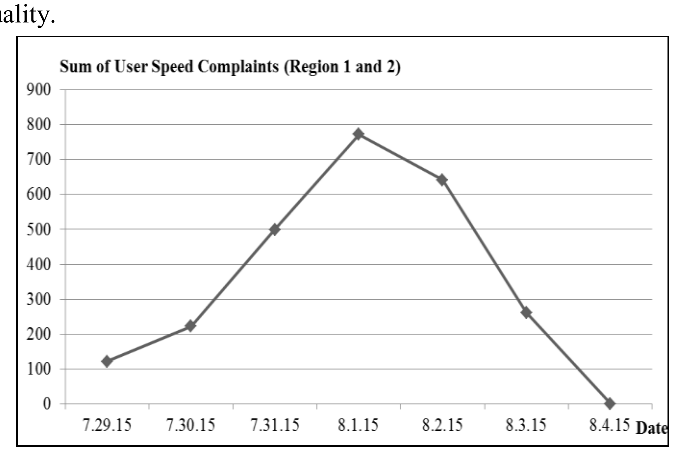

Fig. 8. Speed Satisfaction Feedbacks/Complaints.

2) Detection of Problems Before User Complaints

A Denial-of-Service (DoS) attack targeted the ISP servers providing Internet infrastructure, and as a result users encountered Internet access problems in some regions enumerated in Fig. 9. Users of the Regions 1, 2, 3, 4 and 5 are affected from the DoS attack for 18 minutes. Fig. 10 gives the network speed change for the Region 1 of Fig. 9. As shown in Fig. 10, the DoS attack started at around 12:00 and the speed was suddenly increased. Shortly after, some packet core network servers were unable to serve to mobile users. Our system was operational at that period and detected this DoS attack through user experiences and raised the red/yellow alarms for the respective regions. In a short time of minutes, system administrators were able to intervene. As a result, the problem is detected and fixed right before massive user complaints arrive. Around 12:18 the Internet usage/speed experience came to normal.

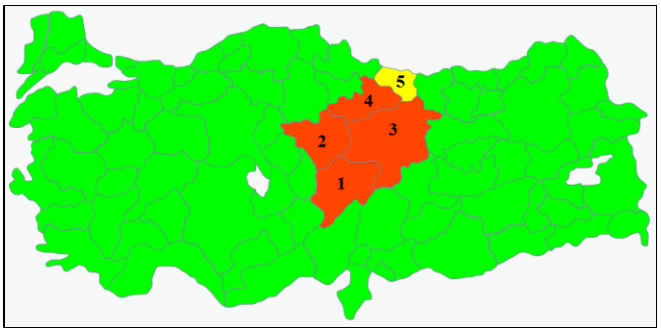

Fig. 9. DoS attack regions.

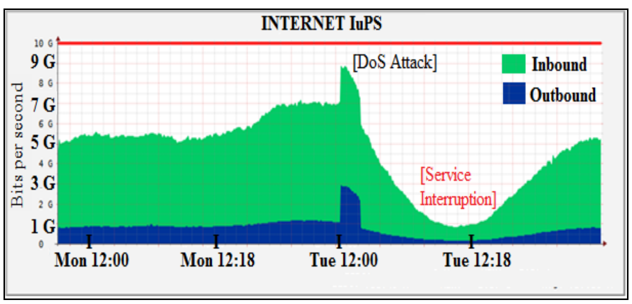

Fig. 10. Internet usage/speed in Region 1.

B. Content Delivery Service Based Deployment

The most important feature of Content Delivery Service Based Deployment is scaling user experience and sentiments for each service separately.

1) User Experience

Some of the web services of packet network were adversely affected after an operational work of maintenance. The affected regions are shown in red in Fig. 11. As a result, Internet usage rates quickly fell in Sunday at around 00:00 (Fig. 12).

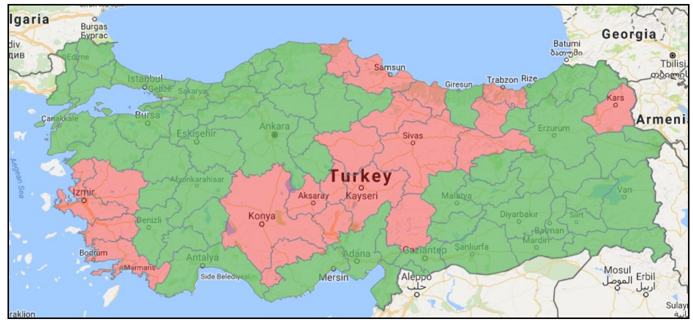

Fig. 11. Operational work in Internet service servers. The maintenance is carried out in regions marked with red.

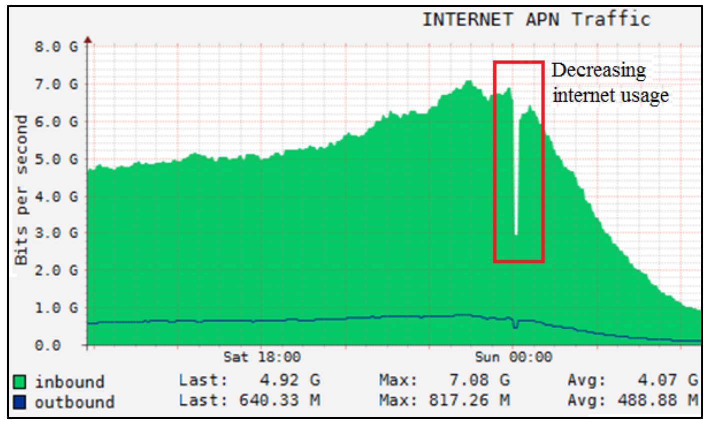

Fig. 12. Internet usage/speed in Region 1.

2) Competitive Intelligence

Competitive intelligence can be gathered from analyzing patterns of customer usage and mobile network data segmentation on the content servers. Our system is able to show reports about patterns of actual customer usage due to real user activities. Therefore, collected logs of the servers include all Internet activities of users such as click patterns, connection streams, performance indicators of network centric measurement, information of data segmentation and page views and similar. Hence, this deployment does not only scale user experience, but also analyses competitive intelligence.

In the months of March, April and May, the deployment was run on the content delivery network to analyze real Internet usage. For our experimental objective, we projected analysis results on different categories of service based, VOIP based, application based, video based and instant messaging based cases in the ISP's mobile network. The results of competitive intelligence are given in Fig. 13 through Fig. 17. Note that these figures are very informative as far as comparing the demand for the related services/tools/applications.

The competitive intelligence reports provide the ISPs with valuable information of highly demanded services. Hence, ISPs can use this information to decide on their future investment plan.

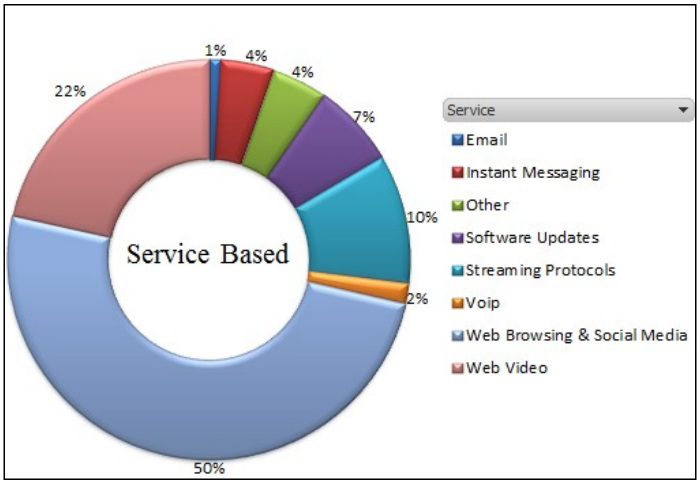

Fig. 13. Service Based Analysis.

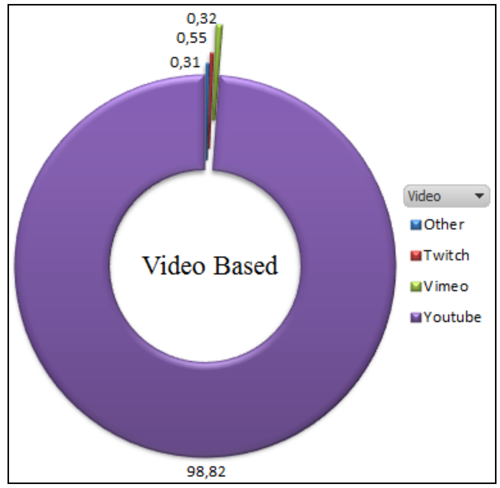

Fig. 16. Video Based Analysis.

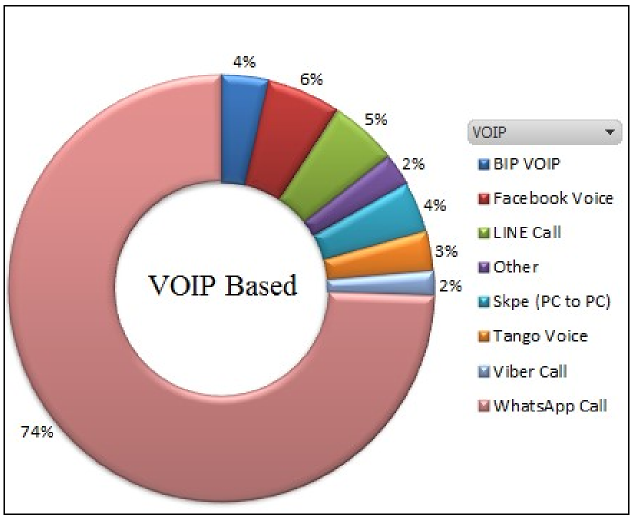

Fig. 14. VOIP Based Analysis.

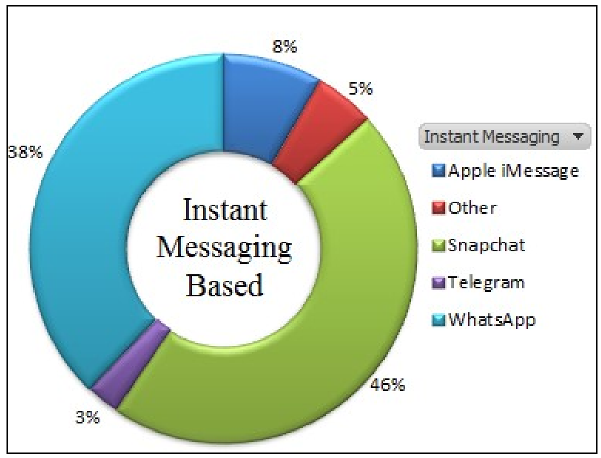

Fig. 17. Instant Messaging Based Analysis.

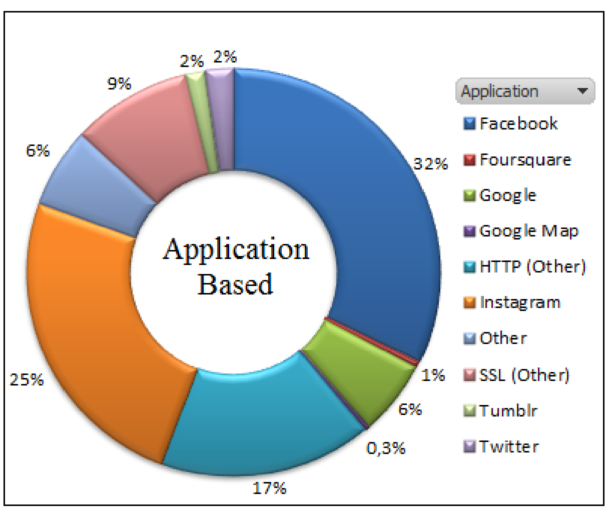

Fig. 15. Application Based Analysis.

V. CONCLUSION

Today wider user community services like social media, online gaming, video streaming and VoIP demand more network speed and high bandwidth capacity. For mobile ISPs, Internet service quality is a very strategic issue to satisfy user demands and to be competitive as well. Therefore, constantly monitoring Internet speed and true user experiences have the key role for customer satisfaction. However, correctly measuring the true user experiences in online fashion and promptly intervening the network after any service degradation involve the coordination of right speed measurement, right speed analysis and right decision making facilities.

In this paper, we proposed an Internet speed analysis system for mobile ISPs. The system called Proactive ISP Speed Meter exploits end-to-end network speed measurement to correctly gauge user experience. The system has two deployments: the client-side initiated geographic region based, and the server-side initiated content delivery service based. The former deployment analyzes Internet access speed from each geographic region while the latter deployment analyzes Internet access speed per content delivery category and can additionally provide competitive intelligence.

Big Internet traffic data coming from various clients of the mobile ISP network are processed and analyzed using

MapReduce programming model on Hadoop platform. The analysis results of regional speed characteristic maps and service usage plots are useful for operational excellence. Most importantly, our system is able to predict potential mass user complaints proactively and provide warnings in time.

The utility of our system has been demonstrated on a live network of a leading mobile ISP of Turkey.

REFERENCES

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf, Last accessed Feb 2020.
[2] H. D. Trinh, N. Bui, J. Widmer, L. Giupponi and P. Dini, "Analysis and modeling of mobile traffic using real traces," 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, ISBN: 978-1-5386-3531-5.
[3] R. Pries, F. Wamser, D. Staehle, K. Heck and P. Tran-Gia, "Traffic Measurement and Analysis of a Broadband Wireless Internet Access," Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th, Barcelona-Spain, ISBN: 978-1-4244-2517-4.
[4] W. Robitza, A. Ahmad, P.A. Kara, L. Atzori, M. G. Martini, A. Raake and L. Sun, "Challenges of future multimedia QoE monitoring for internet service providers," Multimed Tools Appl (2017) 76: 22243, https://doi.org/10.1007/s11042-017-4870-z.
[5] M. Uzun and O. Abul, "End-to-end internet speed analysis of mobile networks with MapReduce," 2016 International Symposium on Networks, Computers and Communications (ISNCC), Yasmine Hammamet, Tunisia, ISBN: 978-1-5090-0284-9.
[6] M. Kyryk, N. Pleskank and M. Pitsyk, "QoS mechanism in content delivery network," 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), 2016, Lviv, Ukraine, ISBN: 978-6-1760-7807-4.
[7] M. Taruk, E. Budiman, Haviluddin and H. J. Setyadi, "Comparison of TCP variants in Long Term Evolution (LTE)," 5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), 2017, Malang, Indonesia, ISBN: 978-1-5386-0355-0.
[8] L. Khoshnevisan, F. R. Salmasi and V. Shah-Mansouri, "An adaptive rate-based congestion control with weighted fairness for large round trip time wireless access networks," 24th Iranian Conference on Electrical Engineering (ICEE), 2016, Shiraz, Iran, ISBN: 978-1-4673-8789-7.
[9] F. Ahmed, J. Erman , Z. Ge , A. X. Liu , J. Wang and H. Yan, "Detecting and Localizing End-to-End Performance Degradation for Cellular Data Services Based on TCP Loss Ratio and Round Trip Time," IEEE/ACM Transactions on Networking, Volume: 25, Issue: 6, p: 3709 –3722, Dec. 2017, ISSN: 1558-2566.
[10] P. Skocir, D. Katusic, I. Novotni, I. Bojic and G. Jezic, "Data rate fluctuations from user perspective in 4G mobile networks," 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2014, ISBN: 978-9-5329-0052-1.
[11] A. S. Khatouni et al, "Speedtest-Like Measurements in 3G/4G Networks: The MONROE Experience," 29th International Teletraffic Congress (ITC 29), 2017, Genoa, Italy, ISBN: 978-0-9883045-3-6.
[12] http://www.speedtest.net/, Last accessed Feb 2020.
[13] L. Velasco, "Managing services in the telecom cloud: An example for CDN," 18th International Conference on Transparent Optical Networks (ICTON), 2016, Trento, Italy, ISBN: 978-1-5090-1467-5.
[14] S. Wang, X. Wang, J. Huang, R. Bie and X. Cheng, "Analyzing the potential of mobile opportunistic networks for big data applications," IEEE Network, Volume: 29, Issue: 5, September-October 2015, ISSN: 0890-8044.
[15] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," In Proceedings of the 6th Symposium on Operating System Design and Implementation (OSDI 2004), pages 137-150, San Francisco, California, 2004.
[16] J. Dean, and S. Ghemawat, "MapReduce: a flexible data processing tool," Communications of the ACM, 53(1), pp.72-77, 2010.
[17] Hadoop, http://hadoop.apache.org/, Last accessed Feb 2020.
[18] P. Zikopoulos and C. Eaton, "Understanding big data: Analytics for enterprise class hadoop and streaming data," McGraw-Hill Osborne Media, 2011.
[19] Y. Qiao, Z. Xing, Z. M. Fadlullah, J. Yang and N. Kato, "Characterizing Flow, Application, and User Behavior in Mobile Networks: A Framework for Mobile Big Data," IEEE Wireless Communications, Volume: 25, Issue: 1, February 2018, ISSN: 1536-1284.
[20] G. H. M. Almeida, E. H. R. Coppoli, E. N. Goncalves, M. M. Afonso and U. C. Resende, "Traffic flow management in a real mobile phone network using linear optimization," IEEE Latin America Transactions, Volume: 16, Issue: 2, Feb. 2018, ISSN: 1548-0992.
[21] Dumpcap, https://www.wireshark.org/docs/man-pages/dumpcap.html, Last accessed Feb 2020.
[22] R. Blum, "Network Performance Open Source Toolkit: Using Netperf, tcptrace, NISTnet, and SSFNet," John Wiley & Sons, 2003.
[23] Tcptrace, http://www.tcptrace.org/, Last accessed Feb 2020.
[24] M. Dighriri, G.M. Lee and T. Baker, "Measurement and Classification of Smart Systems Data Traffic Over 5G Mobile Networks," In: Dastbaz M., Arabnia H., Akhgar B. (eds) Technology for Smart Futures. Springer, Cham, ISBN: 978-3-319-60137-3, 2018.
[25] B. Zhou, J. Li, S. Guo, J. Wu, Y. Hu and L. Zhu, "Online Internet Traffic Measurement and Monitoring Using Spark Streaming," GLOBECOM 2017 - 2017 IEEE Global Communications Conference, Singapore, Singapore, ISBN: 978-1-5090-5019-2.
[26] J. Sandoval, A. Ehijo, A. Casals and C. Estevez, "New Model and Open Tools for Real Testing of QoE in Mobile Broadband Services and the Transport Protocol Impact: The Operator's Approach," IEEE Latin America Transactions, Volume: 13, Issue: 2, Feb. 2015, ISSN: 1548-0992.

BIOGRAPHIES

METE UZUN received his B.S. and M.S. degrees in computer engineering from the TOBB University of Economics and Technology. His research interests include mobile networks and machine learning.

OSMAN ABUL is a professor of Computer Science at TOBB University of Economics and Technology, Ankara, Turkey. He received his Ph.D. degree in Computer Engineering from Middle East Technical University, Ankara, Turkey. He held visiting posts in University of Calgary, Norwegian University of Science and Technology, and Italian Institute of Information Science and Technology. His research interests include data mining, privacy and bioinformatics.