

Analysis of End to End Internet Traffic in Education's Networks: A New Study

Wisam Dawood Abdullah*

Computer Science Department, Computer Science and Mathematics College, Tikrit University, Iraq.

E-mail: wisamdawood@tu.edu.iq

Ali Abdullah Ali

Minister Office of Higher Education and Scientific Research, Iraq.

Layth Rafea Hazim

Computer Science Department, Computer Science and Mathematics College, Tikrit University, Iraq.

Received September 18, 2021; Accepted December 16, 2021

ISSN: 1735-188X

DOI: 10.14704/WEB/V19I1/WEB19223

Abstract

Network Traffic Monitoring and Analysis (NTMA) is the main element to network management, especially to correctly operate large-scale networks such as the Internet on which modern academic organizations heavily depend. Their traffic use increases significantly because students, staff members, and research labs use them to search information. It is necessary to analyze, measure, and classify this Internet traffic according to the need of different stakeholders such as Internet Service Providers and network administrators. Moreover, bandwidth congestions frequently occur, causing user dissatisfaction. This study tries to find different characterizations such as data over hosts, countries, cities, companies, top-level domains, and servers. In addition, this is a new study to find out different patterns and levels of analysis from the device to its international requests. Our findings show that the highest traffic use is on Mondays and Wednesdays. Web server and DNS server drop in response to fault tolerance. Social networks consume most of the bandwidth, such as 42% Facebook followed by 22% WhatsApp in peak hours. The second most accessed sites are search engines. Google is the most used one. About 59% of the host cities are outside Iraq, in particular USA and the UK. In Amara and Baghdad cities, the requested sites are 51% and 49% overseas. About 40% of the traffic is provided by EarthLink Ltd. Communication Internet services (Iraq), 14% EdgeCast. 12% level3, 9% Facebook, 7% Google, Akamai-as and Microsoft-corp-msn-as-block. This study gives guidelines for network administrators to improve their performance and bandwidth at the educational networks.

Keywords

NTMA, Internet Traffics, Traffic Analysis, Education Network, Traffic Measurement.

Introduction

Understanding the use and operation of Internet services is critical. This understanding requires a Network Traffic Monitoring and Analysis (NTMA) (Abbasi, Shahraki, and Taherkordi 2021) (Čermák, Jirsík, and Laštovička 2016) (Alconzo et al. 2019) (Fahad et al. 2014) (Islam, M.R., Koirala, T.K., Khatun 2018). The applications include providing a network traffic view to detect anomalies and unknown attacks while the feeding systems are responsible for utilization, monitoring, and accounting (Sikos 2020). These systems collect historical data required for supporting traffic engineering and troubleshooting. This collection helps planning network evolution and identifying the root of the problem (Alconzo et al. 2019) (Bar et al. 2015). By 2020, cisco estimates that there will be 50 billion "connected" devices. Many people are connected to the Internet 24 hours a day. Cisco defines the Internet of Everything (IoE) as bringing together people, processes, data, and things (IoT) to make networked connections more relevant and valuable than ever before (Kunle and Olubunmi 2017) (Miraz et al. 2015) (Da Costa, Oliveira, and De Souza 2021). As a result, the analysis of network traffic is one of the main network functions to monitor effective network operations and management. Although thoroughly examined, online traffic is still challenging for many reasons. One of the primary challenges in the online traffic is the heavy traffic volume to be analyzed within a finite amount of time due to the increasing network bandwidth (Kim and Sim 2019) (Lee and Lee 2013) (Shamsudin, Katuk, and Abdullah 2017). The characterization of Internet traffic leads to various activities related to network management such as capacity planning and provision, traffic engineering, fault diagnosis, application performance, anomaly detection, and pricing (Wisam Dawood Abdullah 2012). More importantly, the difficulty of managing and securing the campus networks, an important component of the Internet ecosystem, allows cyber attackers to actively explore and compromise end hosts and mobile devices in these networks. It becomes extremely important to develop effective techniques to understand traffic patterns and behavioral dynamics of end hosts and applications in campus networks (Weng et al. 2016) (Das et al. 2014) (Hendawi et al. 2016). Universities also tend to use Voice over IP (VoIP), e-commerce, video conference, file sharing, google classroom, google form, cisco webex, etc. These activities negatively influence network performance particularly during peak working hours in academic institutions. Traffic identification means recognizing traffic which should happen before traffic classification.

Studies of wireless and wired networks are less concentrated on the users and on the comprehensive characterization of requested sites and overseas. Ibrahim et al. (Ibrahim et al. 2016) reviewed the most flexible Internet traffic tools of measurement and analysis to manage the dynamicity of data transferring characteristics. Wu et al. (Wu et al. 2016) have studied the protocol efficiency, delivery of frame, application types and IEEE 802.11 variants (i.e.a/b/g/n/ac/ad). This study analyzed the traffics on two different IEEE 802.11 operating networks: one on campus and the other out the campus. Hendawi et al. (Hendawi et al. 2016) studied the Internet protocol (IP) and Transport Control Protocol (TCP). Adib et al. (Abdullah et al. 2017) concentrated on daily captured behaviors of users, such as application traffic, packet lengths, protocols, loads. Hafiz et al. (Shafiq and Mehmood 2018) analyzed transport, application, and network layer to find a variety of features. They studied multimedia sites like YouTube which use large bandwidth and followed by social media sites, where Facebook is the most widely used. Bhandari et al. (Bhandari et al. 2018) focused on packet analysis and packet sniffing over TCP connection. TCP time-sequence graph, TCP Throughput graph, and TCP round trip time graph are analyzed using Fireshark software (Saxena, P., & Sharma 2017). Rosa and Kadir (Rosa and Kadir 2018) conducted a data analysis of traffic using network traffic behavior method and history of connected traffic. However, comprehensive information on the used internet traffic is monitored for the analysis. In this paper, application traffic (such as social networks, Education, Search engine, Email, Multimedia) are analyzed. Also, various parameters are shown such as data distribution over hosts, web and Domain Name System (DNS) servers, numbers of byte per second, data distribution over countries then thoroughly examine traffics data over cities and companies, and top-level domain statistics. Furthermore, it is clear that widely accessed cites are the web.

There are multiple tools available for capturing Internet traffic such as Tcpcmdump(Goyal and Goyal 2018) (Varanasi and Swathi 2016), Wireshark (Asst et al. 2019) (Siswanto et al. 2019), Ethereal, Ntop, Network Grep (Ngrep), Ipsumdump, etc.(Siswanto et al. 2019), which will capture and store the traffic locally or on remote machines (for scalability) (Kaur and Misra 2019). When Internet Service Providers (ISPs) or network administrators need to analyze, measure or classify this traffic, the proper data management would be required as well as access to the remote machines. The Internet traffic data collected is very huge and will continue to grow. Academic institutions regularly update their Internet services. However, applications of web 2.0 are spreading very quickly. Also, their extensive usages consume high bandwidth, causing congestion of network and degradation of performance. Therefore, it is vital to investigate the use of network resources and Internet traffic flow distributions in educational networks (Abdullah et al. 2017). In Iraq, universities are

connected to the internet by different local companies such as EarthLink, Scope sky, Giganet, etc. It is important to observe and understand networks by measuring them. Many studies perform passive and active measurement modes (Abdullah et al. 2017). This study concentrates on characterizing Internet traffic on the basis of users' preferences in the main campus of Tikrit University (TU) as a case study. Then, the result could be utilized for the generation of guidance and offering a suggestion to help university and institute learning. This university works on providing Internet for more than 29,000 students and 6,500 staff members. Moreover, it is connected to the Internet through EarthLink as the Internet provider. Around 220 GB has been captured from students and staff members.

Data Collection

Lenovo ThinkPad is utilized as a capturing device which is linked to Cisco's catalyst 3750 (which is the switch of the Internet in Cisco Networking Academy supporting mirror porting). This switch is attached to computer center switches. Internet switch port fa1/0/9, fa1/0/17, fa1/0/23, and dot1q encapsulation is mirrored to 1/0/13 port. tcpdump, wireshark, geoLite ASN, city, country, and tcpstat have been used for capturing complete outgoing traffic on the TU gateway link. Then, data is analyzed on a different device to prevent overload of CPU on the gateway. Also, complete traffic in a series are captured where pcap file substitutes big files. This approach contributes to completing this study and in improving the path and dropping incomplete information. The capturing time is 9:00 am to 13:00 pm for one working week on 21, 22, 23, 24, and 25 Mar 2021. The average captured data size a day was about 220 GB as illustrated in Table 1.

Table 1 Captured data information

Days	Date	Time Interval	Size
Sunday	21 / 3 / 2021	9:00am-13:00pm	
Monday	22 / 3 / 2021	9:00am-13:00pm	
Tuesday	23 / 3 / 2021	9:00am-13:00pm	220 GB
Wednesday	24 / 3 / 2021	9:00am-13:00pm	
Thursday	25 / 3 / 2021	9:00am-13:00pm	

Data Analysis

In order to analyze traffic, the feature of Wireshark, geoLite ASN, city, country, tcpstat, geoIP domain name, and netperf have been used in this paper. The capturing tools created a lot of raw data from the levels of connection to application and to the destination levels such as; cites, countries, and companies that are responsible for these traffics or activities. The most important traffics that have been used for analysis are application, overall statistics, traffic analysis for various website categories, geocity analysis, geocountry

analysis, geocompany analysis, server, top-level domains, and performance. Each traffic contains different information and that helps the administrator, network engineer, and cybersecurity staff to enhance the topology and network devices according to the future Internet. Finally, gnuplot tools and excel are utilized to visualize data.

Method

This section explains the proposed method that has been used to analyze the data. Firstly, two main aspects, network topology, and devices are considered. The choice of a suitable place to collect the packet is important to prevent capturing irrelevant packets. Secondly, a complete description is drawn of how we analyze the data distribution of various statistics. Thirdly, how to obtain the distribution of the website category is described. Finally, the locating Geo-location of the Iraqi Internet traffic is explained.

1. Network Environment

Tikrit University (TU) network works on a switched network during the collection of the data. This indicates that the University's Internet center utilizes a single subnet. The cisco academy network is connected to the Internet with a cisco switch catalyst 3750 which is a virtual network on a particular subnet different from that of the campus. Moreover, the cisco academy and computer center are in the same building. Additionally, every building within the campus is provided with Nano beam or lite beam that is connected to a Switch, linked to the Core switch in the Computer Center.

2. Traffic Analysis with Total Statistics

In this study, basic details about packet levels e.g. IP, protocol, bytes transferred and bytes numbers are provided. To obtain statistics of protocols such as HTTP, HTTPs, DNS, LLMNR, etc., benchmarking tools from traffics with protocol utilized information of the received bytes are used. For the transport level time-series of different protocols such as; TCP and UDP, similar steps are conducted. For a complete analysis at various levels, procedures similar to that of statistics other than traffics distribution are followed. The study processes and analyzes data by Linux.

3. Analysis for Different Website Categories

This study obtained the distribution of websites of different categories such as education sites, search engines, social networks, multimedia, email, etc. For instance, social networks contain various common social websites such as; Facebook, Twitter, Viber, and WhatsApp, etc. We find their overall distribution for complete operating hours. Furthermore, managing

multiple DNS names for common websites, www.example.com and www.example.com.pk, is required to prevent multiple results for one website. This issue is also managed in the proposed method for each given website.

4. Traffic Analysis Over Geolocation

To locate traffics activity accessed by TU users, host name look-up queries have to be conducted. Thus, we locate Geo- statistics of the accessed traffics i.e., the number of the hosted traffics in Iraq and abroad, the ones in Baghdad and Amara cities, and in other cities, the traffics provided by Earthlink company and the number of overseas companies. With third party Python API pygeoip, freely available Geo-IP information GeoLite database is employed for this purpose.

5. Analysis of Domains Statistics

The statistics of Top-Level Domains (TLD) for the TU used websites have been found out such as;.com domain,.edu domain,.org domain,.net domain, etc.

Results and Discussion

In general, firstly, web server, and domain name server statistics are established. Then, application, transport, and network layers' important protocols statistics are discussed. This is followed by presenting different website categories. Finally, the Geo cites, countries and companies' information of the used websites are described with the statistics of the TLD.

1. User Application Popularity and Overall Statistics

Nowadays, network traffics analysis and knowledge extraction have become the most important studies due to the large numbers of traffics and the difficulties of extracting knowledge from them. The analysis of the user application and overall statistics of traffics used by TU distribution switch and the specification of the traffic, showed that in Monday the highest traffics (39%) is used in an hour because this day is busy with assignments and lectures, followed by Wednesday with 27%. Then, 21% on Sunday and the lowest percentage was on Thursday. Moreover, the percentage of traffics fluctuates during the week. Where we observed that the highest rates are on Mondays, then it declines on Tuesdays, and rises on Wednesdays and then decreased on Thursdays. Fig 1. describe the entire number of bytes per second.

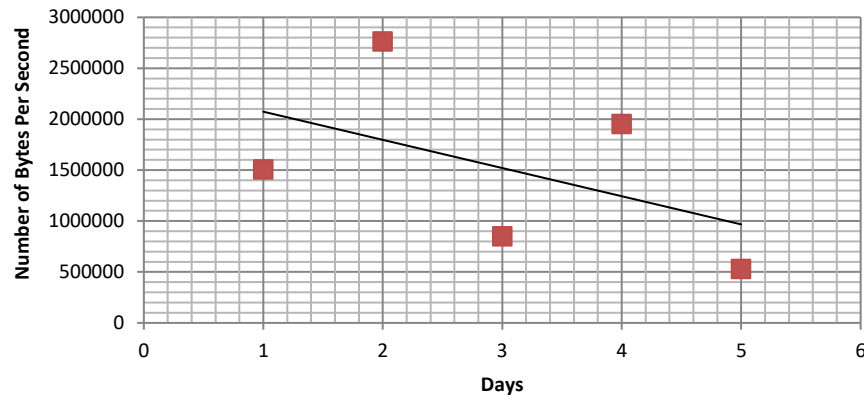


Fig. 1 Number of Bytes A Second

Fig. 2 shows the classification of the traffic according to IP protocols and applications. Connection-Oriented (TCP) is 82% of the total bytes transferred, followed by connectionless (UDP) (14%) and web browsing (HTTP) (2%). Also, it is noticeable that there are some NetBois NBNA service, LLMNR, DNS, HTTPS, and ARP traffics, although it is 1% of the whole transferred traffics.

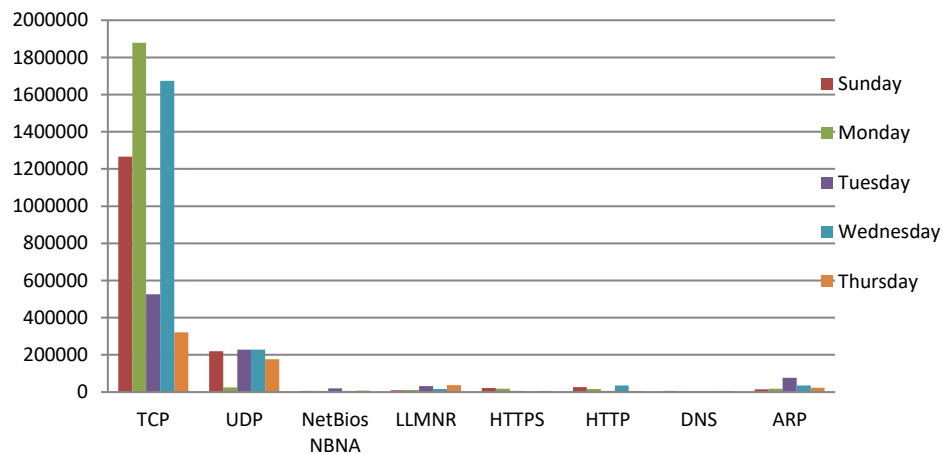


Fig. 2 The User Traffic by IP Protocols and Application

The findings also reveal that the used rates of TCP and HTTP protocols are high, with several errors in the TCP protocol such as out-of-order segment, Zero window packets and Duplicate acknowledgment among others.

Fig.3 shows a proportion of web server services and domain name server services in a week. Here, the web server request is 22% and the web server response is 10%, while the number of drops at web server is 11% of the web server traffics. This means that there is bottleneck and more fragmentation in the network topology. 29% of the requests are created by domain

name servers, 5% of the requests are responded by the server, 23% of traffic is dropped as result to the big problem in physical topology in campus.

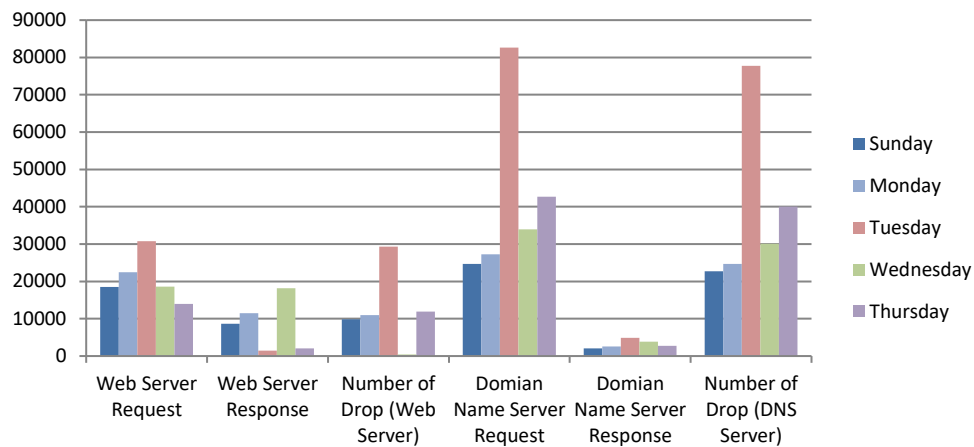


Fig. 3 The Proportion of Web Server and DNS Server

2. Websites Category Distribution Statistics

The application categories and signatures (domains) are shown in a week’s period as illustrated in table 2. This helps to show the proportion of the application traffic, the applications with more interactive users, the applications consuming more TU network bandwidth. It is clear that social networking site traffic is 34% (which is the highest) of the total traffic, in comparison to other sites in five days within an hour. This could be due to the provision of social networking sites for video-sharing devices and URLs, political news, and COVID 19 pandemic in Iraq and worldwide by Multimedia which is indispensable to staff and students. Also, educational sites account for 17% of traffic while Email traffic is 6%. Fig 4 depicts the most preferred categories.

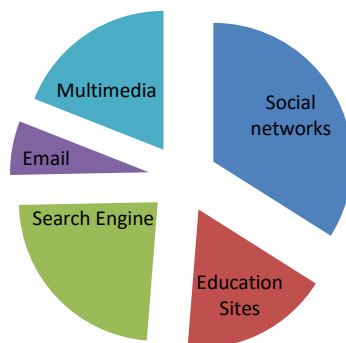


Fig. 4 Preferences of Users

Table 2 Application types and signature

Category	Items
Social Networks	Facebook; Twitter; Viber; WhatsApp; Telegram; LinkedIn
Education Sites	tu.edu.iq; ca.tu.edu.iq; cic.tu.edu.iq; mohesr.gov.iq
Search Engines Sites	Google; Yahoo; Google Scholar; Scopus; IEEE Xplore; Science Direct; SCI-HUB
Email	tu; Gmail; Yahoo
Multimedia	YouTube; Flickr; Snapchat; Vine; Instagram

The findings show that social network sites and video streaming are the most frequently visited sites, indicating the need for limiting such access during business and studying hours. The highest bandwidth consumption quantity starts with the social network, then search engines, multimedia, education, and email. Also, a good balance in the traffic ratios and a good rise in education sites compared to previous studies are reported. This indicates a distinct use of education sites.

Most teaching and learning related activities using web 2.0 tools are shown in the following figures. The percentage of each social website for studying period has been given in Fig 5. Facebook is the highest. This is expected, since Facebook is a worldwide application and one of the first sites that established the idea of sharing among users.

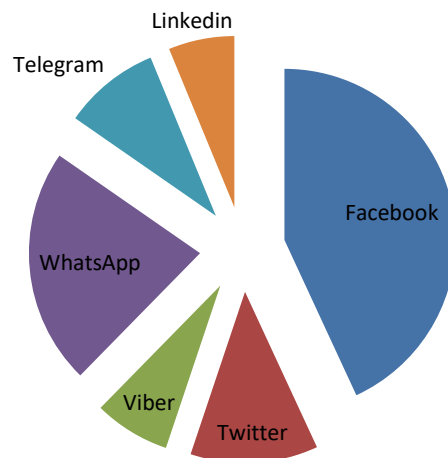


Fig. 5 Social Networking Sites Distribution

From the total social traffic, it appears that 42% of the user's access Facebook site. Thus, it is the first common used application, followed by WhatsApp 22%, Twitter 12%, Telegram 9%, Viber 7%, and LinkedIn 6%.

Fig 6 shows that "tu.edu.iq" is 45%, contains all services of this domain, then "netacad.com" are up by 25% because the cisco academy has lots of classes in that time with exercises.

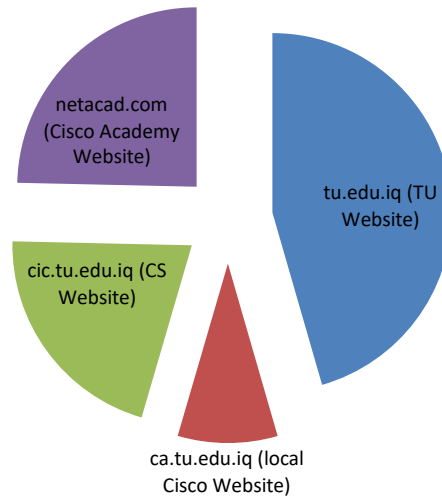


Fig. 6 Educational Websites Distribution

Fig 7 illustrates the statistics for each search engine. Google as a percentage of search engines for the period is 74%. The next frequent use is 14% Google scholar, IEEE Xplore 5%, sci-hub took 3%, sciencedirect and scopus are 2 %.

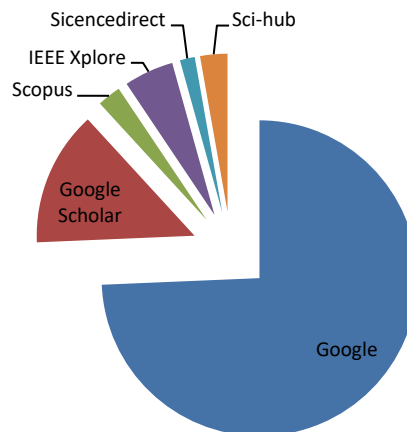


Fig. 7 Search Engines Websites Distribution

Fig 8 quantifies the use of email websites from 9:00 AM to 13:00 PM daily with 70% gmail, 27% tu.edu.iq, and 3% yahoo.

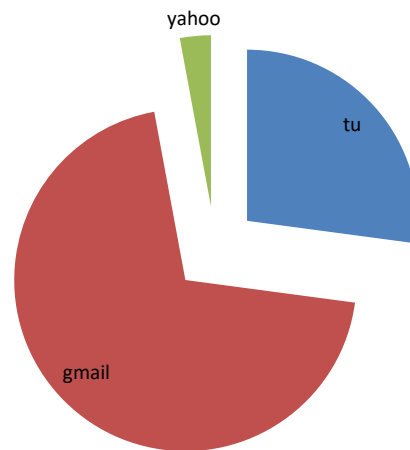


Fig. 8 E-Mail Websites Distribution

In Fig 9, Multimedia uses are expressed in percentages. The highest is YouTube which is 84% followed by 11% Instagram, 3% Snapchat and 1% to vine and flicker.

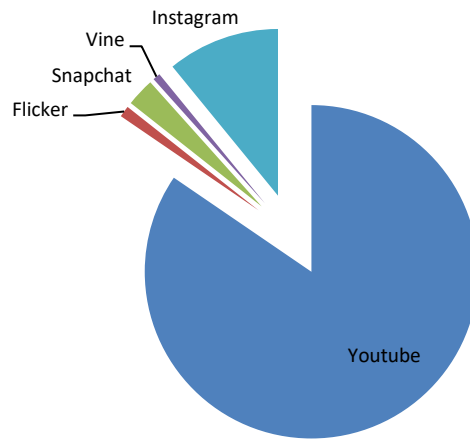


Fig. 9 Multimedia Websites Distribution

3. Traffic Analysis Over Countries

About 41% of traffic is in Iraq, 37% in the United States, 15% in the United Kingdom, 3% in Ireland, 2% in Russia, 1% in Germany, 1% in Oman, and 1% in other countries. In addition, the traffics is distributed over 18 countries during the week as shown in Fig. 10.

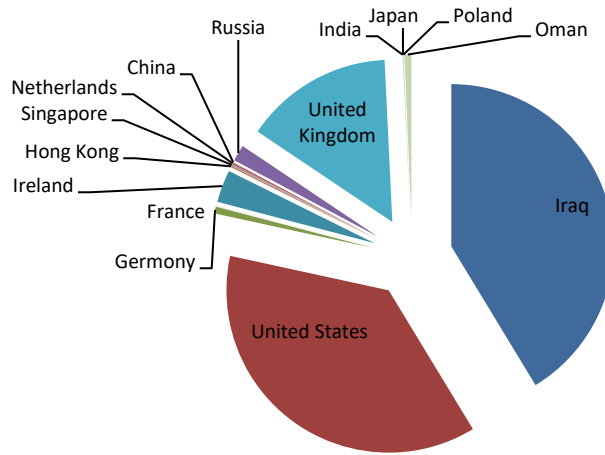


Fig. 10 TU Traffic Distribution Over Countries

At the countries' level, Iraq composes the highest level of data, this is normal because of the study nature in Iraq. While at the international level, the United States percentage is the highest traffic because the main highest rate traffic website servers such as Google, Facebook, and Twitter, are located in the USA, then the United Kingdom comes third.

4. Traffic Analysis Over Cities

The traffic analysis over cities shows that about 44% of traffic belongs to Amara (Iraq) and, 20% to US cities, Additionally, the UK has 15%. This is followed by 7% to Baghdad and 7% Frankfurt am Main, 3% to Dublin, 2% to Moscow, and other countries are less than 1%. In addition, the traffics is distributed over 27 cities in a week as illustrated in Fig. 11.

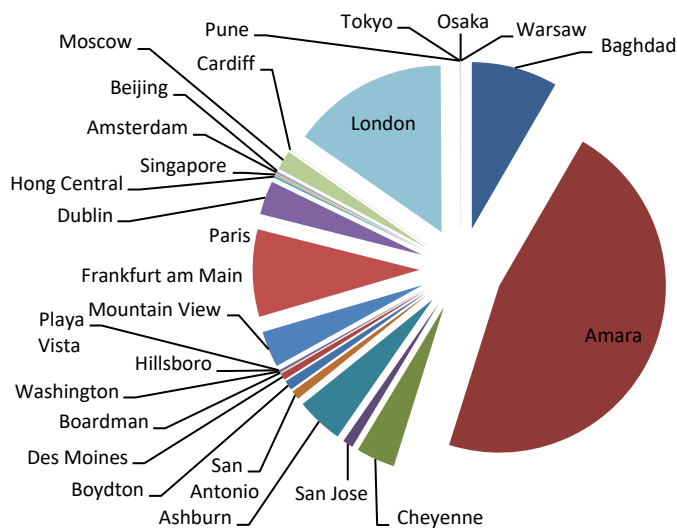


Fig. 11 TU Traffic Distribution Over Cities

At the city level, the city of Amara in Iraq is the highest because most of the servers are provided by EarthLink Company which is in Amara. While at international cities, the cities of the United States have the highest ratio, then London, after that Baghdad, Frankfurt and Main.

5. Traffic Analysis Over Companies

EarthLink Ltd Communication Internet Services in Iraq account for 40%. This is followed by 14% EdgeCast. 12% level 3, 9% Facebook, 7% Google, Akamai-as, and Microsoft-Corp-Msn-as-block. Other companies comprise less than 1%. In addition, the traffics are divided into 28 companies during the week as illustrated in Fig.12.

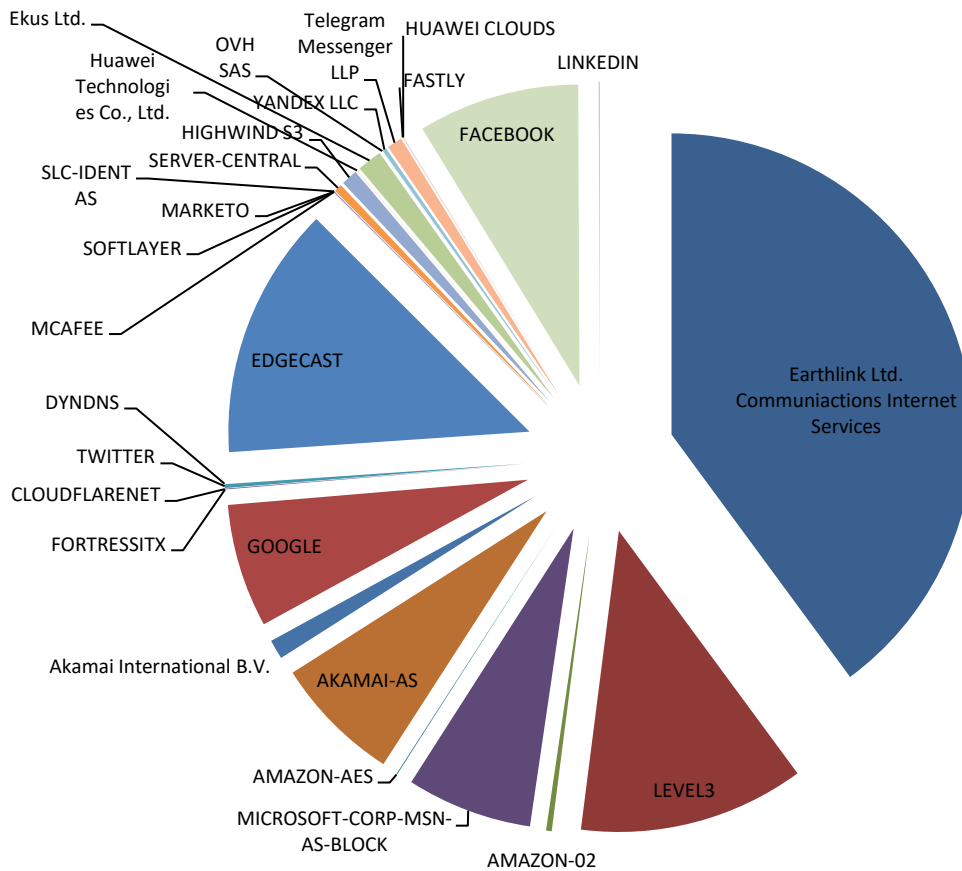


Fig. 12 TU Traffic Distribution Over Companies

At the level of technical companies, EarthLink Ltd. Communication Internet Services in Iraq have the highest percentage of data, then EdgeCast, level3, Facebook, Google, Akamai-as, and Microsoft-corp-msn-as-block. Because most of the educational institutions

in Iraq, including TU, take the Internet service from EarthLink Company, this company has taken the highest percentage.

6. Traffic Analysis Over Top Level Domains

In the TLD analysis of TU users, the bytes of ".com" extension and ".edu" extension are 75% and 16%, respectively. Other observed top TLDs are 4% to ".net" extension, 3% ".info" extension, 1% to ".org" and UK extensions, and less than 1% is ".gov" extension. Most traffic is to ".com" extension, then to ".edu" extension as shown in Fig.13. These percentages are very good considering that the education extension has the second rank.

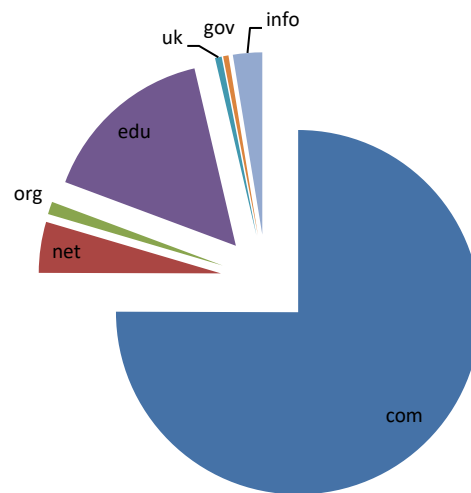


Fig. 13 Traffic Analysis Over TLD

Conclusion

This study is the first one in analyzing traffics for an Iraqi higher education institution. It includes the overall statistics of the main protocols in network. And the application traffic has been examined to measure the extent of their impact on network performance. Additionally, we have analyzed data traffic and measured its performance at the level of countries, cities, companies, and TLD. Such study helps researchers and engineers to develop network elements such as devices, media, protocols, and message. It also extracts knowledge from this data, which enables administrators and engineers to set a correct path in problem-solving. In this study, 220 GB from 6500 staff members are captured. The study aims to determine the time, location, traffic pattern, and cost for the utilized network of educational institutions. It is obtained that users mostly prefer social networking sites and

video streaming. This indicates the need of reducing such type access during working or studying hours.

Some modifications may be required to policies, such as firewall blocks in social networking sites during business and studying times. The application to strict bandwidth to restrict the use of the Internet and monitoring the network to discover the behaviour of new applications may not be useful in the scientific field. These findings could be used as a reference for future comparative studies in network performance. This could help network administrators to improve their network efficiency by identifying negative Internet resources. Also, they provide valuable information for network engineers to design and improve network performance upon requirements. The study divides the application traffics into five types: social networks, education sites, search engines, multimedia, and email. The social networks are used more than search engines, multimedia, education, and email. However, a thorough examination of each category shows that Facebook access is higher than all social network uses, and the Google search engine is higher among all the search engines, while the YouTube site is the highest in the multimedia, TU website is accessed more than other education sites, and Gmail is the most frequently used email. Also, the TCP protocol reaches the highest point during the study period followed by the UDP. Furthermore, the analysis of the application protocol shows that HTTP is the highest, followed by ARP, LLMNR, NetBios NBNA service, and DNS. In the countries of the study, Iraq recorded the highest rate in traffic, then USA, and the UK. Moreover, at the city level, the study reports that Amara has the highest rate of traffic, then the US cities. EarthLink Ltd. Communication Internet Services (Iraq) exceeds other companies, then EdgeCast, level3, Facebook, Google, Akamai-as, and Microsoft-corp-msn-as-block. Finally, in TLD, most traffics belongs to the ".com" extension, then ".edu" extension. Studies, like this, help researchers and engineers to develop network elements such as devices, media, protocols, and messages. It also extracts knowledge from this big data, and enables administrators and engineers to set a correct path in solving the problem.




References

- Abbasi, M., Shahraki, A., & Taherkordi, A. (2021). Deep learning for network traffic monitoring and analysis (ntma): A survey. *Computer Communications, 170*, 19–41.
<https://doi.org/10.1016/j.comcom.2021.01.021>
- Abdullah, W.D., MonzerHabbal, A.M., & Mahmuddin, M.B. (2017). Evaluation of user behavior and network performance in Malaysian Institution of Higher Education (MIHE) of wireless network. *In Annual Conference on New Trends in Information & Communications Technology Applications (NTICT)*, 46-51.

- D'Alconzo, A., Drago, I., Morichetta, A., Mellia, M., & Casas, P. (2019). A survey on big data for network traffic monitoring and analysis. *IEEE Transactions on Network and Service Management*, 16(3), 800-813. <https://doi.org/10.1109/TNSM.2019.2933358>
- Bär, A., Finamore, A., Casas, P., Golab, L., & Mellia, M. (2014). Large-scale network traffic monitoring with DBStream, a system for rolling big data analysis. In *IEEE International Conference on Big Data (Big Data)*, 165-170.
- Bhandari, A., Gautam, S., Koirala, T.K., & Islam, M.R. (2018). Packet sniffing and network traffic analysis using TCP—A new approach. In *Advances in Electronics, Communication and Computing*, Springer, Singapore, 273-280.
- Čermák, M., Jirsík, T., & Laštovička, M. (2016). Real-time analysis of NetFlow data for generating network traffic statistics using Apache Spark. In *NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium*, 1019-1020. <https://doi.org/10.1109/NOMS.2016.7502952>.
- Farias Da Costa, V.C., Oliveira, L., & De Souza, J. (2021). Internet of Everything (IoE) Taxonomies: A Survey and a Novel Knowledge-Based Taxonomy. *Sensors*, 21(2), 568. <https://doi.org/10.3390/s21020568>
- Das, A.K., Pathak, P.H., Chuah, C.N., & Mohapatra, P. (2014). Contextual localization through network traffic analysis. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, 925-933. <https://doi.org/10.1109/INFOCOM.2014.6848021>
- Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A.Y., & Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, 2(3), 267-279. <https://doi.org/10.1109/TETC.2014.2330519>
- Goyal, P., & Goyal, A. (2017). Comparative study of two most popular packet sniffing tools-Tcpdump and Wireshark. In *9th International Conference on Computational Intelligence and Communication Networks (CICN)*, 77-81. <https://doi.org/10.1109/CICN.2017.8319360>
- Hendawi, A.M., Alali, F., Wang, X., Guan, Y., Zhou, T., Liu, X., & Stankovic, J.A. (2016). Hobbits: Hadoop and Hive based Internet traffic analysis. In *IEEE International Conference on Big Data (Big Data)*, 2590-2599.
- Ibrahim, L.T., Hassan, R., Ahmad, K., Asat, A.N., & Omar, H. (2016). Online traffic measurement and analysis in big data: Comparative research review. *American Journal of Applied Sciences* 13(4), 420-431. <https://doi.org/10.3844/ajassp.2016.420.431>
- Islam, M.R., Koirala, T.K., & Khatun, F. (2018). Network traffic analysis and packet sniffing using UDP. In *Advances in Communication, Devices and Networking*, Springer, 907-914. https://doi.org/10.1007/978-981-10-7901-6_97
- Kaur, P., & Misra, N. (2019). A Methodical Review on Network Traffic Monitoring & Analysis Tools. *A Journal of Composition Theory*, 12(9), 1964-1968.
- Kim, J., & Sim, A. (2019). A New Approach to Multivariate Network Traffic Analysis. *Journal of Computer Science and Technology*, 34(2), 388-402. <https://doi.org/10.1007/s11390-019-1915-y>

- Kunle, O.J., Olubunmi, O.A., & Sani, S. (2017). Internet of things prospect in Nigeria: Challenges and solutions. *In IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON)*, 736-745.
- Lee, Y., & Lee, Y. (2012). Toward scalable internet traffic measurement and analysis with hadoop. *ACM SIGCOMM Computer Communication Review*, 43(1), 5-13.
<https://doi.org/10.1145/2427036.2427038>
- Miraz, M.H., Ali, M., Excell, P.S., & Picking, R. (2015). A review on Internet of Things (IoT), Internet of everything (IoE) and Internet of nano things (IoNT). *In Internet Technologies and Applications (ITA)*, 219-224. <https://doi.org/10.1109/ITechA.2015.7317398>
- Rosa, S.L., & Kadir, E.A. (2018). Abnormal internet usage detection in LAN Islamic University of Riau Indonesia. *In Proceedings of the International Conference on Intelligent Science and Technology*, 17-22.
- Saxena, P., & Sharma, S.K. (2017). Analysis of network traffic by using packet sniffing tool: Wireshark. *International Journal of Advance Research, Ideas and Innovations in Technology*, 3(6), 804-808.
- Shafiq, H.M., & Mehmood, M.A. (2018). Internet Traffic Analysis of an Educational Network using Bro IDS. *In International Conference on Frontiers of Information Technology (FIT)*, 76-81. <https://doi.org/10.1109/FIT.2018.00021>
- Shamsudin, S., Katuk, N., & Abdullah, K. (2017). Analysis of wireless network usage at Universiti Utara Malaysia: a preliminary study towards bandwidth management. *In Proceedings of the International Conference on High Performance Compilation, Computing and Communications*, 102-106. <https://doi.org/10.1145/3069593.3069613>
- Sikos, L.F. (2020). Packet analysis for network forensics: A comprehensive survey. *Forensic Science International: Digital Investigation*, 32.
- Siswanto, A., Syukur, A., & Kadir, E.A. (2019). Network traffic monitoring and analysis using packet sniffer. *In International Conference on Advanced Communication Technologies and Networking (CommNet)*, 1-4. <https://doi.org/10.1109/COMMNET.2019.8742369>
- Varanasi, A., & Swathi, P. (2016). Comparative Study of Packet Sniffing tools for HTTP Network Monitoring and Analyzing. *International Journal of Science, Engineering and Computer Technology*, 6(12), 406-409.
- Weng, W., Lei, K., Xu, K., Liu, X., & Sun, T. (2016). Internet Traffic Analysis in a Large University Town: A Graphical and Clustering Approach. *In International Conference on Web-Age Information Management*, Springer, 378-389.
- Abdullah, W.D. (2012). *Characterization of Internet Traffic in UUM Wireless Networks* (Doctoral dissertation, Universiti Utara Malaysia).
- Wu, S., Abed, S.H., Yang, Q., & Wang, H. (2016). IEEE 802.11 Traffic Measurement and Analysis. *In Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications*, 1-6.

Biographies of Authors

	<p>Wisam Dawood Abdullah received his B.Sc. degree in computer science from Tikrit University, Iraq and his M.S. degree in Information Technology (with concentration in Telecommunications and Networks) from the University Utara Malaysia (UUM). He received an expert certification from Cisco Networking Academy CCNP, CCNA, CCN Security, IoT, Entrepreneurship, Grid, Voice, Wireless Cloud, Linux, CCN Cybersecurity and IT, also he is a NetAcad administrator in Cisco Networking Academy Iraq, currently he is lecturer in the Tikrit University, Cisco Networking Academy member in IEEE. Research interest: Protocol Engineering, Network Analysis, Intern Architecture and Technologies, Wireless Performances, Network Traffic Engineering, Data Mining, Future Internet, Internet of Things, AI, ML.</p>
	<p>Ali Abdullah Ali received his M.Sc. degree in computer science from Iraqi Commission for Computers and Informatics, Iraq. Currently he is an assistant prof. in the Minister Office of Higher Education and Scientific Research, Baghdad, Iraq. His research and professional interests include network communication and security technologies.</p>
	<p>Layth Rafea Hazim is an Iraqi assistant teacher at the Cisco Networking Academy, Tikrit University, Iraq. He received his BSc degree in Computer Science from Tikrit University in 2007, M.Sc. degree from the Altinbas University, Turkey in 2018. He is a lecturer in the Tikrit University, College of Computer Science and Mathematics, Department of Computer Science He worked as a head of Electronic Computer Center ECC at Tikrit University during the period 2020 until now.</p>