Network Traffic Characterisation Using Flow-Based Statistics

Petr Velan, Jana Medková, Tomáš Jirsík, Pavel Čeleda
Institute of Computer Science
Masaryk University
Brno, Czech Republic
{velan, jana.medkova, jirsik, celeda}@mail.muni.cz

***Abstract*—Performing research on live network traffic requires the traffic to be well documented and described. The results of such research are heavily dependent on the particular network. This paper presents a study of network characteristics, which can be used to describe the behaviour of a network. We propose a number of characteristics that can be collected from the networks and evaluate them on five different networks of Masaryk University. The proposed characteristics cover IP, transport and application layers of the network traffic. Moreover, they reflect strong day-night and weekday patterns that are present in most of the networks. Variation in the characteristics between the networks indicates that they can be used for the description and differentiation of the networks. Furthermore, a weak correlation between the chosen characteristics implies their independence and contribution to network description.**

I. Introduction

A lot of research is being performed in the areas of network traffic classification, anomaly detection, and network security in general. Researchers involved in these areas often evaluate their methods using live network traffic. However, performing research on live network traffic includes several caveats. First, repeating experiments on live traffic is infeasible. Second, the traffic has to be thoroughly described, since most methods heavily depend on the properties of the observed traffic. Last but not least, the network traffic's properties can change during the research cycle, which can lead to suboptimal results.

To be able to repeat the experiments, a packet trace must be captured. Considering the speed and utilisation of current uplinks of the research organisations, hundreds or even thousands of gigabytes of network data would have to be stored. Together with privacy issues, this makes sharing such data sets almost impossible.

A description of data sets is an important part of any work concerning network traffic. The properties of the traffic highly correlate with the results of any experiments. When the researchers are not able to share their data set, its description should allow other researchers to find a similar data set which would exhibit similar results. Moreover, such a description can be checked for consistency thus ensuring that the results do not unduly fluctuate over time.

The goal of this work is to provide a simple method for discerning different types of network traffic. This method allows us to compare network traffic from live networks and even packet traces and determine which traffic samples show similar properties. Finding similar traffic is highly desirable as it allows the independent evaluation of published experiments. Moreover, when applying the method continuously to a specific network, changes in the properties of the network can be not only observed, but also quantified.

978-1-5090-0223-8/16/$31.00

Flow-based characteristics [1] are used by our method as they are easier to obtain than packet-based characteristics. Moore et al. [2] list a number of discriminators that can be used for classifying traffic. We select a small subset of discriminators obtained from NetFlow v5 [3] records, that are available from most network devices. Our approach is generic and the set of discriminators can be extended for specific purposes. We show that the chosen discriminators provide enough information to distinguish between various networks.

We analyse the traffic of five campus networks within the Masaryk University. The networks have different properties as they contain a different number of servers, workstations, personal client stations, and portable devices. The networks and the methodology are described in detail so that other researchers can repeat our steps and evaluate their own network traffic.

We show that it is possible to distinguish between networks based on simple characteristics derived from flow information. Day-night and weekday patterns in the traffic are important phenomena that need to be taken into consideration when deriving characteristics. The characteristics used in this paper do not show any significant correlation which indicates that all of them contribute to the description of the network traffic.

The rest of the paper is structured as follows. Related work is surveyed in Section II. We describe the methodology for characterising network traffic in Section III. Results of applying the proposed method on several different networks are shown in Section IV. The paper is concluded in Section V.

II. Related Work

Statistical properties of Ethernet traffic were studied by Leland et al. [4]. They discovered that the traffic is statistically self-similar, which was later confirmed by several studies [5], [6]. These studies also showed that detailed characteristics, such as packet inter-arrival times, show large deviations and burstiness. Our study makes use of traffic flow properties

which are aggregated over the whole network traffic and therefore much more stable.

Fraleigh et al. [7] performed packet-level traffic measurement on the Sprint IP backbone. They showed that the measurement interval affects peaks in packet rate and bit rate. Using flow-based analysis makes it possible to normalise the data. Otherwise, it would be more difficult to compare the results from different networks. The authors also showed that different networks have different week and day-night patterns, as well as traffic types and packet size distribution. There was also a discrepancy between flows per second and bits per second for different networks. The authors argued that the network properties depend on the customer type of the link and its location.

Fomenkov et al. [8] studied network traffic behaviour on long-term data samples. The samples were captured from different networks, one to eight times a day, once per month for 60 to 120 seconds. Only packet headers were stored. The authors used packet, byte and flow rates as well as a number of source-destination IP addresses to describe long term changes in Internet traffic characteristics. The burstiness of the traffic in combination with short measurement intervals impeded the day-night and weekday patterns. Even the expected long term growth of the traffic was not observed. This advocates using longer measurement intervals in our own work. Despite the intricate nature of the samples used, the authors reported that the average packet length increases with traffic growth. They also observed differences in the composition of traffic transport protocol usage between different networks.

Benson et al. [9] described traffic characteristics of data centres. They performed a top-down analysis of ten data centres identifying applications used and their communication patterns. The authors observed flow-level and packet-level communication characteristics such as active flows per interval, the distribution of flow sizes and lengths, packet sizes and packet inter-arrival times. They noticed day-night and weekly patterns in the communication and that there is a difference between core and edge links as well as some of the data centres. Their analysis supports our belief that it should be possible to differentiate between network links based on traffic characteristics.

A characterisation of ISP traffic was performed by García-Dorado et al. [10]. The authors attempted to provide accurate, extensive, and quantitative measurements of application usage, bandwidth utilisation, and user preferences. They compared customers of different networks and access technologies for long periods of time. Unlike previous findings the daily patterns are reported to be quite invariant, although the weekdays were different from weekends in a campus network. Application traffic shares differed significantly between the individual networks. Moreover, notable changes were observed on the same link during the measurement period. This supports our assumption that measurements on live networks should be well documented and validated.

All of the above-cited works focus on describing the properties and characteristics of network traffic. The key difference to our work is that we identify characteristics that vary between the networks and show that they can be used to differentiate and describe the networks in a uniform manner.

TABLE I: Measured networks.

Network	Packets	Bytes	Flows
Faculty of Informatics	227.1 G	236.4 T	3.6 G
Institute of Computer Science	107.3 G	106.2 T	0.7 G
University Campus Bohunice	449.8 G	473.9 T	4.1 G
Virtual Switching Segment	1 119.2 G	1 158.3 T	11.7 G
Masaryk University	1 366.6 G	1 427,7 T	20.1 G

III. Methodology

To be able to describe network traffic and find key properties that allow us to discern different networks, we collected data from several different parts of our campus network. Understanding the purpose of each network is imperative for the correct interpretation of the data collected, therefore we describe the networks used for collecting the data in detail. This section describes the processes and tools used to collect the data and the collected data itself.

A. A Description of the Networks

We continually collect data from individual networks in our campus. We chose five networks that are very different by their nature and utilisation. For each network we selected two months of data, January and March. The networks are expected to exhibit different characteristics between the two months, since there is an exam period at Masaryk University in January and March is a standard midterm month. A summary of networks and the collected data is presented in Table I.

The **Faculty of Informatics** (FI) network represents a network of a university faculty. It connects staff offices, computer labs, and faculty servers. The faculty has its own Eduroam infrastructure which can be observed in the network. The network also contains servers with the information system for the entire Masaryk University.

The **Institute of Computer Science** (ICS) has its own network connecting staff offices and a small server infrastructure to support office computers such as remote storage or update servers. Unlike the Faculty of Informatics, it does not connect to computer labs.

University Campus Bohunice (UCB) is a large campus building with hundreds of offices, computer labs and a large library. The faculty of Sports Studies and Faculty of Medicine are situated in the campus building. The Central European Institute of Technology (CEITEC) is also located on these premises and it generates a large volume of data due to intensive scientific computing.

The **Virtual Switching Segment** (VSS) contains a server segment and also the Eduroam wireless network concentrator. Every Eduroam connection at the university goes through this network. Servers supporting the Masaryk University IT infrastructure, such as the Economic and Administrative Information System of Masaryk University or digital libraries, are also located in this network.

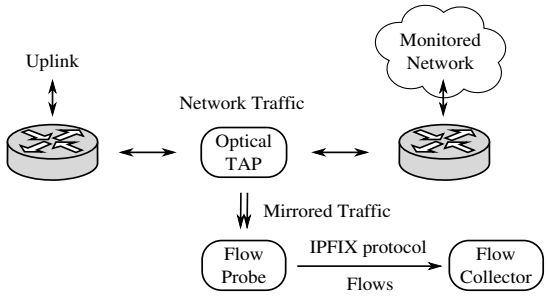

Fig. 1: Flow Monitoring Architecture Schema.

Masaryk University's (MU) network is measured as a whole at the uplink to ISP. The communication of every university subnet is observed at this point with the exception of internal communication.

B. Data Collection

We have a flow probe [1] monitoring each of the networks in our campus. The data from the probes is sent to a flow collector where it is stored for further processing. The flow monitoring architecture is depicted in Figure 1.

Flow exporters are configured to create flows with an active timeout of 300 seconds and an inactive timeout of 60 seconds. The flows are terminated using only these timeouts, TCP FIN and RST flags are not used for the flow termination. The exporters use a standard flow key consisting of IP addresses, L4 protocol and its ports. It is important that exporters have the same configuration, since different configurations may result in a different number of exported flow records. As we will show, the number of flow records is an important characteristic of the network.

The flows from all probes are collected on a single collector. No sampling is used in the flow creation or collection process; therefore we have a complete flow trace of the measured network in a given period of time.

C. Data Preprocessing

This subsection describes how the raw flow data was processed. First, we generate statistics for short intervals so that the amount of data is reduced. These statistics do not carry any personal information, therefore they can be shared with other researchers without privacy concerns. Then we process these statistics, look at daily and weekly patterns, their averages and deviations. We identify characteristics which differ between the networks, describe them and use them to differentiate between the networks. To be able to compare the traffic characteristics of different networks, we need to normalise the volumetric properties of the traffic.

Individual flow records are not needed to describe network traffic. Instead, the records must be aggregated and statistical indicators must be derived from the raw data. Since most flow-based frameworks process data in 5 minute intervals [11], we chose the same interval for generating aggregated statistics.

We generate five types of statistics from the original flow records: *Basic volume*, *Advanced volume*, *Layer 3*, *Layer 4* and *Layer 7* statistics. These statistics are listed in Table II. There are 41 collected statistics in total when the bytes, packets and flows statistics are counted separately.

TABLE II: Collected statistics.

Type	Statistic
Basic Volume	Bytes per second Packets per second Flows per second
Advanced volume	Flow size in packets, bytes Flow connection length
Layer 3	Source host count Destination host count IPv4 / IPv6 bytes, packets, flows
Layer 4	IPv4 / IPv6 TCP bytes, packets, flows IPv4 / IPv6 UDP bytes, packets, flows IPv4 / IPv6 ICMP bytes, packets, flows IPv4 / IPv6 other protocol bytes, packets, flows
Layer 7	Top 10 ports by packets, bytes, flows

The Advanced volume statistics represent the detailed volumetric characteristics of the network. Therefore, we compute mean, variance, standard deviation, median and all percentiles from 0.55 to 0.95 with increment of 0.05. Percentiles lower than the median are superfluous since half of the flows in every network are composed from single packet flows, which amount to zero length flows with small packet sizes (typically 40 bytes). We also computed covariance and correlation for each pair of these statistics. The port statistics are stored as a table of top 10 ports sorted by either packets, bytes or flows. We also the computed overall top port statistic for our analysis.

IV. Results

We carefully considered all the acquired data and divided it into seven areas of network behaviour that are studied in more detail. This section describes our findings and decisions taken when analysing the collected data.

A. Day-night Pattern

The day-night pattern is a key element of change in every network with human users. It is evident that we cannot compare traffic captured during the day with traffic captured at the night. The reason is that the traffic pattern during the night generally contains much less user generated traffic and the properties are very different, as we present further in our analysis. The day-night pattern is best observed in flows per second since it expresses the number of connections. Using bytes or packets per second distorts the pattern since there are heavier flows during the day than at the night [12]. It is possible to use the number of hosts to show the day-night pattern but the results are very strongly correlated, therefore we decided to use flows per second.

The day-night behaviour of the network discloses a lot of information about the purpose of the network. Figure 2 shows flows per second in University Campus Bohunice network in January. The data from all the days is superimposed in the graph so that the day-night pattern is clearly visible. The red colour shows the weekends and the black is for workdays. It

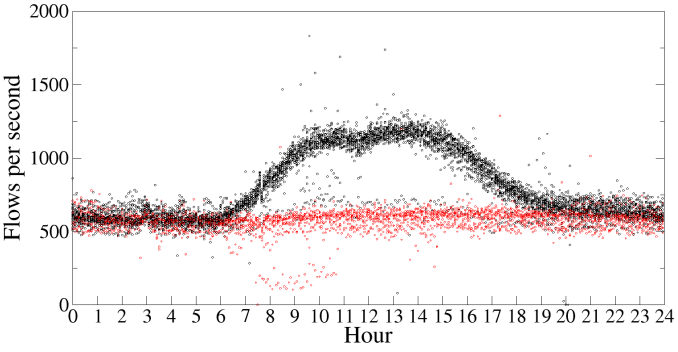

Fig. 2: Flows per second in the UCB network.

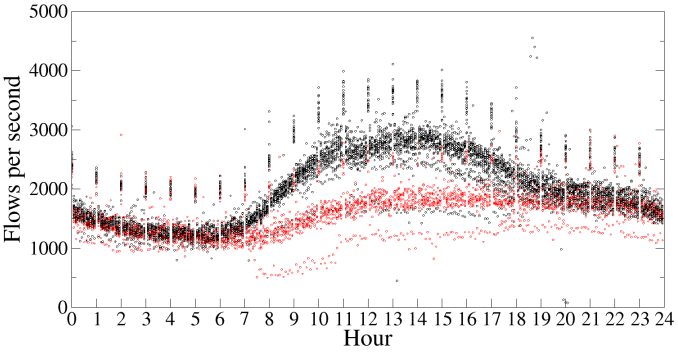

Fig. 3: Flows per second in the VSS network.

is possible to even see a lunch break in the data at half past eleven. The weekends are flat and do not exhibit the day-night patterns since there are not many users at the weekend.

Figure 3 shows the Virtual Switching Segment network which has quite different properties. Since there are Eduroam users and servers that are accessed from outside the university, we can see that there is a day-night pattern visible even at the weekends. Moreover, we can see periodical communication which is caused by automated server backups and updates.

TABLE III: Day to night and workday to weekend ratio according to flows per second.

Network	Day	Night	Day-night Ratio		Week Ratio	
			January	March	January	March
UCB	13	5	1.96	2.08	1.85	1.87
ICS	13	1	2.25	1.70	2.24	1.52
MU	14	5	3.08	3.62	1.45	1.95
FI	10	5	2.04	2.16	1.47	1.64
VSS	14	5	2.16	2.73	1.43	1.77

For each network the most and the least busy hour of the day according to flows per second was determined (as shown in Table III). Given the property of the network, the *day-night ratio* will be, from now on, computed as a ratio between the average of the property during the busiest hour in the day and the least busy hour at night. Using day-night ratios specific to each network allows us to compare the behaviour of each network regardless of the absolute size of the network.

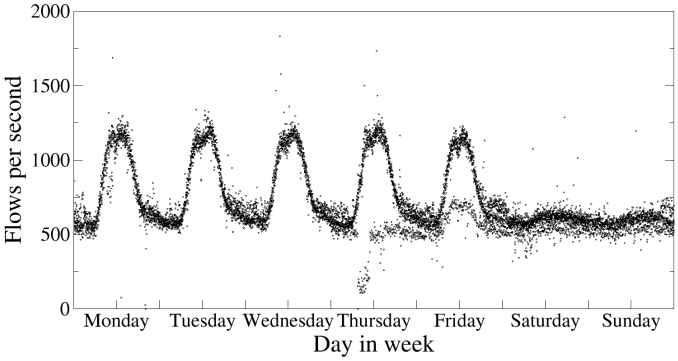

Fig. 4: Flows per second in the UCB network.

Table III also describes the day-night pattern for all networks. We processed the data from January and March 2015 separately to show the difference between the exam period and the teaching period. The day-night ratio expresses the ratio of flow count as described earlier. The largest difference between the day-night ratios in January and March is in the traffic from the entire university and the VSS traffic. This is a seasonal change caused by students leaving the campus and returning only for their exams. The traffic from faculty networks (FI, VSS) increases only slightly in the teaching period, indicating that the faculty networks are not highly utilised by students. The ICS network shows the opposite trend, which is likely caused by project deadlines and people working harder after the Christmas vacation to meet them.

We also computed an interval which has the maximal ratio of average flows per second to the rest of the day. We have observed two types of network. The activity in networks FI, MU and VSS start at circa 7:00 and ends at 23:00, while the activity in networks UCB and ICS start at 8:00 and ends at circa 17:00. This can be attributed to the fact that that UCB and ICS networks are used while people work and the rest of the networks are used while people are awake.

The day-night pattern provides important information about the networks and their usage. We can use figures from Table III to describe the day-night pattern of these networks.

B. *Weekday Pattern*

Another significant pattern that can be observed in network traffic is the weekday pattern. Figures 4 and 5 show the weekday pattern in January for UCB and VSS. The graphs are constructed by superimposing the data from individual weeks. There are significant differences between the networks. The weekend traffic at UCB is almost flat but VSS shows behaviour that suggests that there is a significant number of users communicating over the network even at weekends.

The difference between work days and weekends can be expressed as a ratio of average flows per second measured during the busiest hours on workdays and at weekends. Table III shows the results for January and March. This characteristic shows a correlation with the day-night pattern. We presume that if the network is accessed by other users, such as students,

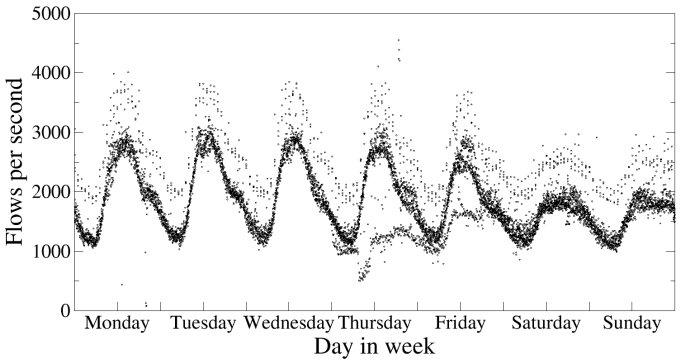

Fig. 5: Flows per second in the VSS network.

TABLE IV: Average length of flow and packets per flow.

Network	Length of The Flow		Packets per Flow	
	Week avg.	Day-night rat.	Week avg.	Day-night rat.
UCB	10.07 s	2.18	112.86	1.28
ICS	10.34 s	1.68	205.36	0.66
MU	13.09 s	2.13	71.79	0.81
FI	5.40 s	1.77	67.60	0.70
VSS	7.14 s	2.04	106.25	0.94

TABLE V: Source and destination hosts day-night ratios.

Network	Source Hosts		Destination Hosts	
	January	March	January	March
UCB	2.66	2.51	1.91	1.90
ICS	2.76	3.11	1.99	2.35
MU	2.49	2.66	1.65	2.08
FI	1.24	1.35	1.22	1.31
VSS	2.93	4.93	2.48	4.12

outside working hours, it will probably be accessed also on weekends by the same users. However, we can still employ this statistic to differentiate between networks, since the correlation is not strong and the statistic still adds new information.

C. Flow Characteristics

Using basic network characteristics such as packets per flow, bytes per packet, flow length or number of hosts is difficult due to day-night and weekly patterns. The variance of these variables renders averaging the values useless. Therefore, we have taken another approach. Average values are computed for whole weeks and the averages are used as base values. The variance of week averages computed over the months is on average less than 10 %, which is low enough to use these values to characterise the networks.

Bytes per packet statistic had little variance over the measured networks and it gives almost no information to discern the networks. We found that flow length and packet per flow statistics differ more significantly, as shown in Table IV, and can be used to discern the networks. The day-night ratios show that the networks with similar averages might have different day-night behaviour for the same statistics. Therefore, the ratios add new information to the week averages.

The average number of hosts is a statistic directly related to the size of the network and cannot be used to describe the type of the network without normalisation. Therefore, we use only the day-night host ratios, as shown in Table V, which are independent of the absolute number of hosts. Source and destination directions are determined by the direction of the flow. There is a significant difference between host ratios in January and March for the VSS network. The UCB is quite indifferent in comparison, as the ratios remain almost the same. This shows that the statistic can be used to find differences between the networks.

We have shown that comparing basic characteristics on networks must be done in sufficiently large intervals that are the same for all networks. In our case the average values taken over an entire week proved to be stable enough to be used as a network characteristic. The ratio between characteristics measured during the day and night period also provides relevant information about the behaviour of the network and can also be used as an important network characteristic.

TABLE VI: IPv6 utilisation.

Network	Flows	Packets	Bytes
UCB	0.02 %	0.01 %	0.00 %
ICS	5.58 %	12.35 %	13.57 %
MU	12.98 %	1.94 %	1.41 %
FI	3.22 %	2.04 %	2.10 %
VSS	4.66 %	0.23 %	0.17 %

TABLE VII: TCP and UDP share in day and night in flows.

Network	Day		Night	
	Tcp	Udp	Tcp	Udp
UCB	38.52 %	59.76 %	19.44 %	77.66 %
ICS	41.26 %	56.88 %	28.20 %	68.84 %
MU	55.55 %	43.03 %	42.99 %	54.25 %
FI	49.96 %	49.07 %	25.15 %	73.49 %
VSS	30.67 %	67.76 %	19.30 %	78.00 %

D. IPv6 Utilisation

The utilisation of the IPv6 protocol in the network is a Layer 3 characteristic that indicates the technological readiness of the network. The adoption of IPv6 is slow and IPv4 traffic is still dominant. We use this fact to differentiate between networks with different levels of IPv6 readiness. The amount of IPv6 traffic is affected by IPv6 ready servers in the networks and IPv6 services accessed by users outside the network. Table VI shows the ratio of IPv6 to total traffic in flows, packets and bytes. The packets and bytes ratios are strongly correlated, however, they differ from the flows statistic. The UCB, ICS and MU networks clearly stand out in these characteristics.

E. Protocol Share

We investigated the differences in Layer 4 protocol usage in the data sets. TCP and UDP are the dominating protocols in all networks and the shares of ICMP and other protocols are so negligible that they can hardly be used to describe the networks. The shares of TCP and UDP differ significantly between day and night time. Table VII shows the average TCP and UDP shares during the day and night in flows.

TABLE VIII: The most common ports by flows in January.

Port / Network		UCB	ICS	MU	FI	VSS
DNS	**53**	9.7 %	30.2 %	21.5 %	12.2 %	42.7 %
HTTP(S)	**80**	9.2 %	7.4 %	20.2 %	16.8 %	5.9 %
	443	6.5 %	8.3 %	14.7 %	20.1 %	4.0 %
Mail	**25**	–	–	1.0 %	0.6 %	–
	993	–	1.7 %	–	0.6 %	–
Samba	**445**	1.0 %	–	–	–	0.7 %
SSH	**22**	–	–	1.4 %	0.3 %	–
NTP	**123**	–	0.9 %	7.1 %	43.9 %	–
SNMP	**161**	52.8 %	11.9 %	–	–	23.5 %
Telnet	**23**	1.0 %	1.3 %	1.6 %	0.4 %	–

The increase in TCP traffic during the day is caused by users since most of the services use the TCP protocol. We also investigated protocol shares in packets and bytes and found that they match the overall packet to flow and byte per packet ratio. The results show that the TCP to UDP ratio differs between the networks and can be used for their description.

F. Most frequent ports

Traffic analysis by port numbers provides evidence of application usage in the network. Although the port numbers are not considered to be accurate enough for application identification [13], most of the traffic adheres to well-known port numbers, which is accurate enough for our purpose. We computed the top 10 port statistics by flows and packets in January and March for all networks. The flows statistic is more stable and more suitable for analysing the behaviour of the network. We selected ten well-known ports that were most often observed in the statistic. Table VIII shows the percentage of flows that belong to these ports. We can observe that the networks have very different usage of the ports which makes port usage an important characteristic of the network. Note that we did not study day-night variance of port usage, but we expect that its fluctuation may be used to refine the results.

V. Conclusions

We have presented an analysis of network traffic measured at five different campus networks at Masaryk University. Our goal was to show that the properties of the networks can be extracted and quantified and that we can use the results to differentiate and describe the networks. We have made several observations during our analysis which affect the derivation of network characteristics.

- Flow-based statistics are more stable than byte or packet based statistics. Therefore it is more practical to use flow statistics to track long-term changes in network behaviour.
- Day-night and weekday patterns must be taken into consideration when computing network characteristics.
- Long enough samples of the traffic must be available.
- Using ratios between day and night helps to compensate for the day-night pattern and the size of the network. However, there is still a slight correlation between network size and the day-night ratios.
- To describe a network used to perform an experiment, the description of behaviour using network characteristics should be complemented with absolute volumetric information.

Our measurements have confirmed that it is possible to differentiate between networks based on the observed characteristics. Moreover, the campus networks showed quite stable characteristics over time even though the measurements were taken during the winter exam and spring teaching periods. Furthermore, we tested the correlation of the described characteristics and found that they are only very weakly correlated. This shows that each of the characteristics is valuable and carries information about the network. However, the presented characteristics are not by any means complete. We analysed a subset of possible characteristics derived from flow records, which can be easily measured on a network. Thus we believe that more work is required to identify other useful characteristics and utilise them to describe the behaviour of the networks.

References

[1] R. Hofstede, P. Čeleda, B. Trammell, I. Drago, R. Sadre, A. Sperotto, and A. Pras, "Flow Monitoring Explained: From Packet Capture to Data Analysis with NetFlow and IPFIX," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 2037–2064, 2014.

[2] A. Moore, M. Crogan, and D. Zuev, "Discriminators for use in flow-based classification," Queen Mary, University of London, Tech. Rep., 2005. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.7450&rep=rep1&type=pdf

[3] Cisco, "NetFlow Export Datagram Format," 2015, [cited 2015-07-27]. [Online]. Available: http://www.cisco.com/en/US/docs/net_mgmt/netflow_collection_engine/3.6/user/guide/format.html

[4] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the Self-similar Nature of Ethernet Traffic," in *Conference Proceedings on Communications Architectures, Protocols and Applications*, ser. SIGCOMM '93. New York, NY, USA: ACM, 1993, pp. 183–193.

[5] M. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: evidence and possible causes," *Networking, IEEE/ACM Transactions on*, vol. 5, no. 6, pp. 835–846, Dec 1997.

[6] V. Paxson and S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 226–244, Jun. 1995.

[7] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and S. Diot, "Packet-level traffic measurements from the Sprint IP backbone," *Network, IEEE*, vol. 17, no. 6, pp. 6–16, Nov 2003.

[8] M. Fomenkov, K. Keys, D. Moore, and K. Claffy, "Longitudinal Study of Internet Traffic in 1998-2003," in *Proceedings of the Winter International Synposium on Information and Communication Technologies*, ser. WISICT '04. Trinity College Dublin, 2004, pp. 1–6.

[9] T. Benson, A. Akella, and D. A. Maltz, "Network Traffic Characteristics of Data Centers in the Wild," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '10. New York, NY, USA: ACM, 2010, pp. 267–280.

[10] J. Garcia-Dorado, A. Finamore, M. Mellia, M. Meo, and M. Munafo, "Characterization of ISP Traffic: Trends, User Habits, and Access Technology Impact," *Network and Service Management, IEEE Transactions on*, vol. 9, no. 2, pp. 142–155, June 2012.

[11] P. Haag, "NfSen," 2015. [Online]. Available: http://nfsen.sourceforge.net/

[12] L. Quan and J. Heidemann, "On the characteristics and reasons of long-lived internet flows," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '10. New York, NY, USA: ACM, 2010, pp. 444–450.

[13] A. W. Moore and K. Papagiannaki, "Toward the Accurate Identification of Network Applications," in *Proceedings of the 6th International Conference on Passive and Active Network Measurement*, ser. PAM'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 41–54.