NETWORK TRAFFIC CHARACTERISATION USING FLOW-BASED STATISTICS

Wednesday 27th April, 2016

Petr Velan

Jana Medková, Tomáš Jirsík,
Pavel Čeleda

CSIRT-MU

Introduction

- We need to be able to describe the network traffic.
  - The researchers have to be able to describe the network traffic for their experiments.
  - Most methods heavily depend on the properties of the observed traffic.
  - The network traffic's properties can change during the research cycle.
- However:
  - We cannot store packet traces.
- The goal of this work is to provide a simple method for discerning different types of network traffic.

Collected Data

- NetFlow data were collected from five campus networks within the Masaryk University.
- Two whole months of data were collected during January and March 2015.
- 41 L2, L3 and L4 statistics were computed from collected data and evaluated based on their importance in describing the traffic in network.

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

CSIRT-MU

A Description of the Networks

University Campus Bohunice (UKB)

- Offices, computer labs and a large library
- The Central European Institute of Technology located here generates a large volume of data due to intensive scientific computing

Virtual Switching Segment (VSS)

- Every Eduroam connection at the university goes through this network
- Servers supporting the Masaryk University IT infrastructure

A Description of the Networks

Faculty of Informatics (FI)

- Staff offices, computer labs, and faculty servers
- Faculty Eduroam infrastructure
- Servers with the information system for the entire university

Institute of Computer Science (ICS)

- Staff offices
- Server infrastructure to support office computers such as remote storage or update servers

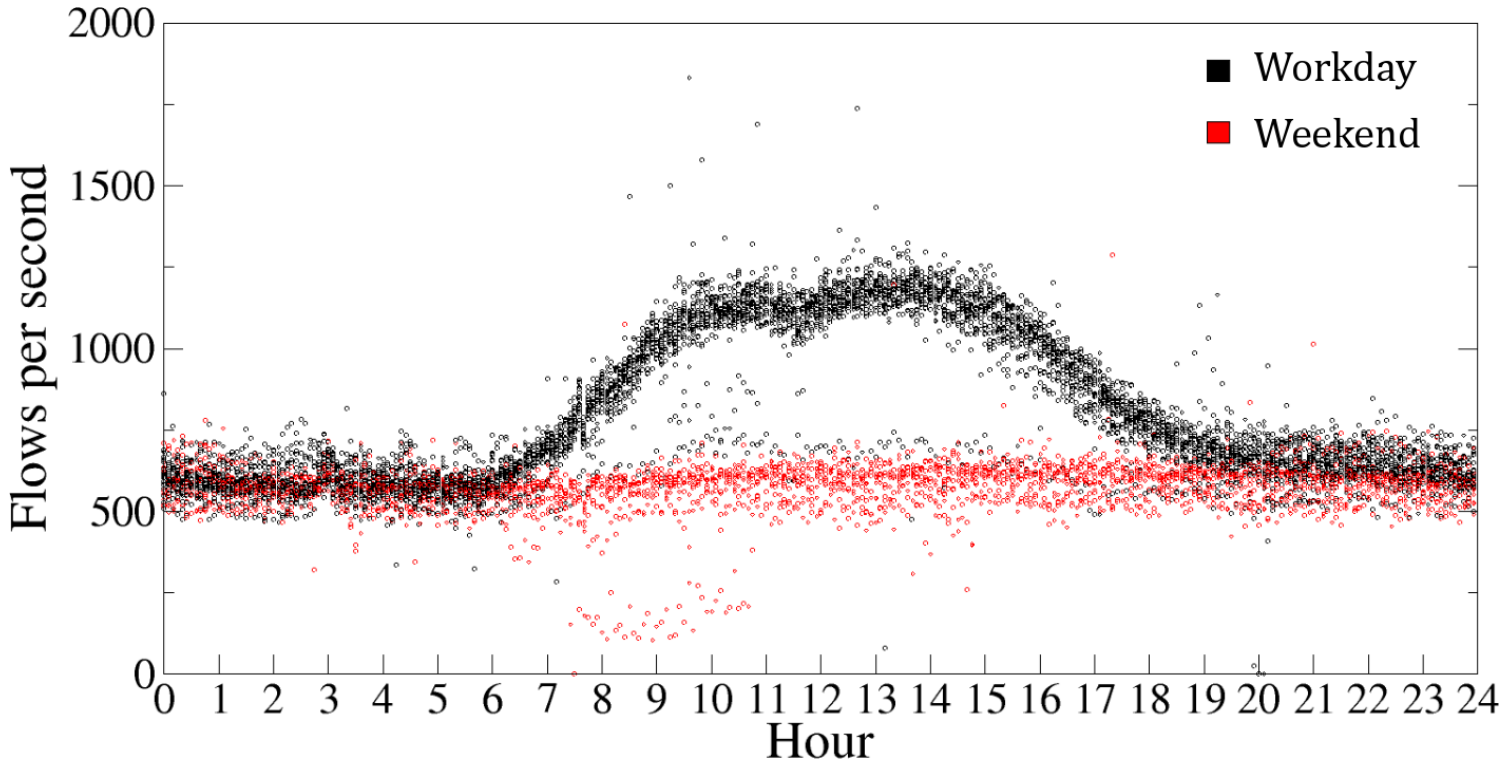
Masaryk University (MU)

- Measured at the uplink to ISP
- The communication of every subnet is observed except for internal communication

Day Night Pattern

University Campus Bohunice

The day is from 8:00 till 17:00.

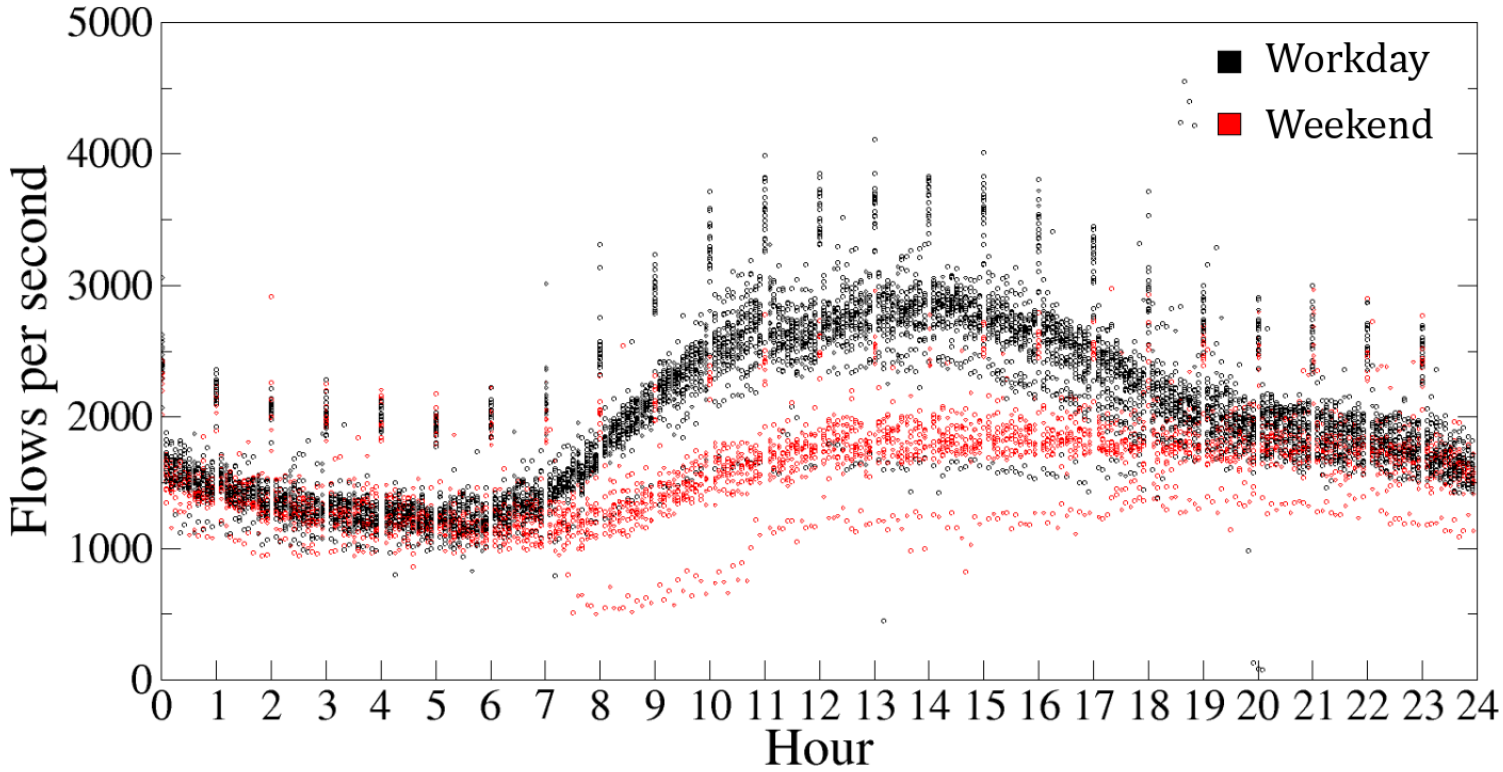


CSIRT-MU

Day Night Pattern

Virtual Switching Segment

The day is from 7:00 till 23:00.



CSIRT-MU

Day Night Pattern - day-night ratio

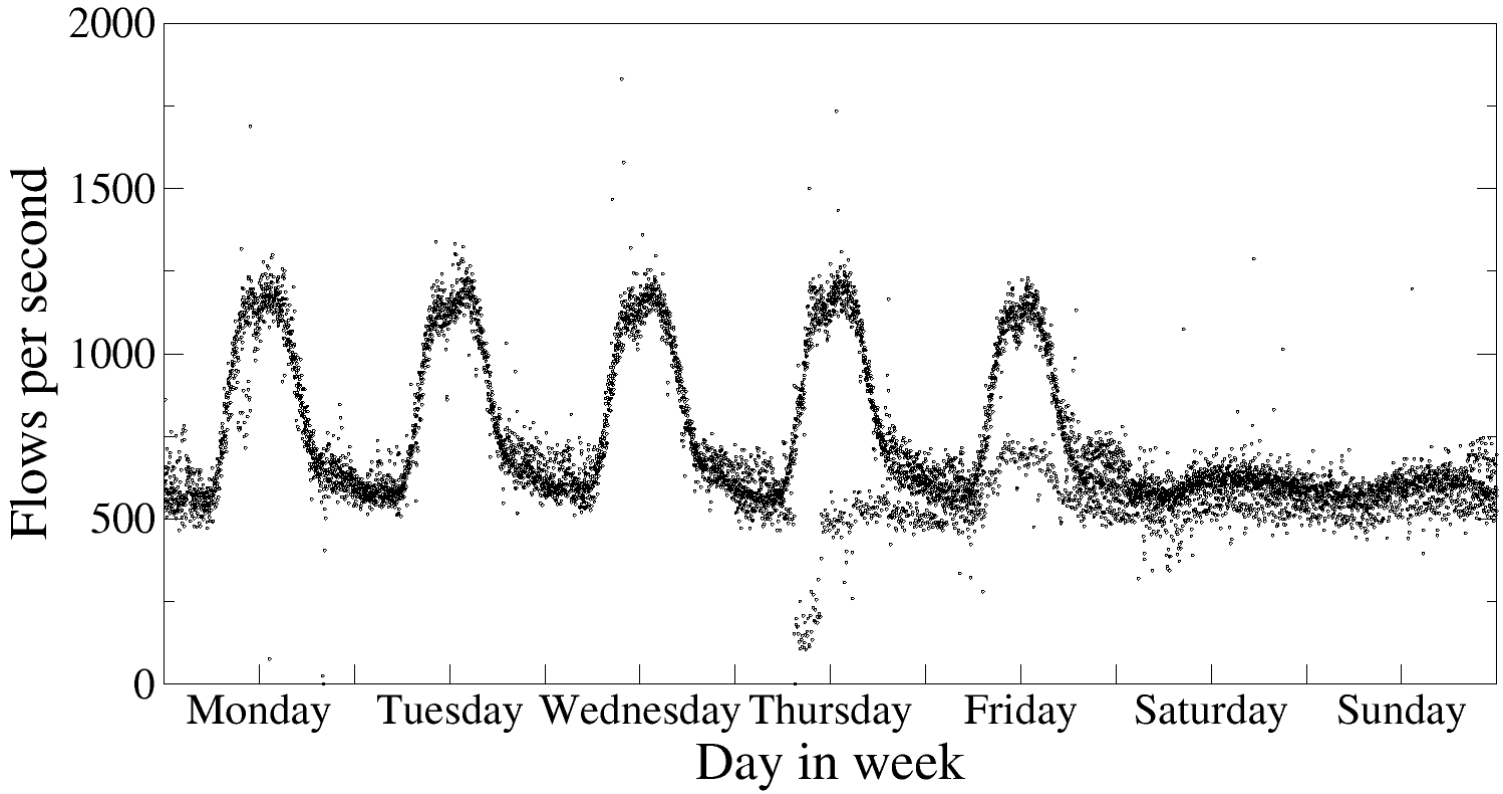
- The *day-night ratio* was computed as a ratio between the average of the property during the busiest hour in the day and the least busy hour at night on workday.

$$dr = \frac{\#\text{flows in the most busy hour of the day}}{\#\text{flows in the least busy hour of the night}}$$

Network	Day	Night	Day-night Ratio
UCB	13	5	1.96
ICS	13	1	2.25
MU	14	5	3.08
FI	10	5	2.04
VSS	14	5	2.16

Weekday Pattern

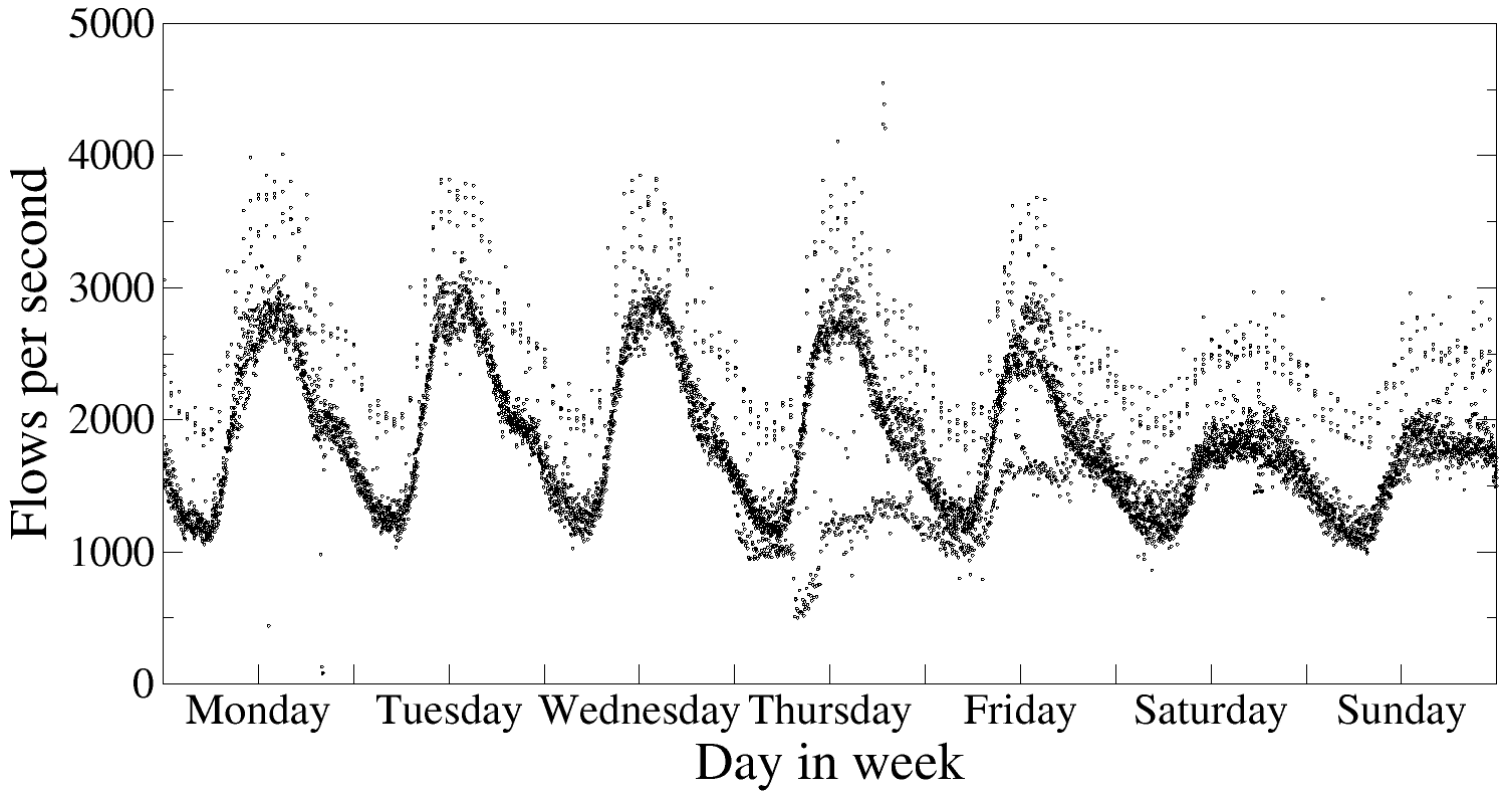
University Campus Bohunice



CSIRT-MU

Weekday Pattern

Virtual Switching Segment



CSIRT-MU

Weekday Pattern

- It affects only traffic at daytime. The traffic during night is the same at weekends and on workdays.
- The *week ratio* was computed as a ratio between the average flows per second during the busiest hour on workdays and at weekends.

$$wr = \frac{\#\text{flows in the most busy hour of the day on workdays}}{\#\text{flows in the most busy hour of the day on weekends}}$$

Network	Week Ratio
UCB	1.85
ICS	2.24
MU	1.45
FI	1.47
VSS	1.43

Flow Characteristics

- Because of day-night pattern and workday-weekend pattern, averages have to be taken for whole weeks.
- Bytes per packet statistic has small variance over the networks and is not usable to discern the networks.

Network	Length of The Flow		Packets per Flow	
	Week avg.	Day-night rt.	Week avg.	Day-night rt.
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

IPv6 Utilisation

- The utilisation of the IPv6 protocol indicates the technological readiness of the network.

Network	Flows	Packets
UCB	0.02 %	0.01 %
ICS	5.58 %	12.35 %
MU	12.98 %	1.94 %
FI	3.22 %	2.04 %
VSS	4.66 %	0.23 %

Protocol Share

- TCP and UDP are the dominating protocols in all networks.
- The ratio between UDP and TCP at day and at night differ between networks and can be used to characterise the network.

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 %

Most frequent ports

- The networks have very different usage of the ports, so port usage can be used to characterise the network.

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 %	–

Summary

At least the following information should be given when describing network traffic

- Total number of bytes, packets and flows per week
- The day and night interval in the network
- Day-night ratio and week ration
- Average week length of flow and packets per flow
- IPv6 protocol usage in flows and packets
- UDP and TCP shares in day and at night
- Top 10 ports and the ratio of traffic on these ports

THANK YOU FOR YOUR ATTENTION!

Petr Velan

{velan, jana.medkova, jirsik, celeda}@mail.muni.cz