

COVID-19 Pandemic and Internet Traffic in Poland: Evidence from Selected Regional Networks

Michał P. Karpowicz

NASK National Research Institute, Warsaw, Poland

<https://doi.org/10.26636/jtit.2021.154721>

Abstract—The COVID-19 pandemic has forced governments all over the world to impose lockdowns keeping citizens at home in order to limit the virus spread rate. The paper compares weekly traffic samples captured in the selected nodes of the network managed by NASK – National Research Institute during the pre-lockdown period, i.e. between January 27 and February 3, 2020, with those captured between March 30 and April 6, 2020, i.e. after the lockdown was announced. The presented results show changes in network traffic observed during the periods of time in question and illustrate the evolution in the popularity of top network services.

Keywords—COVID-19, Internet traffic.

1. Introduction

The COVID-19 pandemic has reshaped both the economy and our daily routines. Be it education, shopping, sports, traveling, work, health care, entertainment or social interactions – the pandemic has left its mark on all areas of our lives.

To limit the virus spread rate, governments all over the world imposed lockdowns keeping citizens at home. This has resulted in numerous everyday activities being performed online. Consequently, the demand for network services and resources surged. This paper studies network traffic variations observed in the network of the NASK National Research Institute at the beginning of the first lockdown introduced in Poland.

1.1. Related Work

The impact of the COVID-19 pandemic on Internet traffic is a subject of extensive studies. Indeed, reports published so far reveal numerous interesting and similar effects that are correlated with the introduction of lockdowns.

As reported in [1], a sharp surge in traffic was observed worldwide in late March and early April 2020, i.e. after the introduction of the first lockdowns in Europe. The lockdowns led to a surge in the popularity of streaming services, causing a visible change in demand for network resources. On the one hand, that impact was similar to surges caused by planned worldwide events, such as New Year's Eve

celebrations or concerts, or events like natural disasters or flash crowds. On the other hand, numerous differences have been spotted as well. Specific symptoms of ongoing changes in users' habits could be observed. Evolution of e-commerce is visible as well, market models change, remote education is gaining momentum, and interest shifts emerge.

The lockdowns have also left their mark on packet transmission latency. This phenomenon affected many latency-sensitive applications, including online games, video calls, VoIP, and IP geolocation. Inferior service quality was experienced mainly during the evenings. In other words, latency increased due to recreational activities, rather than due to remote working or distance learning [2].

In [3], an academic campus is taken as an example to identify changes in traffic patterns. The study shows that incoming traffic decreased drastically. In contrast, outgoing traffic doubled in order to support online learning and working from home, particularly with the use of streaming platforms, VPNs, and remote desktop services. For a related study of remote learning strategies during the COVID-19 pandemic, see also [4].

Finally, [5] offers a comprehensive study of the lockdown effect observed from the point of view of a Central European ISP, three major European Internet exchange points (IXPs), and one metropolitan educational network in Spain. After the lockdowns were imposed in the second half of March, much more traffic was generated in the mornings and during late evening hours. The study provides details of the changes visible in the transport and application layers. The collected data prove that the distribution of traffic source and destination ports has changed due to lockdowns. Those changes, clearly visible in traffic samples, may be attributed directly to remote working, education, VPN-based communications, and video conferencing. Also, interestingly enough, workday patterns became similar to weekend patterns. That phenomenon needs a further explanation that takes into account the characteristics of local economies and social structures, as the sections below suggest.

This paper presents traffic patterns observed in the NASK – National Research Institute network during the first Polish lockdown.

1.2. Dataset and Ethical Considerations

The dataset used in this study consists of sFlow (RFC 3176) traffic samples collected between January 27 and February 3, 2020, and between March 30 and April 6, 2020, as a part of routine network analysis procedure performed by NASK. It describes the activity of a stable group of (~1,000) commercial customers, government and public web services, and selected academic network users. In the period considered, there were no significant changes to the network infrastructure the study is concerned with. Likewise, the network user base did not change either.

The collected data contain randomly sampled packet headers and do not reveal any packet payload information. Also, special care was taken to anonymize IP addresses in compliance with Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal on the free movement of such data. Therefore, the anonymization process was applied permanently and irreversibly and removed any potential links between IP addresses and both legal entities and natural persons to whom the addresses may relate.

The following procedure was used while working with the dataset. First, the traffic samples collected were pre-processed with the use of a cryptography-based prefix-preserving anonymization algorithm (CryptoPan) to replace the IP addresses observed with new addresses [6]. Next, we replaced the last bits of each new IP address with zeros. As a result, the dataset contained only information regarding the flow of packets between anonymized networks. The activity of any individual IP address is not visible in the dataset. Finally, as presented in the following sections, aggregated statistical data and time series were obtained.

2. Network Traffic Data Study

This section illustrates and comments on the correlation between the changes in network traffic characteristics and the decision to impose the lockdown in Poland in the second half of March 2020.

Firstly, it compares weekly traffic patterns during the pre-lockdown period with weekly traffic patterns during the lockdown. The aggregated traffic is discussed together with its top components, including TCP/UDP and HTTP/HTTPS traffic.

Secondly, it compares the activity of leading source and destination ports before and after the decision to impose the lockdown. This comparison offers an overview of changes attributed to remote working, education, VPN-based communications, and video conferencing. Observations regarding network security are given as well.

To calculate relative changes in traffic patterns, the total number of bytes transferred during the periods studied was compared. The following formula was applied:

$$\Delta = \frac{Y - X}{X} \cdot 100\% , \tag{1}$$

where X denotes pre-lockdown and Y lockdown measurements. The statistics were generated with an improved version of of nfdump software, introducing the correct handling of port 0 traffic.

2.1. Traffic Rate Shifts

Figure 1 illustrates the network traffic rate, measured in bits per second (bps) and packets per second (pps), during the week preceding the lockdown and the week after the lockdown was announced. For the sake of clarity, the presented time series were smoothed with a low-pass zero-phase second order Butterworth filter. As we can see, the overall traffic increased by more than 40%. That number stems from the comparison of the total number of bytes and packets transferred during the periods under analysis.

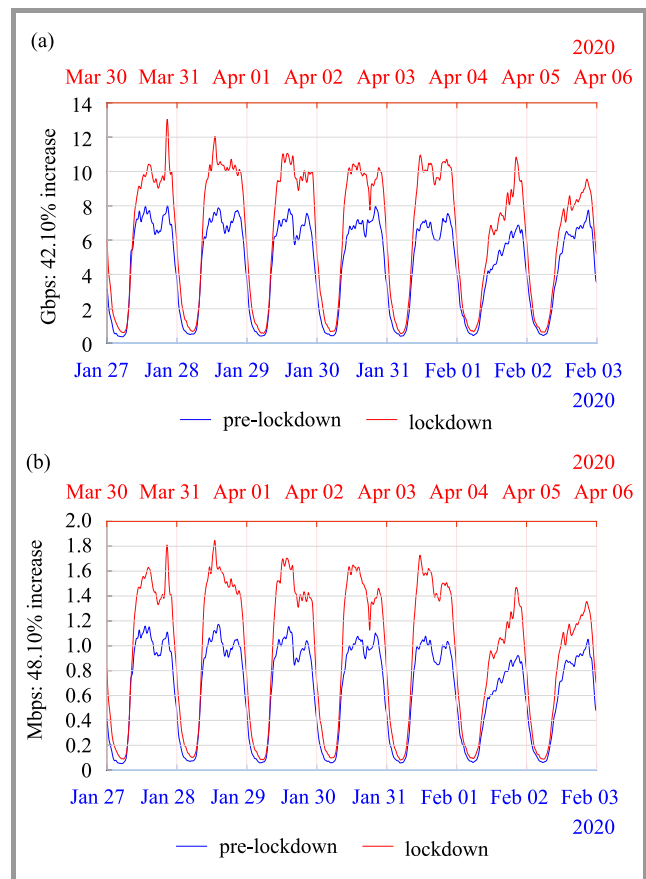


Fig. 1. Aggregated network traffic before and after the lockdown: bits per second (a), packets per second (b). (see the digital edition for color images)

On workdays, traffic increased during the first part of the day, before lunchtime. After lunch, traffic increased again, reaching its peak late in the evening, during the second part of the day. Traffic reached its maximum before lunch. On

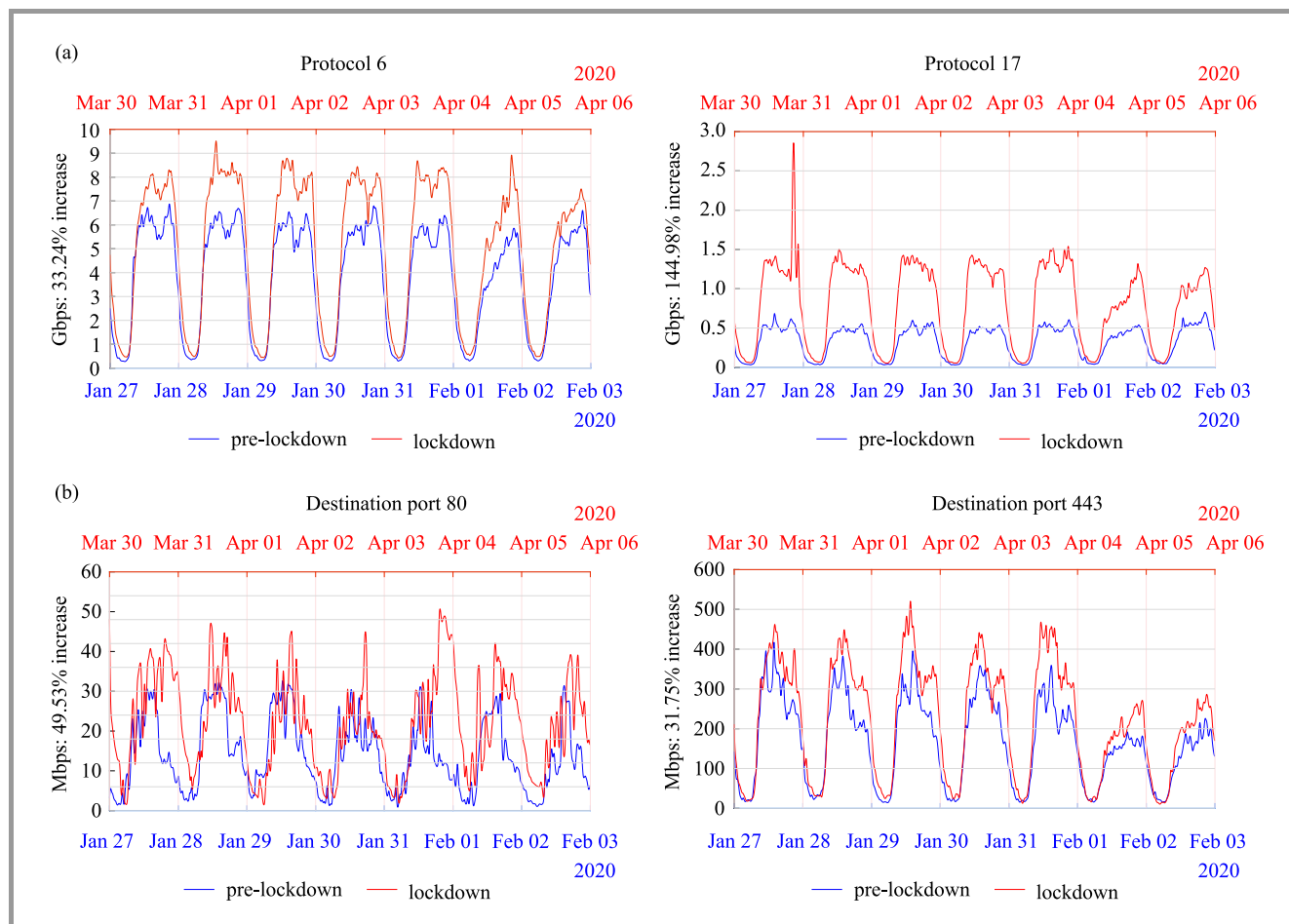


Fig. 2. Traffic shift details: TCP/UDP [Gbps] (a), HTTP/HTTPS [Mbps] (b).

weekends, in contrast, traffic increased gradually to reach its peak in the evening. An increase may also be observed in minimal traffic rates, both on workdays and during weekends.

Figure 2 reveals more details of the traffic shifts. Top figures show TCP (proto 6) and UDP (proto 17) traffic during the period concerned. In the first case, traffic increased by 33%. However, in the second case, that increase reached almost 145%. This may be explained by a rapid increase in UDP-based traffic generated by data streaming and VPN services. HTTP/HTTPS traffic is presented in the middle. The rate of HTTP traffic rose almost by 50%, whereas HTTPS traffic rose by nearly 32%. In the case of HTTP traffic, the peaks moved towards late evening hours, a symptom suggesting user behavior changes.

2.2. Application Activity Shifts

This subsection compares the activity of top source and destination ports before and after the lockdown. A comparison of the top fifteen source ports is presented in Table 1. Top destination ports, in turn, are compared in Table 2. The last column in each table presents activity shifts calculated based on the number of bytes transferred from or to a given port within the period under consideration.

Both before the lockdown and the just after its introduction, ports 443 and 80 remained the most active ones. The dominant role of such application layer protocols as HTTP (port 80) and HTTPS (port 443) may be explained easily. These are the application-layer protocols delivering web page content and transferring data over the Internet. Since the lockdown moved our lives to the online environment, the increase in traffic at ports transmitting HTML documents, images, or videos comes as no surprise. As far as the HTTPS protocol is concerned, source traffic increased by 11.6%, whereas destination traffic increased by 24.1%. In the HTTP protocol, the increase reached 28.8% and 33.8%, respectively.

The remaining part of the list of top ports illustrates critical changes in application activity patterns. The most significant growth in traffic volume was recorded by IPsec, OpenVPN, and new communication services, such as MS Teams or Google Hangouts.

Activation of virtual private networks (VPNs) was the first and straightforward consequence of the lockdown. This explains the increase in IPsec (port 4500) and OpenVPN (port 1194) traffic. In the first instance, that increase reached over 190%, while in the second – over 96% in terms of source traffic. Similarly, since streaming and communica-

Table 1
Top source ports, ordered by bytes

Pre-lockdown				Lockdown				
Source port	Application	TB	Gpackets	Source port	Application	TB	Gpackets	Δ [%]
443	HTTPS	826.5	634.7	443	HTTPS	922.3	729.1	11.6
80	HTTP	326.9	224.0	80	HTTP	421.0	288.3	28.8
53	DNS	10.4	15.8	4500	IPSec	14.5	22.8	190.0
993	IMAPS	8.1	9.4	53	DNS	11.9	16.8	14.4
0		7.0	8.0	0		11.2	14.5	60.0
37777	Video	5.7	4.2	1194	OpenVPN	10.2	12.5	96.2
1194	OpenVPN	5.2	7.0	37777	Video	8.6	6.2	50.9
4500	IPSec	5.0	7.5	5544		6.8	5.2	88.9
6908	Bittorrent	4.5	3.4	993	IMAPS	6.7	7.3	-17.3
6907	Bittorrent	4.0	6.3	8080	HTTP	5.1	3.9	
8080	HTTP	3.7	2.7	8801	Backup	4.9	10.2	
5544		3.6	2.9	8000	Streaming	4.5	3.6	45.2
8000	Streaming	3.1	2.5	6908	Bittorrent	4.5	3.4	0.0
22	SSH	2.6	2.4	10443	SSL/dogtag	3.2	4.4	
995	POP	2.4	1.9	34765		2.8	2.2	

Table 2
Top destination ports, ordered by bytes

Pre-lockdown				Lockdown				
Destination port	Application	TB	Gpackets	Dst port	Application	TB	Gpackets	Δ [%]
443	HTTPS	77.5	323.5	443	HTTPS	96.2	396.8	24.1
80	HTTP	13.6	90.7	80	HTTP	18.2	128.1	33.8
6666	IRC	8.0	5.7	4500	IPSec	13.8	19.7	181.6
0		7.0	8.0	0		11.2	14.5	60.0
1194	Open VPN	5.8	7.5	6666	IRC	9.6	6.9	20.0
4500	IPSec	4.9	7.5	1194	Open VPN	7.9	11.4	36.2
6180		3.7	2.4	38188		4.6	3.4	
1935	RTMP	2.3	4.7	6180		2.9	1.9	-21.6
25	SMTP	2.2	2.3	6901	MSN Messenger	2.8	2.2	33.3
6901	MSN Messenger	2.1	1.7	6902	MSN Messenger	2.8	2.2	
33451	Webex	2.0	1.5	25	SMTP	2.8	2.5	
54536		1.8	2.9	19305	Hangouts	2.1	3.9	
48508		1.8	1.4	1935	RTMP	2.0	3.4	-13.0
48611		1.6	2.6	20	FTP	2.0	1.4	33.3
20	FTP	1.5	1.1	33001	Aspera	1.8	1.2	

tions services supporting online meetings and file sharing became crucial for the remote working model, they significantly boosted traffic rates as well. That is the case for ports 37777, 8000, 6901, 19305, or 20. Additionally, that observation may also explain the decrease in port 993 traffic supporting mailing over SSL. New communication services, such as MS Teams or Google Hangouts, seem to have taken its place.

Interestingly, the data collected reveal also an increase in traffic that is correlated with ports often used by malicious software. This includes port 5544 and port 0. In fact,

during the period from March to May, numerous attacks were registered in the network under observation. Figures 1–2 show a traffic spike caused by a DDoS attack. A campaign of DDoS attacks may also be seen in Fig. 3, illustrating the panel of the FLDX DDoS protection system developed and used in the networks managed by NASK. Detected distributed denial of service (DDoS) attacks can be seen as traffic spikes (each of them was successfully attenuated).

Statistics for port 0 traffic need a more detailed explanation. While being illegal in general cases, port 0 is well known to

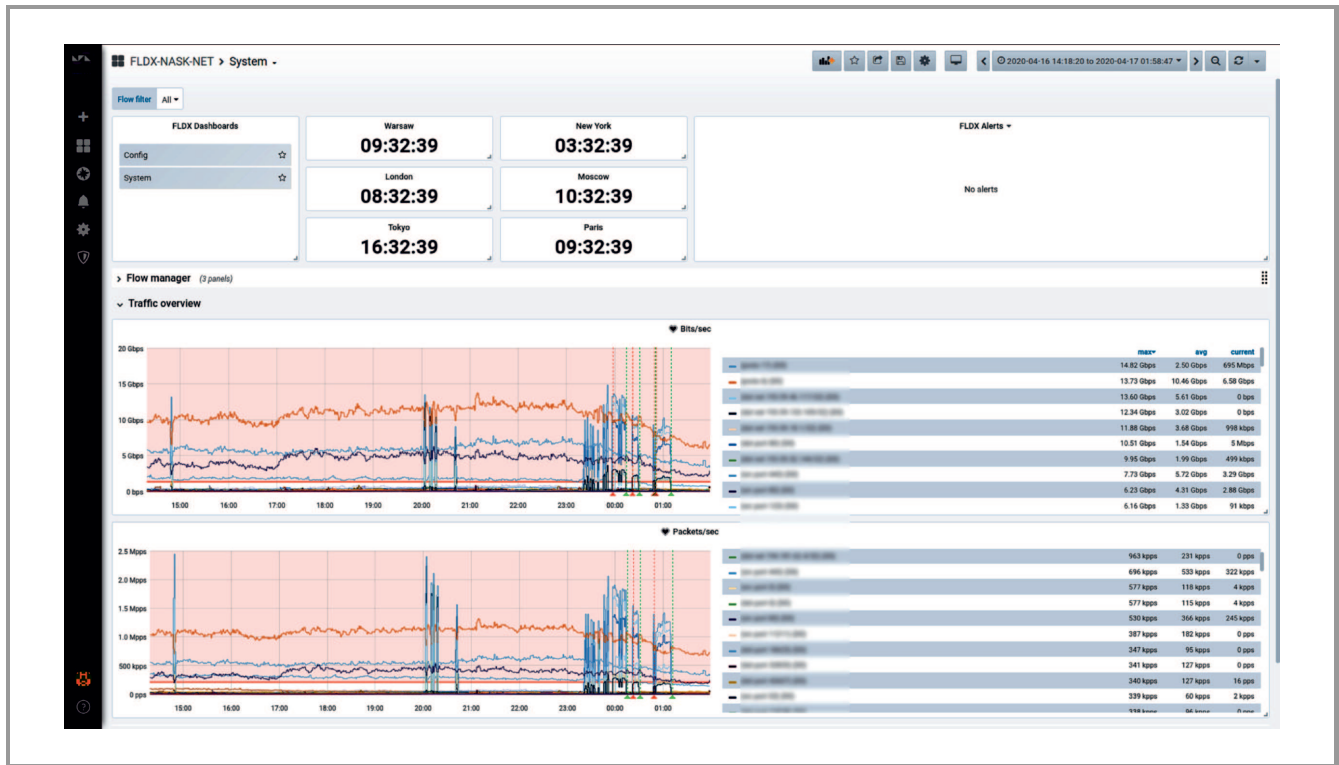


Fig. 3. Example of a series of DDoS attacks (seen as traffic spikes) detected and attenuated by the FLDX DDoS protection system in the second half of April 2020.

have a large traffic share because it aggregates fragmented, malicious, scanning, and wild packets. It is often seen in many network attacks, including DDoS attacks originating from large botnets. However, it is associated with legitimate traffic as well. To be more precise, packets generated by protocols that do not use ports happen to be marked by network devices and packet analyzing software precisely as port 0 traffic. Therefore, much care should be taken when analyzing that particular case. Statistics presented in Tables 1, 2 and 3 compensate for that effect, exposing only traffic with valid source and destination ports. In the

period studied, the overall traffic share of port 0 increased approximately by 60%. For more details, see also [7].

3. Summary

The COVID-19 pandemic has introduced significant changes to our lives. Many of these changes may stay with us for longer or may turn out to be permanent. Remote work and remote education, online medical appointments, online shopping, movie premieres at home, dancing practice in your living room, all these activities have evolved rapidly due to lockdowns. It is so because they help us face the challenges associated with isolation. At the same time, many forms of activity have disappeared as a result of lockdowns.

Changes in network traffic reflect changes in our habits. The volume of network traffic has significantly increased, and we tend to be online for longer. Communication platforms, streaming services and VPNs supporting remote work and remote education have gained in importance. A careful observation of web applications and protocols helps us to understand the ongoing changes and pinpoint the potential threats.

Acknowledgements

I express my gratitude to the FLDX system development team, especially to Arkadiusz Piórkowski and Janusz

Table 3

Service port 0 traffic profile: destination ports (top) and protocols ordered by bytes (bottom)

Pre-lockdown		Lockdown	
Destination port	Bytes	Destination port	Bytes
0	7T	0	11.2T
80	60.2M	80	49.7M
443	4.5M	443	29.4M
53	1.9M	6680	2.8M
12001	1.9M	12812	1.5M
Protocol	Bytes	Protocol	Bytes
UDP	6.9T	UDP	11.1T
TCP	61.3T	TCP	110.1G
IPv6	142M	IPv6	109M

Janiszewski, for their continuous and innovative work and support in analyzing data. I am also deeply indebted to Marek Dawidiuk for securing the traffic samples during the difficult period of remote work. Finally, my sincere appreciation goes to Urszula Brochwicz for developing legal recommendations on data anonymization.


References

- [1] T. Boettger, G. Ibrahim, and B. Vallis, "How the Internet reacted to COVID-19: A perspective from Facebook's Edge Network", in *Proc. of the ACM Internet Measurement Conf.*, Virtual Event, USA, 2020, pp. 34–41 (DOI: 10.1145/3419394.3423621).
- [2] M. Candela, V. Luconi, and A. Vecchio, "Impact of the COVID-19 pandemic on the Internet latency: A large-scale study", *Computer Networks*, vol. 182, 2020 (DOI: 10.1016/j.comnet.2020.107495).
- [3] T. Favale, F. Soro, M. Trevisan, I. Drago, and M. Mellia, "Campus traffic and e-Learning during COVID-19 pandemic", *Computer Networks*, vol. 176, 2020 (DOI: 10.1016/j.comnet.2020.107290).
- [4] T. Gonzalez *et al.*, "Influence of COVID-19 confinement on students' performance in higher education", *PloS one*, vol. 15, no. 10, 2020 (DOI: 10.1371/journal.pone.0239490).
- [5] A. Feldmann *et al.*, "Implications of the COVID-19 Pandemic on the Internet Traffic", in *Broadband Coverage in Germany; 15th ITG-Symposium*, pp. 1–5, Online Conf.: VDE, 2021 (ISBN: 9783800754748).
- [6] J. Fan, J. Xu, M. H. Ammar, and S. B. Moon, "Prefix-preserving IP address anonymization: measurement-based security evaluation and a new cryptography-based scheme", *Computer Networks*, vol. 46, no. 2, pp. 253–272, 2004 (DOI: 10.1016/j.comnet.2004.03.033).
- [7] A. Maghsoudlou, O. Gasser, and A. Feldmann, "Zeroing in on port 0 traffic in the wild", in *Proc. of the 2021 Passive and Active Measurement Conference (PAM '21)*, Online Conf., 2021, pp. 547–563 (DOI: 10.1007/978-3-030-72582-2_32).



Michał P. Karpowicz, Ph.D. (2010), D.Sc. (2020), is a Professor with the NASK National Research Institute, head of the IT Systems Engineering Department and Assistant Professor with the Institute of Control and Computation Engineering, Warsaw University of Technology. His research interests include control theory,

signal processing, and linear algebra. He is an author of more than 30 articles, co-author of one book, designer of two network control systems supporting cybersecurity operations in Poland's nation wide networks.

 <https://orcid.org/0000-0003-1431-3078>

E-mail: michal.karpowicz@nask.pl
 NASK National Research Institute
 ul. Kolska 12
 01-045 Warszawa
 Poland

