ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

UDC 004.738.5.056.5 **doi: 10.32620/reks.2022.4.05**

Olexander ZADEREYKO[1], Olena TROFYMENKO[1], Yuliia PROKOP[2], Nataliia LOGINOVA[1], Anastasiia DYKA[1], Serhii KUKHARENKO[1]

[1] *National University "Odessa Law Academy", Odessa, Ukraine*
[2] *Czech Technical University in Prague, Prague, Czech Republic*

RESEARCH OF POTENTIAL DATA LEAKS IN INFORMATION AND COMMUNICATION SYSTEMS

This article discusses the problem of ensuring the protection of user data in information systems. It is shown that classic information systems are represented by stationary and mobile communication devices focused on data exchange with digital space. The fundamental principles of user data exchange in the digital space are considered. It has been established that leading technology IT corporations collect data from user communication devices. It is shown that the organization of data collection is carried out by redirecting the DNS traffic of the communication device to the DNS servers of IT corporations, followed by its encryption using the DoH protocol. This makes it impossible for authorized services and departments of national states to control the users' DNS traffic and ensures the monopoly position of IT corporations in the global digital market for collecting and analyzing user data. It is shown that the collection of user data is carried out with the aim of further monetization and influencing decisions made by users. DNS traffic of devices for communication with the digital space of the Internet is fixed. An audit of the recorded DNS traffic was performed, and as a result, specialized Internet resources were identified to be responsible for collecting and processing user data. It has been proved that the identified specialized Internet resources belong to IT corporations. Methods of identification of communication devices in digital space were considered. It is shown that the identification of communication devices is based on the collection of a unique set of data from each communication device. Based on each unique data set, a digital fingerprint of the communication device is formed, which is used for its further identification in the digital space. These approaches allow organizing protection against user data collection in information systems. Software and hardware implementations for protection against data collection from communication devices are proposed. It has been experimentally established that the combined use of the proposed software and hardware models provides the most effective protection against data collection from communication devices and does not affect the functionality of information systems.

Keywords: *data leaks; digital space; DNS queries; DNS servers; communication device; data collection.*

Introduction

Due to objective processes of the formation of a single global information space, various types of users' social, financial, and economic activities are shifted into the digital space.

The main danger of this process is that it contributes to an intensive increase in the volume of personalized user data circulating in the digital space. This circumstance naturally determines the collection and subsequent analysis by the leading subjects of the digital space – IT companies. As a result, this leads to the formation of an unofficial "market" for users' data. The state structures of most countries have a very indirect relationship to the regulation of this market, despite the measures taken in the form of adopting laws on the protection of personal data in their territories [1].

The energetic efforts of the world's leading IT companies such as Google, Microsoft, Amazon, Cloudflare, Facebook, etc. have led to the section and partial monopolization of the digital market for user data. This fact determines the global process of monopolization of the market for collecting, storing, and processing user data in the digital environment [2]. The analysis of user data is based on the organization of the processes of their collection, storage, and processing. It is the basis for the development of (personalized or group) control impact on the participants of the digital space (Fig. 1).

The ultimate goal of the expansion of IT companies into the digital market for user data is to profit from the virtually unlimited opportunities they have formed for the secret study, control, and management of users of the digital space.

In other words, IT companies have integrated a complex means of influence (information weapons) into the digital environment. Its qualitative result is the modification of an information system's properties, which usually means an object that interacts with information – a person with a communication device on which the corresponding software is installed.

The main effect of information weapons is that their use inevitably leads to a change in the consciousness of participants in the digital space, to the reprogramming (unification) of their system of assessing events. This leads to the complete controllability of their behavior [3-4].

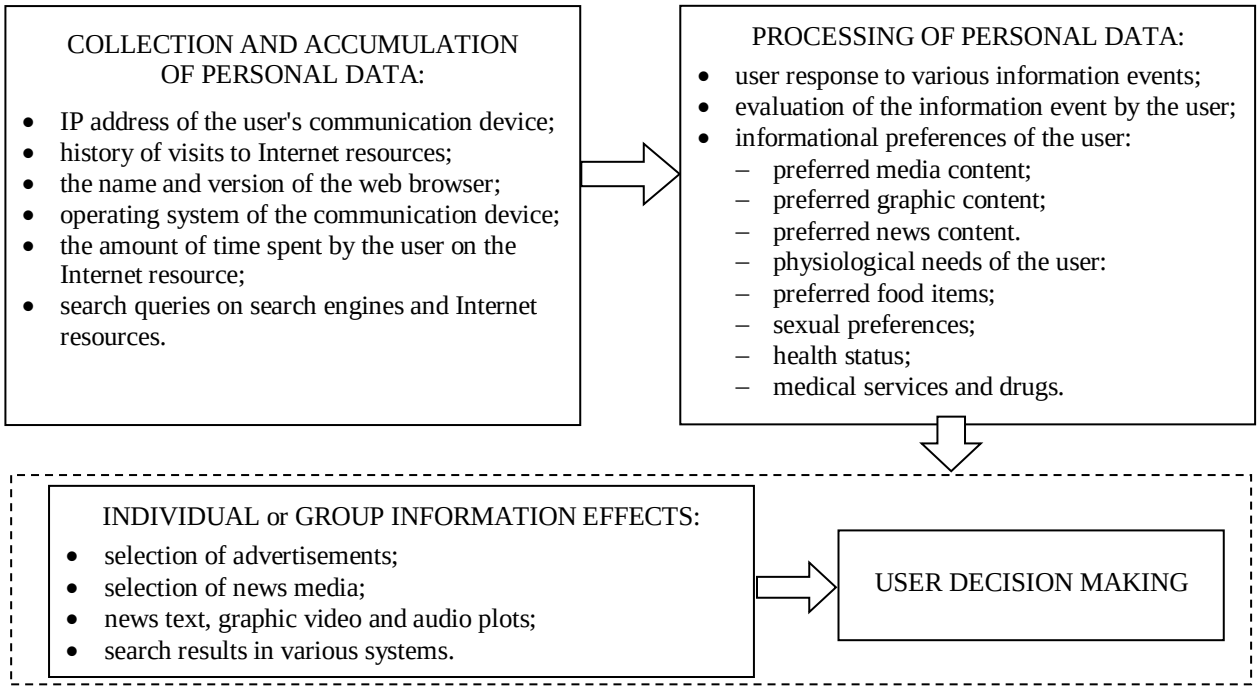

Fig. 1. Management of participants in the digital space

IT companies systematically create and implement information and communication, television, video, photo, and radio broadcasting platforms, integrated simultaneously into human society and the digital space [5]. This multiplies the effect of controlling and manipulative influence on the mass consciousness of the digital audience [6, 7].

This leads to an urgent need for an in-depth analysis of the principles of user data exchange in the digital space. The results of such an analysis will help protect the participants of the digital space from the manipulative influence of IT companies on the users' decisions.

The goal of the paper is to audit the traffic of the user's classic information system and develop a set of measures to protect against leaks of user data in the digital space.

To achieve this goal, it is necessary to solve the following tasks:

1. To analyze the principles of data exchange between a typical user information system and the digital space of the Internet.

2. To analyze the methods (principles) of identification of communication devices in the digital space.

3. To fix the outgoing and incoming connections of the information system with the digital space.

4. To analyze the connections of the user's information system with the digital space.

5. To determine the criteria for classifying the connections of the information system with the digital space.

6. To develop a set of measures to organize protection against leaks of user data in information systems.

1. Data exchange principle between the information system and the digital space

The classical definition of a computer information system (IS) defines it as a system for storing, searching, and processing information, and the corresponding organizational resources (human, technical, financial, etc.) that provide and disseminate information [8].

Regarding the problem, the IS can be defined as the environment, which consists of computing systems (computers and smartphones), technological and software tools (operating systems (OS) and applied software (software)), computer networks, and their users.

In the applied sense, the IS will be considered as a hardware and software complex designed to meet the requirements of end users in the search, retrieval, and storage of information.

As a rule, the main task of such IS is to meet the information needs of the user within the selected subject area. The most widespread are desktop and mobile ISs, in which all hardware and software components for communication are located on one device (computer, smartphone), which will henceforth be called communication devices (CD) (Fig. 2). The modern digital space provides free cross-border data exchange between the user's CD and hosts for various purposes [9, 10]. The physical connection of the user's CD to the digital space begins with sending DNS queries to DNS servers.

To understand the essence of the process of data exchange between the CD and the digital space, let us consider the classical scheme of interaction of a typical DNS client with the domain namespace (Fig. 3).

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

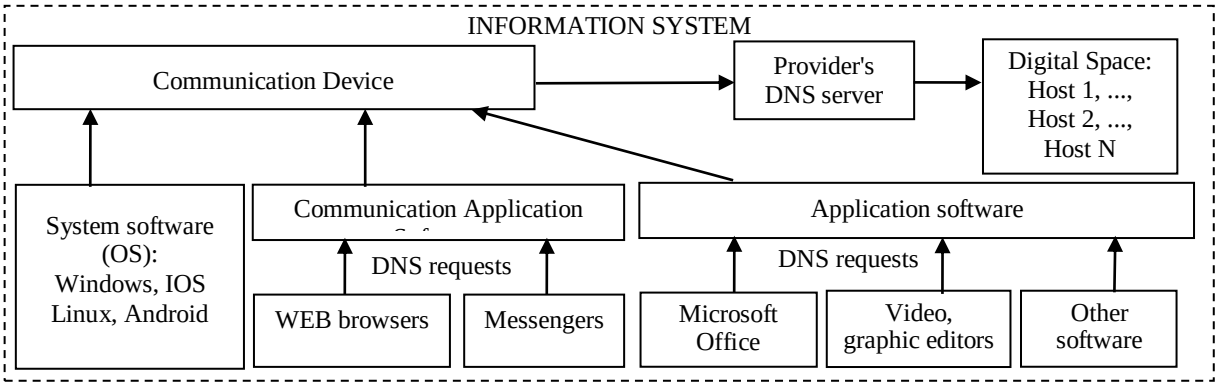

Fig. 2. Interaction of IS modules with digital space

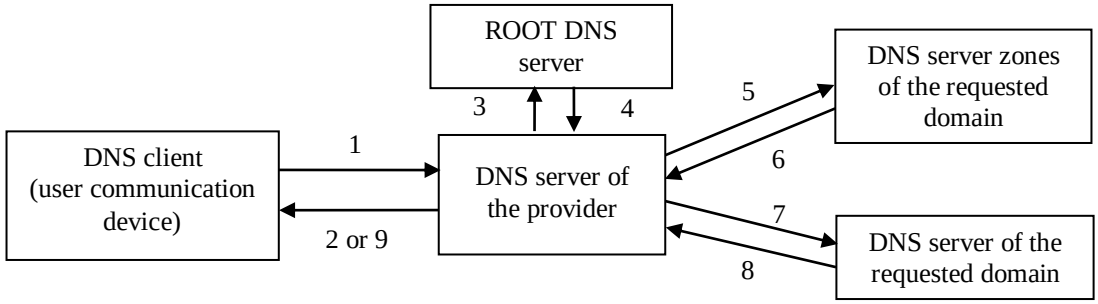

Fig. 3. The scheme of interaction of a typical DNS client with the domain namespace of the Internet

DNS queries in the domain space are processed in a strictly defined order:

1 – the DNS client makes a request of the "protocol://domain" type to the provider's DNS server to obtain the IP address, hosting, where the Internet resource corresponding to the requested domain is located;

2 – the provider's DNS server searches its log for the requested domain. If it is absent, the client receives a DNS – IP address corresponding to the completed DNS request;

3 – the provider's DNS server contacts the root DNS server with a request for the IP address of the DNS server responsible for the domain zone corresponding to the DNS request of the DNS client;

4 – the root DNS server returns to the provider's DNS server the IP address of the DNS server responsible for the domain zone corresponding to the DNS request of the DNS client;

5 – the provider's DNS server addresses the DNS server responsible for the domain zone corresponding to the request of the DNS client with the request for the IP address of the DNS server responsible for hosting the domain of the Internet resource requested by the DNS client;

6 – DNS server responsible for the domain zone corresponding to the DNS request of the DNS client returns the IP address of the DNS server responsible for hosting the domain requested by the DNS client;

7 – the provider's DNS server addresses the DNS server responsible for hosting the domain requested by the DNS client with a request for the hosting IP address corresponding to the domain requested by the DNS client;

8 – the DNS server responsible for hosting the domain requested by the DNS client returns the IP address corresponding to the requested domain to the ISP's DNS server;

9 – the provider's DNS server returns to the DNS client the IP address corresponding to the requested domain.

From the CD data exchange scheme (see Fig. 3), it can be concluded that all DNS client requests are recorded in the provider's DNS server logs. Their analysis gives comprehensive information to the provider and the government agencies that control it about the actions of the user or his CD in the digital space. This allows users' DNS queries to be treated as user data. IT companies and special services for various purposes in all states are showing an increasing interest in the collection, storage, and analysis [11].

An equally important aspect is the fact that the initial openness of the DNS facilitates monitoring, interception, and data spoofing. DNS queries of users are transmitted in an open unencrypted form via UDP and TCP (port 53), which greatly simplifies their fixation or interception not only by providers of various levels but also by third parties. Therefore, to protect users' DNS traffic from leaks and interception, the DNS-over-TLS (DoT) and DNS-over-HTTPS (DoH) protocols were developed and implemented. Their implementation does not require any changes to the existing domain name

system but adds a secure way of working for DNS for its clients [12]. A logical continuation of the use of the DoH protocol in the digital space was the introduction of it by the majority of web browser developers into their software products. In addition, web browser developers have provided their users with the ability to directly connect (bypassing providers) to the public DNS servers of leading IT companies (Table 1), which also support the DoH protocol [13].

Table 1

Web browsers with DoH support and public DNS servers

Web browsers	DoH support	Built-in support for public DNS servers		
		Google	Cloudflare	Cisco
Mozilla	+	+	+	+
Opera	+	+	+	+
Chrome	+	+	+	+
Edge	+	+	+	+
Safari	+	+	+	+
Yandex	+	+	+	+

Thus, each CD user using any of the web browsers (Table 1) can independently, without any help, choose their preferred DNS server and activate the use of the DoH protocol for the secure transmission of their data.

The implementation of these measures in the digital space has already actually led to:

– impossibility of interception and control of DNS traffic by authorized services and departments of national states;

– monopolization of collection and analysis of data on DNS traffic of users in favor of leading IT companies in the global digital data market.

Another important fact is that each host (Internet resources) collects a wider range of data about the user's CD and his actions in the digital space. The owners overwhelmingly connect the hosts to the data collection and analysis services Google Analytics, Yandex Metrika, Liveinternet, etc. This, in turn, by definition leads to an absolute monopolization of the process of collecting data on users by IT companies (Fig. 4).

Analyzing the structure of interaction between the CD and the digital space (Fig. 4), we can conclude that the ultimate beneficiaries of data collection are IT companies.

Processing and analysis of the received data are carried out in automatic mode and allow you to unambiguously establish [14-16]:

– the IP address from which the CD "enters" the Internet;

– OS under which the CD operates;

– installed applied software;

– CD screen resolution;

– user's search queries;

– consumption sphere;

– relationships with financial institutions;

– time spent on Internet services;

– the subject of viewed content;

– demographic / gender indicators of the user audience;

– the age group to which the user can be assigned.

2. Principles of identification of communication devices

In the digital space, the formation of a digital fingerprint of the user's CD is inextricably linked with its unambiguous identification. CD identification is performed both by active and passive methods [17, 18].

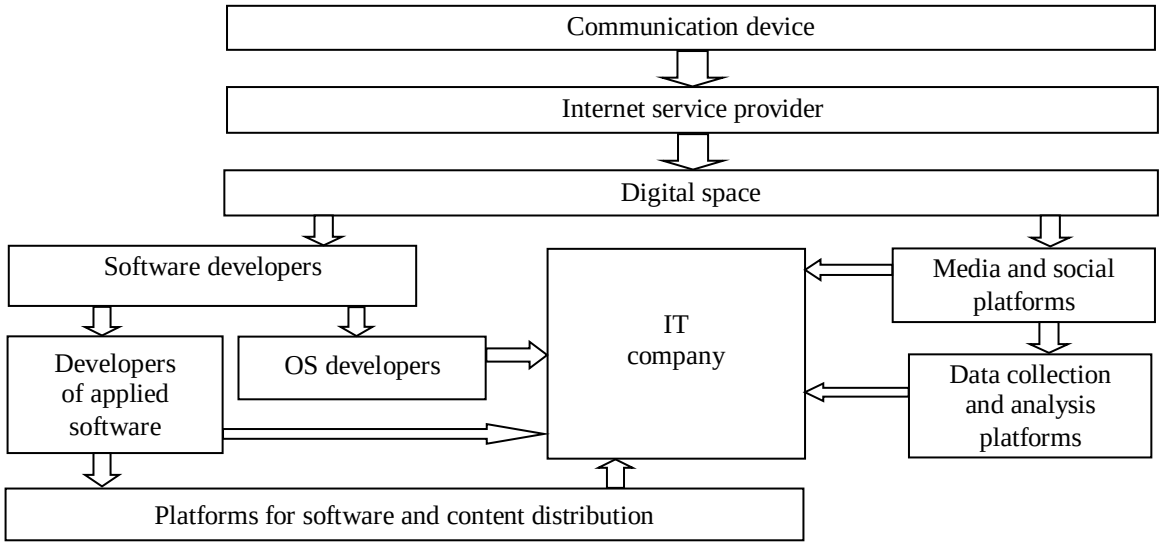

Fig. 4. Scheme of data collection from the CD

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Passive identification is based on the principles of collecting information without sending specialized requests to the CD. Mandatory data sent by each CD contains information about the OS, installed software, and information about its network adapters (MAC address, etc.).

Active identification is based on the principles of active networking, which are implemented by running JavaScript codes from remote hosts. The information received should include the size of the application window or web browser, the installed set of fonts, the set of plug-ins in the web browser, the OS language, and the time zone, as well as information about the hardware configuration of the CD.

At its core, in the process of the CD identification, the following is fixed: MAC addresses of network adapters, OS version, and installed software, especially with the help of which communication with the digital space is carried out: web-browsers, instant messengers, etc. for the digital fingerprint of the CD (Fig. 5).

The formation of a digital fingerprint of a CD is based on collections of data sets collected about it (see Fig. 5), which are combined and converted into a hash function – a unique ID that is stored in external databases, usually owned by IT companies. The received data are impersonal and unique for each management company [18].

The presence of a digital camera in the CD, and, as a consequence, the digital images obtained with their help, placed in the digital space, greatly simplify the process of identifying the CD. This is because each digital camera has its unique characteristics: the level of digital noise and the frequency-contrast characteristic of the optical system and photosensor [19].

3. Fixation of outgoing connections of the information system with the digital space

As numerous studies have shown [12-20], when connecting to the digital space, the CD, its system, and applied software establish outgoing connections not only with the hosts of its developers but also with non-functional hosts. These hosts have no direct relation to them. Software connections to developer hosts can be explained by collecting data on its health and operating modes, checking licensing terms of use, collecting telemetric data, or searching for updated software versions. But connections with non-functional hosts look more than suspicious against this background [20-22]. Moreover, as the practice of blocking such connections shows, the performance of the software is not impaired [23]. So, for example, in mobile OS, most applications in the overwhelming majority of cases require access to the contact list, digital camera, file (photo) gallery, etc. during installation. Without these permissions, such applications are often impossible to install, or they completely or partially lose their functionality. It is this position of the mobile application developers that allows them to get legal access to a wide range of user data. The list of collected data can be found in the license agreement for the software, which most users do not study and a priori accept its conditions to be able to operate it.

Collected user data is sent to system and application software using the TCP / UDP protocol stack. In practice, this leads to an increase in the number of outgoing connections with different IP addresses and, as a result, to an increase in the volume of transmitted Internet traffic. Moreover, it should be noted that selective blocking of some system and application software connections with developer hosts also does not affect their performance and functionality.

The above circumstances make it necessary:

– to carry out experimental fixation of all outgoing connections of the CD with the digital space, which are under the control of mobile and stationary systems;

– performing their analysis, the results of which will allow classifying connections into functional and non-functional;

– perform blocking of all outgoing third-party connections of the CD with the digital space while maintaining the operability of its system and application software (SW).

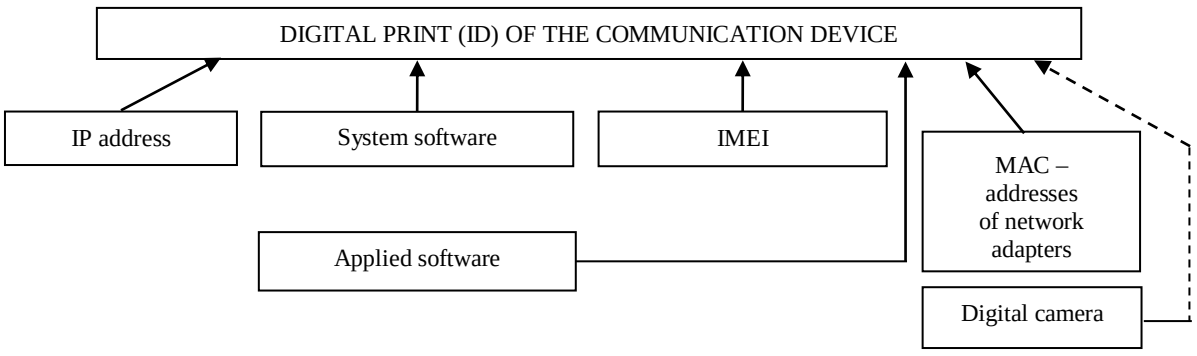

Fig. 5. Formation of a digital fingerprint of the CD

The subject of the research is the CD running the Android mobile OS and the CD running the stationary OS Windows 7/10.

Research tools: NoRoot Firewall (Android OS) [24]; extension of the built-in firewall Windows Firewall Control (OS Windows 7/10) [25], Wireshark network traffic analyzer (OS Windows 7/10) [26].

4. Methodology for classifying incoming and outgoing connections of system and application software

The most common way to organize connections between the application and system SW of mobile CDs and the basic services of their developers is to use special tools of cloud services and technologies [29-30]. This approach is less expensive and allows you to reduce the load on your servers by organizing access to cloud scalable servers for the developed application and system SW and its databases (Fig. 6).

In most cases, when using cloud services, application and system software developers redirect traffic through their DNS servers. In this case, it becomes possible for the cloud service to intercept all traffic, including those transmitted over the HTTPS protocol, which in itself is a privacy threat for users [27, 28].

From the point of view of ensuring security standards to eliminate possible leaks of user data in cloud services, the transfer, storage, and processing of user data should be carried out on the servers of the system /application software developer (Fig. 7) [1, 2].

To reduce possible user data leaks, it is necessary to classify the working connections of system and application SW into functional and non-functional ones [29], where:

1. **Functional connections** will be understood as such connections that are established exclusively with the hosts (servers) of the developer and ensure the implementation of the inherent functionality of the system and application SW. Forced blocking of such connections leads to a violation or partial loss of the inherent functionality of the system and application SW.

2. **Non-functional connections** will be understood as those that are established with hosts (including cloud services) that are not related to the implementation of the inherent functionality of the system and application SW. Forced blocking of such connections does not lead to a violation or partial loss of the inherent functionality of the system and application SW (Fig. 8).

Audit of outgoing connections of system and application software includes the following:

1. Registration of IP addresses and port numbers of outgoing and incoming connections for system and application SW;

2. Consistent blocking of connections to IP addresses of hosts:

- cloud services;
- services that do not belong to the system and application software servers;

3. Control of the functionality incorporated by developers into the system and application software after each lock is performed.

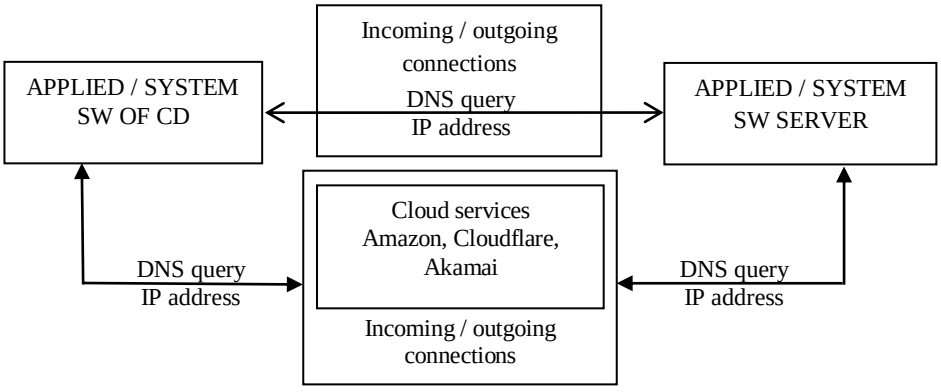

Fig. 6. Diagram of software access to the Internet

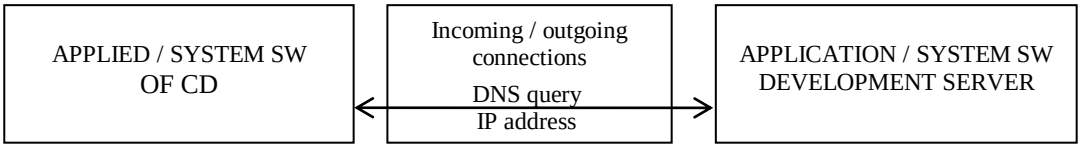

Fig. 7. Scheme of secure organization of system /application SW connections with the system /application SW server

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

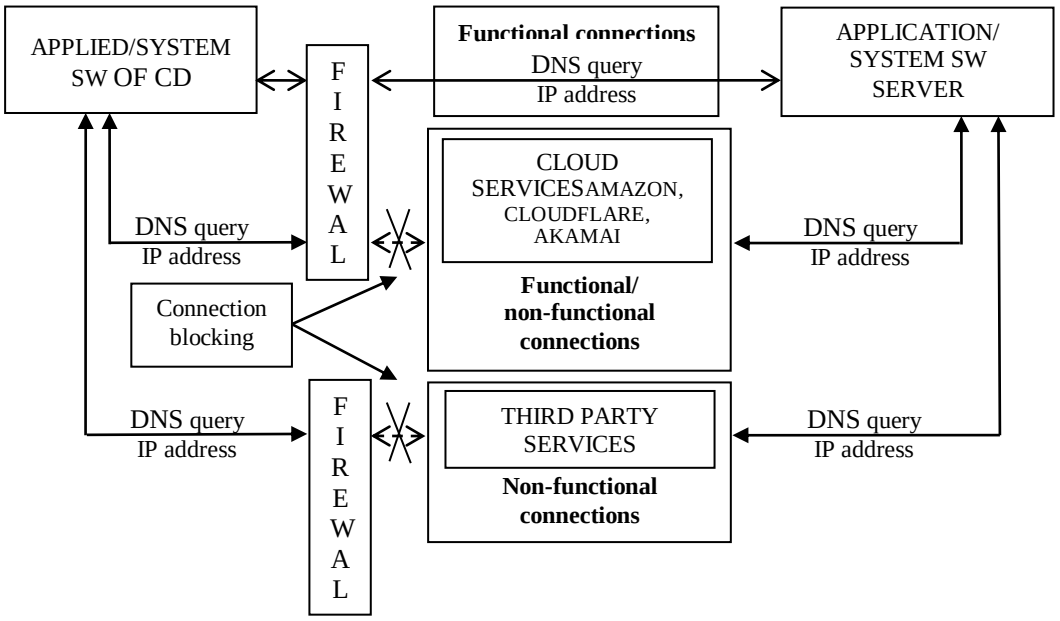

Fig. 8. Blocking scheme for non-functional system / application SW connections to the system /application SW server

Such an organization of outgoing connections will significantly reduce the probability of user data leaks, by identifying and blocking non-functional connections of the system and application software of the CD [30].

Classification of outgoing connections of system and application software was based on the audit of fixed and mobile traffic of various MCs (Fig. 9).

The results of the classification into functional and non-functional connections are presented in tables 2 – 5.

In tables 2 – 5, the "On" marker defines the inclusion, and the "Off" marker defines the disabling of the connections blocking.

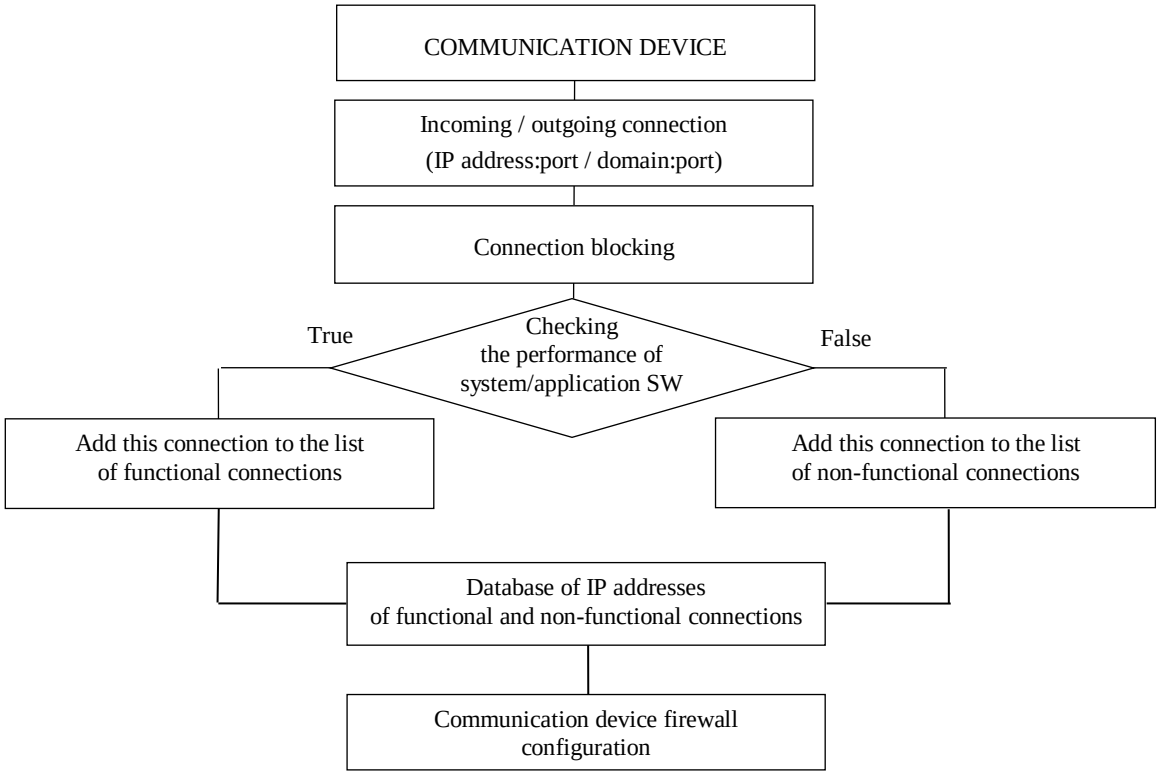

Fig. 9. Classifying incoming / outgoing system / application SW connections into functional and non-functional connections

Table 2

Functional and non-functional connections of the CD running Android

System and applied software		Incoming / outgoing connections (IP addresses / host domains)				IT companies	Block functional connection (On/Off)	Block non-functional connection (On/Off)	Work-ability software (On/Off)
		Functional	Port	Non-functional	Port				
1		2	3	4	5	6	7	8	9
Messengers	Telegram	149.154.167.50	443			Telegram Messenger Inc	Off		On
		149.154.167.151	443						
		149.154.167.96	443						
		149.154.167.55	443						
		149.154.167.51	443						
		149.154.167.91	443						
Messengers	Telegram			149.154.167.*	5222	Telegram Messenger Inc		On	On
				149.154.165.*	5222				
				5.28.194.8	443				
				5.28.194.8	5222				
				91.108.56.192	5222				
				104.16.248.249	443	Cloudflare		On	On
				196.55.217.*	443	African Network Information Center			
				dns.google	443	Google			
		whatsapp-chatd-edge-shv-02-atl3.facebook.com	5222			Facebook			
		whatsapp-chatd-edge-shv-02-waw1.facebook.com	5222						
		whatsapp-cdn-shv-01-waw1.fbcdn.net	443						
		whatsapp-cdn-shv-01-frt3.fbcdn.net	443						
		whatsapp-cdn-shv-01-arn2.fbcdn.net	443						
				whatsapp-chatd-edge-shv-02-waw1.facebook.com	80			On	On
				*.compute-1.amazonaws.com	5222	Amazon			
	Viber	185.151.204.13	443			Rakuten	Off		On
		185.151.204.7	443						
		185.151.204.15	443						
		185.151.204.9	443						
		ec*.compute-1.amazonaws.com	4244						
		ec*.compute-1.amazonaws.com	443						
				*.deploy.static.akamaitechnologies.com	443	Akamai Tech-тologies		On	On
				*.bc.googleusercontent.com	443	Google			
				*.1e100.net	443				
				*.r.cloudfront.net:	443	Cloudflare			
				s3-website.eu-central-1.amazonaws.com	80	Amazon			
				52.0.252.*	5242				
				151.101.114.110	443	Fastly Inc			
				192.48.236.*	443	Twitter			
				65.9.69.*	443	Amazon CloudFront			
				65.9.71.*	443				
Mobile applications	Facebook	*.facebook.com	443			Facebook	Off		On
		edge-star-mini-shv-01-waw1.facebook.com	443						
		edge-star-shv-01-waw1.facebook.com	443						
		edge-mqtt-shv-01-waw1.facebook.com:	443						
		edge-mqtt-mini-shv-01-waw1.facebook.com	443						
		*.fbcdn.net	443						
		facebook.dataline.net.ua	443						
		edge-star-mini-shv-01-frx5.facebook.com	443						

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

End of Table 2

1		2	3	4	5	6	7	8	9
Mobile applications	Facebook	edge-mqtt-mini-shv-01-frx5.facebook.com	443						
		edge-star-shv-01-otp1.facebook.com	443						
		edge-star-mini-shv-01-otp1.facebook.com	443						
		facebook-casche.te.net.net	443						
				*.compute-1.amazonaws.com	4244	Amazon		On	On
				*.r.cloudfront.net	443				
				*.eu-west-1. compute.amazonaws.com	4244				
				104.16.221.74	443	Cloudflare			
				104.16.199.73	443				
				104.20.110.39	443				
				91.198.36.60	443	LLC Digital Ventures			
				*.utel.net.ua	443	Internet Invest Ltd			
				*.vps-default-host.net	443	Hosting Ukraine LLC			
				ua*.host.hit.gemius.pl	443	Gemius S. A			
				es49.mirohost.net	443	Internet Invest, Ltd			
				*.ip.kyivstar.net	443	Kievstar			
				pl-sh.host4.biz	443	Ukrnames			
				*.dialup.umc.net.ua	443	Vodafone Ukraine			
				*.1e100.net	443	Google			
				te.net.net	443	Domains By Proxy, LLC			

Table 3

Functional and non-functional connections in Android OS

System software services / modules	Incoming / outgoing connections (Host IP addresses)				IT companies	Block functional connection (On/Off)	Block non-functional connection (On/Off)	Workability software (On/Off)
	Functional	Port	Non-functional	Port				
1	2	3	4	5	6	7	8	9
analytics			47.241.109.173	443	Alibaba		On	On
			47.241.21.203	443				
			47.241.67.215	443				
			47.241.109.173	443				
			47.241.109.173	443				
			161.117.183.182	443	Alibaba Cloud		On	On
			107.155.53.108	443	Zenlayer (Kingsoft cloud corporation limited)		On	On
			107.155.53.85	443				
msa			47.241.108.101	443	Alibaba		On	On
Battery & Performance, Security, Geodata from Multiple Sources, Screen Recorder, Interface, Ribbon Widgets, Settings, Report, Call Control, Input Devices Settings Storage			47.241.69.153	443	Alibaba		On	On
			47.88.222.244	80				
			65.9.71.10	443	Amazon CloudFront			
			65.9.71.64	443				
			65.9.71.75	443				
			65.9.71.43	443				
			161.117.96.220	443	Alibaba Cloud			
			161.117.160.12	443				
			161.117.204.141	443				
			161.117.97.83	80				
			161.117.71.89	80				
			161.117.71.92	80				
			161.117.97.84	80				
			161.117.57.135	443				

End of Table 3

1	2	3	4	5	6	7	8	9
Battery & Performance, Security, Geodata from Multiple Sources, Screen Recorder, Interface, Ribbon Widgets, Settings, Report, Call Control, Input Devices Settings Storage			161.117.94.82	443				
			161.117.183.226	443				
Google Services, Framework, Google Backup, Google Play Services	142.251.1.188	443			Google	On		On
	172.253.113.188	5228				On		On
	142.250.203.138	443			Google Cloud	On		On
MTP Host, Downloads, Media Storage			161.117.71.156	443	Alibaba Cloud		On	On
			161.117.71.89	80				
			161.117.71.92	80				
			161.117.71.84	80				
			161.117.71.83	80				
			161.117.96.220	80				
			161.117.94.82	80				
			161.117.160.82	80				
			47.74.233.137	80	Alibaba		On	On
			47.74.233.244	80				
Gallery (photo and video)			161.117.97.84	80	Alibaba		On	On
			161.117.97.83	80				
			161.117.71.92	80				
			161.117.71.89	80				
			161.117.204.141	80				
			47.74.233.137	80				
			47.88.222.244	80				
			161.117.96.220	80	Alibaba Cloud		On	On
			161.117.94.82	80				
			161.117.160.12	80				

Table 4

Functional and non-functional connections of the CD in OS Windows 7/10

System software services / modules	Incoming / outgoing connections (Host IP addresses)				IT compa-nies	Block functional connection (On/Off)	Block non-functional connection (On/Off)	Work-ability software (On/Off)
	Functional	Port	Non-functional	Port				
1	2	3	4	5	6	7	8	9
svchost.exe	64.4.0.0 – 64.4.31.255	443			Microsoft Corporation	On		On
	65.55.32.0 – 65.55.47.255	443						
	65.55.0.0 – 65.55.31.255	443						
	65.55.128.0 – 65.55.255.255	443						
	65.55.96.0 – 65.55.127.255	443						
	64.4.0.0 – 64.4.31.255	443						
	64.4.48.0 – 64.4.63.255	443						
	213.199.179.0 – 213.199.179.31	443						
	52.167.0.0 – 52.167.255.255	443						
	207.68.160.0 – 207.68.167.255	443						
	204.79.196.0 – 204.79.197.255	443						
	134.170.0.0 – 134.170.63.255	443						
	134.170.176.0 – 134.170.191.255	443						
	131.253.48.0 – 131.253.63.255	443						
svchost.exe	64.4.0.0 – 64.4.31.255	443				On		On
	65.55.32.0 – 65.55.47.255	443						
	65.55.0.0 – 65.55.31.255	443						
	65.55.128.0 – 65.55.255.255	443						
	65.55.96.0 – 65.55.127.255	443						
	64.4.0.0 – 64.4.31.255	443						
	64.4.48.0 –64.4.63.255	443						

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

End of Table 4

1	2	3	4	5	6	7	8	9
	213.199.179.0 – 213.199.179.31	443						
	52.167.0.0 –52.167.255.255	443						
	207.68.160.0 –207.68.167.255	443						
svchost.exe	204.79.196.0 – 204.79.197.255	443						
	134.170.0.0 –134.170.63.255	443						
	134.170.176.0 – 134.170.191.255	443						
	131.253.48.0 – 131.253.63.255	443						
			23.57.96.0 –23.57.111.255	443				
			23.48.104.0 –23.48.111.255	443				
			23.36.32.0 –23.36.47.255	443				
			23.223.20.0 –23.223.20.255	443				
			23.218.192.0 – 23.218.223.255	443				
			23.205.208.0 – 23.205.215.255	443				
			23.204.64.0 –23.204.79.255	443	Akamai Tech-nologies, Inc			
svchost.exe			23.102.0.0 –23.102.63.255	443			On	On
			2.22.61.0 –2.22.61.255	443				
			104.96.128.0 – 104.96.159.255	443				
			172.228.0.0 – 172.231.255.255	443				
			23.36.160.0 – 23.36.175.255	443				
			104.77.160.0 – 104.77.163.255	443				
			104.95.176.0 – 104.95.183.255	443				
			23.48.164.0 – 23.48.167.255	443				
			23.40.112.0 – 23.40.119.255	443				
	207.46.220.0 – 207.46.223.255	443						
	207.46.112.0 – 207.46.119.255	443						
	207.46.96.0 –207.46.111.255	443						
	157.55.48.0 – 157.55.63.255	443						
	137.117.128.0 –137.117.255.255	443						
	137.116.0.0 –137.116.127.255	443						
	111.221.64.0 –111.221.67.255	443						
	191.237.208.0 –191.237.223.255	443			Microsoft Azure	On		On
	191.232.80.0 – 191.232.80.63	443						
	191.232.138.0 –191.232.139.255	443						
	191.232.138.0 –191.232.139.255	443						
svchost.exe	20.96.0.0 – 20.127.255.255	443						
	52.169.0.0 – 52.169.255.255	443						
	204.79.196.0 – 204.79.197.255	443						
			195.138.255.0	80	Core Back Bone		On	On
			62.140.236.163	80	Fiord Networks, UAB		On	On
			62.140.236.170	80				
			35.186.238.101	80	Google Cloud		On	On
			152.199.0.0 - 152.199.63.255	80	Verizon Internet Services		On	On

Table 5

Services for collecting user data on the Internet

IT companies	Data collection service	Domain	IP address	Port
Yandex LLC	Yandex Metrika	mc.yandex.ru	87.250.250.119	443
Google LLC	Google Analytics	googletagmanager.com	142.250.185.200	443
		www-google-analytics.l.google.com	172.217.16.46	53
	Google	pagead46.l.doubleclick.net	216.58.215.66	53
			216.58.208.194	443
	Google Cloud	googlesyndication.com	142.250.185.68	53
United Network LLC	Liveinternet	counter.yadro.ru	88.212.201.204	443
Mail.Ru LLC	Relap	relap.io	95.163.37.253	80
Mediascope Joint Stock Company	TNS-Counter	tns-counter.ru	194.226.130.226	443
Hetzner Online GmbH	Openstat	openstat.com	138.201.159.191	53
Rambler Internet Holding LLC	Rambler TOP	top100.rambler.ru	81.19.89.1	443
Amazon	Easycounter	easycounter.com	52.1.22.171	80

Table 4 shows the functional connections of the CD under the control of OS Windows 7/10. In this work, the study of the traffic of the system and application software of the CD with the digital space was carried out. This made it possible to establish the IP addresses and domains of the user data collection services. They are owned by leading IT companies in the digital market for user data. Table 5 shows the domain names of user data collection services owned by IT companies. Data collection presented in Table 2-5 was carried out by comprehensive monitoring of DNS traffic of various stationary and mobile CDs for 180 days performed by specialized applied software [24-26].

5. Analysis of connections of information systems with digital space

Analysis of the experimentally obtained domains and IP addresses with which the system and applied software of the CD establish functional and non-functional connections, made it possible to obtain the following data.

5.1. Android operating system application software (see Table 2)

– Domain 1e100.net – connected to ns-servers google.com. Belongs to Google Inc.;

– Domain *.bc.googleusercontent.com is connected to Google's virtual machines as part of the Google Compute Engine project, a component of the service infrastructure (IaaS) of the Google cloud platform;

– Domain te.net.net is owned by Bodis, LLC, which provides monetization and domain traffic management services;

– Domain akamaitechnologies.com – connected to ns-servers akamaistream.net. Owned by Akamai Technologies, a content delivery platform, and website acceleration application provider;

– Domain cloudfront.net – connected to ns-servers: awsdns-35.org, awsdns-07.co.uk, awsdns-52.com, awsdns-19.net. Owned by Amazon, a cloud service provider;

– Domain host.hit.gemius.pl owned by Gemius, a media consumption research company that develops tools used to optimize advertising campaigns;

– Domains compute-1.amazonaws.com и eu-central-1.amazonaws.com owned by Amazon IT company;

– Domain fbcdn.net – connected to ns-servers: facebook.com. Belongs to the Facebook social network. Used to serve the delivery of static content (videos and photos from CDN);

– pl-sh.host4.biz – is owned by Host4Biz sp. zo.o., which provides virtual hosting services (Poland);

– es49.mirohost.net – Internet service provider Internet Invest Ltd (Ukraine).

5.2. Android operating system (see Table 3)

– IP address 161.117.98.205 – "device search" system software module. The owner is Alibaba Cloud Corporation, which provides cloud solutions for suppliers and integrators of intelligent IoT systems;

– IP address 47.241.108.101:443 – "msa" system software module. The owner is Alibaba, an e-commerce and cloud computing and entertainment company;

– IP address 107.155.53.108 – is owned by cloud computing company Kingsoft Cloud.

5.3. Windows operating system (see Table 4)

– the *svchost.com* module accesses the pool of IP addresses belonging to the IT company Akamai Technologies Inc., Core Back Bone, Fiord Networks UAB, Verizon Internet Services, Google Cloud;

– the *svchost.com* module accesses a pool of IP addresses owned by Microsoft IT and its Microsoft Azure division.

5.4. Analysis of the experimental results obtained

The analysis of the results obtained for the experimental verification of the operability of the system and applied software of the CD when blocking its functional and non-functional connections (see Table 2-4) made it possible to establish the following:

1. In the Android operating system:

1.1. Facebook applied software establishes connections with transit Internet providers, as well as with hosts owned by organizations specializing in monetizing user domain traffic.

1.2. Telegram applied software attempts to redirect its traffic through Google's DNS servers.

1.3. Viber establishes connections with Amazon cloud services, Cloudflare, Akamai, Google, and Fastly, as well as with the social platform Twitter.

1.4. All applied software running under Android OS, when activated, makes outgoing connections to the host 1e100.net, owned by Google Inc.

1.5. Functional connections of modules and services of the Android OS system software are performed automatically upon detection of any connection of the CD to the digital space of the Internet, regardless of the network interface used.

1.6. Android OS system software (see Table 3) establish non-functional connections with IP addresses belonging to IT companies Kingsoft cloud corporation limited, Alibaba, Alibaba Cloud, and Amazon CloudFront providing traffic transit, and cloud computing services.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

1.7. Blocking the installed functional and non-functional connections of the Android OS system software does not lead to a violation of its performance.

2. In the Windows operating system:

2.1. OS Windows 7/10 (see Table 4) system software establishes functional connections with IP addresses belonging to its developer, the IT company Microsoft.

2.2. OS Windows 7/10 system software establishes non-functional connections with IP addresses belonging to IT companies Akamai Technologies Inc, Core Back Bone, Fiord Networks UAB, providing traffic transit, cloud computing, and content delivery services.

2.3. Blocking functional and non-functional connections of the OS Windows 7/10 system software modules does not lead to a violation of its performance.

6. Protection of data from leaks in information systems

Cloud services provide a wide range of monitoring tools for hosted applications implemented using PaaS and IaaS platforms. This allows the owners of such applications to receive detailed statistics not only about user actions when interacting with their applications but also about user actions performed on the Internet [31, 32]. Of course, this fact allows us to assert the receipt and use of the same data by the owners of cloud services, regardless of the license terms for using their PaaS and IaaS platforms. This becomes especially relevant when cloud services are affiliated divisions of IT companies, such as Amazon, Google, etc. (Fig. 10) [33-35].

Therefore, to reduce the probability of collecting user data, it is necessary to use system and application software as a priority, which retains its functionality while blocking non-functional connections to cloud services (see Fig. 9 and Table 2-4).

To protect user data from leaks and collection by IT companies, it is necessary to:

1. Organize redirection of DNS requests of system and application software bypassing the provider's DNS servers (Fig. 11) through public DNS servers (Table 6);

2. Use the DoH protocol at the level of system/application software of the CD.

3. Organize outgoing / incoming IS network traffic filtering:

– use firewalls in the CD that operate at layer 3 (IP address) and layer 4 (TCP / UDP) of the OSI network model;

– at the hardware level of the IS, it is necessary to use firewalls operating at layer 7 of the OSI network model [36, 37].

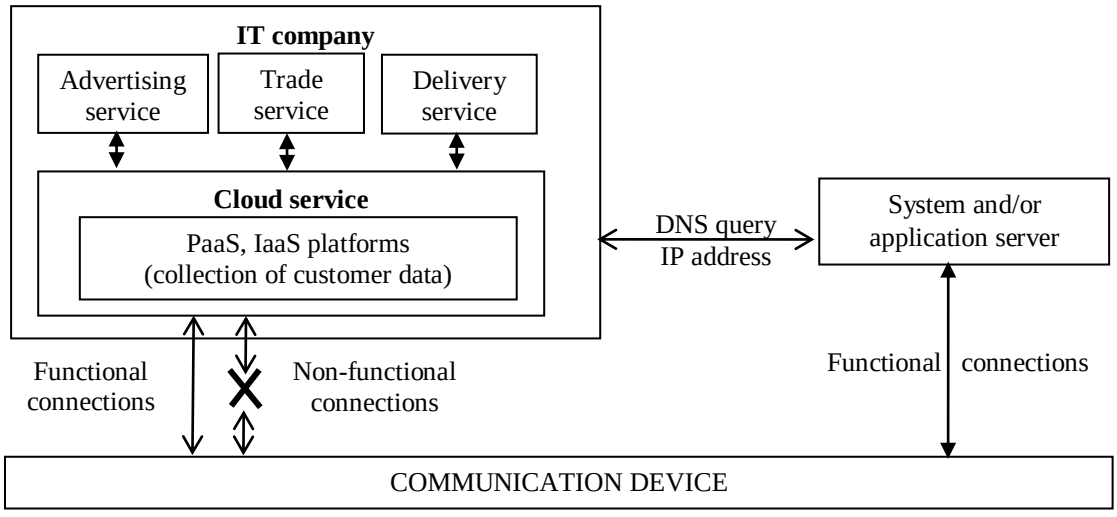

Fig. 10. Data collection by a cloud service when interacting with the CD

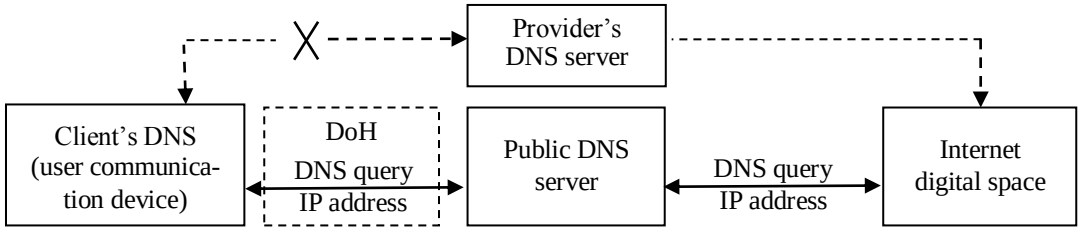

Fig. 11. Redirecting DNS requests to bypass the DNS servers of the provider

Table 6

Public DNS servers

DNS server	IP address
Cisco Umbrella (OpenDNS)	208.67.222.222 208.67.220.220
Cloudflare	1.1.1.1, 1.0.0.1
Comodo Secure DNS	8.26.56.26, 8.20.247.20
Quad9 DNS	9.9.9.9, 149.112.112.112
Adguard DNS	94.140.14.14, 94.140.15.15
FreeDNS	37.235.1.174, 37.235.1.177
DNS.Watch	84.200.69.80, 84.200.70.40
Norton ConnectSafe	199.85.126.10, 199.85.127.10
DNS Advantage	156.154.70.1, 156.154.71.1
Freenom World	80.80.80.80, 80.80.81.81

This will allow you to block all incoming and outgoing connections of the application and system software of the CD that corresponds to:

– IP addresses / domains of non-functional traffic of system and applied software;

– IP addresses / domains of data collection services.

In general, the organization of filtering outgoing / incoming network traffic of the IS using a firewall is shown in Fig. 12.

To achieve the stated goal of the study, it is necessary to organize the process of filtering network traffic of the IS responsible for:

– data collection by applied software (see Table 2);

– data collection by system software (see Table 3-4);

– collection of user data by specialized services (see Table 5).

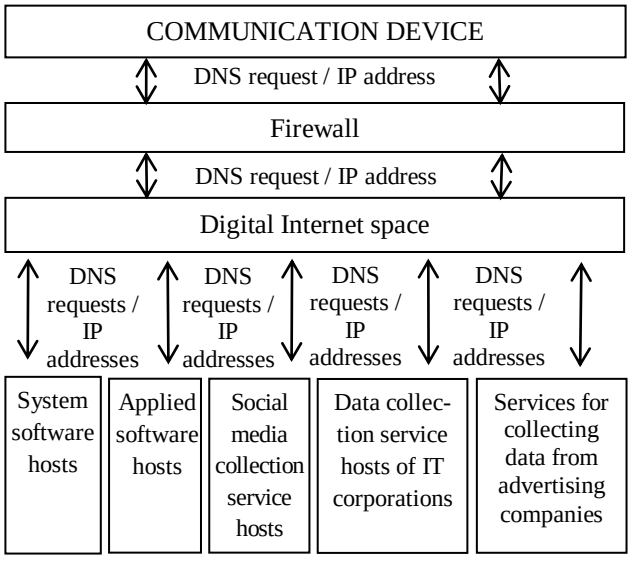

Fig. 12. Organization of IS protection from data collection and leakage

Filtering IS traffic in practice is possible using three approaches to protect against data collection.

The first software-based approach is implemented based on the software organization of the filtering of IP connections. It is implemented with the help of specialized applied software for the CD (Fig. 13):

– for Android OS, NoRoot Firewall is used [24];

– for the Windows OS family, the built-in firewall extension Windows Firewall Control is used [25].

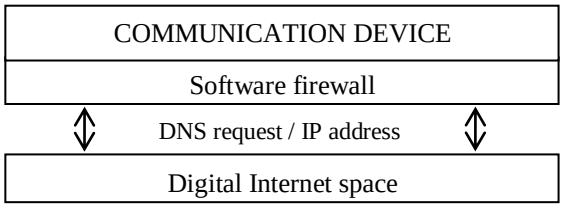

Fig. 13. Software-based IS protection from data collection

The second hardware-based approach is implemented based on the hardware organization of filtering IP connections using the settings of the built-in firewall of the organization's network equipment – a router, a server, etc. (Fig. 14).

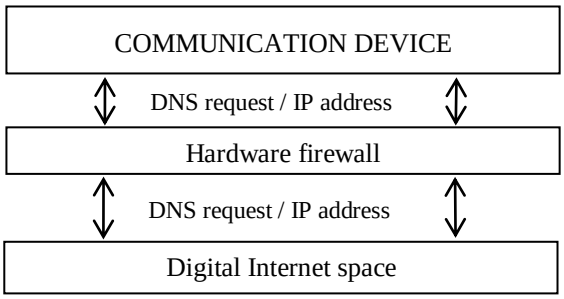

Fig. 14. Hardware-based IS protection from data collection

The third mixed approach is implemented based on the mixed use of software-based and hardware-based filtering of IS connections (Fig. 15).

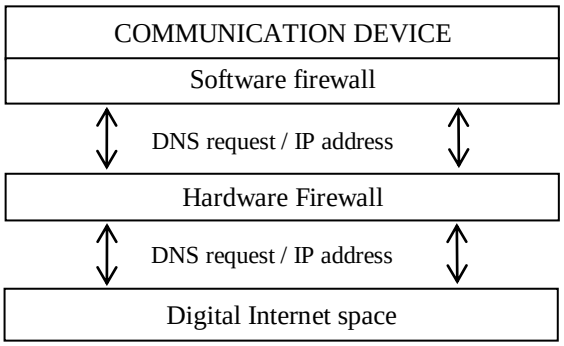

Fig. 15. Mixed IS protection from data collection

Experimental verification of the proposed approaches to IS protection showed that hardware filtering of information system connections should not exclude software filtering implemented directly on the IS itself or vice versa. On the contrary, the use of a combined approach (third model) of filtering network traffic of IS

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

will allow, with a higher probability, to exclude leaks of user data. The effectiveness of the third model of IS protection is because in the CD under the control of the mobile OS Android, its spontaneous shutdown of the firewall is observed. This is due to the incompatibility of some versions of Android OS with the current version of the firewall.

In Windows OS, the Task Scheduler component performs the following functions:

– task management;
– event processing;
– collection of telemetry data about user actions.

The collected data is sent to Microsoft servers:

– db5sch103102609.wns.windows.com
– dev.virtualearth.net
– df.telemetry.microsoft.com
– diagnostics.support.microsoft.com
– disc101-prod.do.dsp.mp.microsoft.com
– telecommand.telemetry.microsoft.com
– eu.vortex.data.microsoft.com
– fs.microsoft.com
– geover-prod.do.dsp.mp.microsoft.com
– i1.services.social.microsoft.com
– kv101-prod.do.dsp.mp.microsoft.com
– ls2web.redmond.corp.microsoft.com
– mobile.pipe.aria.microsoft.com
– oca.telemetry.microsoft.com
– reports.wes.df.telemetry.microsoft.com
– redir.metaservices.microsoft.com
– services.wes.df.telemetry.microsoft.com
– settings.data.glbdns2.microsoft.com
– settings.data.microsoft.com
– settings-sandbox.data.microsoft.com
– settings-win.data.microsoft.com
– settings-win-ppe.data.microsoft.com
– sqm.df.telemetry.microsoft.com
– spynet2.microsoft.com
– spynetalt.microsoft.com
– web.vortex.data.microsoft.com
– wes.df.telemetry.microsoft.com
– win10.ipv6.microsoft.com
– win1710.ipv6.microsoft.com
– vortex.data.microsoft.com
– v10.vortex-win.data.microsoft.com
– vortex.data.glbdns2.microsoft.com
– telemetry.microsoft.com
– telemetry.urs.microsoft.com
– teredo.ipv6.microsoft.com
– test.activity.windows.com
– vortex-sandbox.data.microsoft.com
– vortex-win.data.microsoft.com
– watson.microsoft.com
– watson.ppe.telemetry.microsoft.com
– watson.telemetry.microsoft.com
– storecatalogrevocation.storequality.microsoft.com
– survey.watson.microsoft.com
– sqm.telemetry.microsoft.com
– tsfe.trafficshaping.dsp.mp.microsoft.com

To reduce the amount of information collected in Windows OS, it is necessary to perform a more flexible configuration of its component – Task Scheduler [38]. With its help, management, processing, and fixing user actions in the OS is carried out.

Table 7

List of Task Scheduler tasks recommended for deactivation

№	Description	Task Type
1.	The task aggregates and uploads Application Telemetry information	\Microsoft\Windows\Application Experience\AitAgent
2.	The task collects data for SmartScreen in Windows	\Microsoft\Windows\AppID\SmartScreenSpecific
3.	The task collects technical data about the operation of devices and related software	\Microsoft\Windows\Application Experience\Microsoft Compatibility Appraiser
4.	The task collects information about installed applications and their updates	\Microsoft\Windows\Application Experience\ProgramDataUpdate
5.	The task collects and uploads autochk Software Quality Management data	\Microsoft\Windows\Autochk\Proxy
6.	The task collects additional data and information about processes running in kernel mode (Kernel CEIP) and sends this data to Microsoft	\Microsoft\Windows\Customer Experience Improvement Program\KernelCeipTask
7.	The task collects Universal Serial Bus related statistics and information about computer and sends it to the Windows Device Connectivity engineering group at Microsoft	\Microsoft\Windows\Customer Experience Improvement Program\UsbCeip
8.	The the Windows Disk Diagnostic reports general disk and system information to Microsoft	\Microsoft\Windows\DiskDiagnostic\Microsoft-Windows-DiskDiagnosticDataCollector
9.	The task collects information about Software Quality Management and sends usage data to Microsoft	\Microsoft\Windows\IME\SQM data sender

End of Table 7

№	Description	Task Type
10.	The task works by using preferred indexing options for fast searches. It runs in the background and writes a kind of table of contents for the files stored on the computer	Microsoft\Windows\Media Center\ActivateWindowsSearch
11.	The task analyzes the system looking for conditions that may cause high energy use and battery life problems	Microsoft\Windows\Power Efficiency Diagnostics\AnalyzeSystem
12.	The task uses the error reporting feature to provide customers with troubleshooting information, solutions, or updates for their specific problems. Microsoft developers can use this infrastructure to obtain information that they can use to improve their applications	Microsoft\Windows\Windows Error Reporting\QueueReporting
13.	The task learns about managing applications, including how to remove background task resource restrictions	Microsoft\Windows\Application Experience\ProgramDataUpdater
14.	The task collects network configuration information	Microsoft\Windows\NetTrace\GatherNetworkInfo
15.	The task collects and transmits data about remote tracking applications	Microsoft\Windows\Application Experience\AitAgent

All recorded data about user actions are sent using the *svchost.com* module to Microsoft servers.

To deactivate recommended tasks in the Task Scheduler (see Table 7), you need to execute the *schtasks* command with the *change* key in the following format at the Windows OS command line with administrator rights:

schtasks /change /tn "Task type" /disable

After completing the deactivation of all tasks, you must reboot the OS.

Conclusions

1. The principles of data exchange between a typical IS and the digital space of the Internet have been analyzed. It has been shown that DNS queries of the system and applied software of the CD are transmitted in clear text and recorded in the logs of the provider's DNS server, which reduces the privacy of users. The ways of increasing the privacy of users by introducing measures to protect their DNS traffic have been identified.

2. The principles of constructing digital prints of the CD have been analyzed. It is shown that the process of building digital prints is based on the use of identification methods. It was found that the implementation of the process of the CD unique identification is inextricably linked with the organization of collecting the maximum possible information about the hardware and software configuration of the CD.

3. It has been established that non-functional connections of the system and application software of the CD are established with hosts that belong to IT companies that provide services for the transit, collection, processing, and monetization of user data.

4. The analysis of outgoing connections of the personal computer and smartphone and applied software of the CD has been carried out. Criteria are proposed that make it possible to unambiguously classify functional and non-functional connections of the system and applied software of the CD. An experimental check of the operability of the system and applied software of the CD was carried out when organizing the blocking of functional and non-functional IS connections. It was found that when their blocking is performed, the operability of the IS is not impaired.

5. A set of measures has been developed to organize protection against leaks of user data in IS. Its implementation is based on the use of a firewall that blocks functional and non-functional connections of the system and applied software of the CD, determined experimentally and systematized in Table 2-5.

6. A set of approaches is proposed for the software and hardware organization of protecting user data from leaks in the IS, taking into account the possible unstable operation of the application software. An experimental check of the proposed approaches for organizing data protection has been carried out. It has been experimentally established that the most effective is the model based on the simultaneous use of software and hardware protection of user data.

Practical application of the proposed approaches for protecting data from leaks will allow users of the CD:

– increase privacy in the digital space of the Internet;

– reduce the accuracy of digital profiling of the user's CD;

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

– to preserve the functionality and operability of the system and applied software of the CD;

– reduce mobile traffic of the CD with the digital space of the Internet, which will reduce the cost of paying for the services of mobile providers;

– prevent the collection of data by IT companies;

– reduce the likelihood of manipulative influences from IT companies on decisions;

– reduce the likelihood of manipulative influences from IT companies on the decisions made by users of the CD.

The obtained results can be used in a line of existing and promising approaches in the design of difficult, complex, hybrid, software, and technical systems for protecting user data from leaks in the digital space. Further research should be devoted to improving methods for detecting and blocking data leaks from communication devices. Improving the set of measures that allow to organize protection against leaks of user data in information systems should consider the changing modern legislative norms of each country for the protection of the citizens' personal data.

The results of the experiments were evaluated and analyzed. The following safety indicators have been achieved:

– reduction in the volume of non-functional connections for OS Android amounted to 10-30 MB per day;

– reduction in the volume of non-functional connections for OS Windows amounted to 15-60 MB per day;

– reduction in the number of showing ads by 50%;

– reduction in the number of showing relevant ads by 70%.

Contribution of the authors: setting goals and objectives for developing a set of measures to protect against leaks of user data in the digital space, developing the concept and methodology of the study – **Olexander Zadereyko**; review and analysis of references, suggesting the algorithm for classifying incoming / outgoing system / application SW connections into functional and non-functional connections, and analysis and presentation of results – **Olena Trofymenko**; analysis of the data exchange principles between the information system and the digital space, analysis of the methods of identification of communication devices in the digital space – **Yuliia Prokop**; determination of the criteria for classifying the connections of the information system with the digital space, analysis of the connections of the user's information system with the digital space – **Nataliia Loginova**; analysis of the results obtained for the experimental verification of the operability of the system and applied software of the CD when blocking its functional and non-functional connections – **Anastasiia Dyka**; fixing of the outgoing and incoming connections of the information system with the digital space, analysis of the principles of data exchange between a typical user information system and the digital space of the Internet. It is proposed to increase the privacy of Windows users. This is implemented by deactivating tasks in the Task Scheduler – **Serhii Kukharenko**.

All the authors have read and agreed to the published version of the manuscript.

References (GOST 7.1:2006)

1. General Data Protection Regulation (GDPR) Compliance guidelines [Electronic resource]. – Access mode: https://gdpr.eu/?cn-reloaded=1. – 4.08.2022.

2. Toscano, J. Data Privacy Issues Are the Root of Our Big Tech Monopoly Dilemma [Electronic resource] / J. Toscano. – Access mode: https://www.forbes.com/sites/joetoscano1/2021/12/01/data-privacy-issues-are-the-root-of-our-big-tech-monopoly-dilemma/?sh=cd05cdb3cfd7. – 4.08.2022.

3. Gugelmann, D. On Data and Privacy Leakage in Web Traffic [Text] / D. Gugelmann // Doctoral Thesis. – 2015. – 186 p. DOI: 10.3929/ethz-a-010615756.

4. Трофименко, О. Г. Еволюція поглядів на інформаційні війни в епоху інформаційного суспільства [Текст] / О. Трофименко, Я. Дубовой // Порівняльно-аналітичне право: електронне наукове фахове видання. – 2017. – № 1. – С. 189-192.

5. Privacy Nudging in Search: Investigating Potential Impacts [Text] / S. Zimmerman, A. Thorpe, C. Fox, U. Kruschwitz // Human Information Interaction and Retrieval: Proceedings of the Conference. – 2019. – P. 283-287. DOI: 10.1145/3295750.3298952.

6. Google в твоей голове [Electronic resource]. – Access mode: https://eurasia.film/2019/08/google-v-tvoej-golove/. – 4.08.2022.

7. Esteve, A. The business of personal data: Google, Facebook, and privacy issues in the EU and the USA [Text] / A. Esteve // International Data Privacy Law. – 2017. – Vol. 7(1). – P. 36-47. DOI: 10.1093/idpl/ipw026.

8. ISO/IEC 2382:2015 Information technology – Vocabulary [Electronic resource]. – Access mode: https://www.iso.org/standard/63598.html. – 4.08.2022.

9. Kolisnyk, M. Vulnerability analysis and method of selection of communication protocols for information transfer in Internet of Things systems [Text] / M. Kolisnyk // Radioelectronic and Computer Systems. – 2021. – No 1(97). – P. 133-149. DOI: 10.32620/reks.2021.1.12.

10. Technique for IoT malware detection based on control flow graph analysis [Text] / K. Bobrovnikova, S. Lysenko, B. Savenko, P. Gaj, O. Savenko // Radioelec-

tronic and Computer Systems. – 2022. – No 1(101). –P. 141-153. DOI: 10.32620/reks.2022.1.11.

11. Leveraging eBPF to preserve user privacy for DNS, DoT, and DoH queries [Text] / S. Rivera, V. Gurbani, S. Lagraa, A. Iannillo, R. State // Availability, Reliability and Security (ARES'20): Proceedings of the 15th International Conference. – 2020. – No 78. – P. 1-10. DOI: 10.1145/3407023.3407041.

12. Bumanglag, K. On the Impact of DNS Over HTTPS Paradigm on Cyber Systems [Text] / K. Bumanglag, H. Kettani // Information and Computer Technologies (ICICT'20): Proceedings of the 3rd International Conference, San Jose, CA, USA. – 2020. – P. 494-499. DOI: 10.1109/ICICT50521.2020.00085.

13. Charanjeet, S. How to Enable DNS Over HTTPS in Chrome, Firefox, Edge, Brave & More? [Electronic resource] / S. Charanjeet // Fossbytes. – 2020. – Access mode: https://fossbytes.com/how-to-enable-dns-over-https-on-chrome-firefox-edge-brave/. – 4.08.2022.

14. Imana, B. Institutional privacy risks in sharing DNS data [Text] / B. Imana, A. Korolova, J. Heidemann // Proceedings of the Applied Networking Research Workshop (ANRW'21), Virtual Event, USA. – 2021. –P. 69-75. DOI: 10.1145/3472305.3472324.

15. Chang, D. Hide and Seek: Revisiting DNS-based User Tracking [Text] / D. Chang, J. Chen, Z. Li, X. Li // 2022 IEEE 7th European Symposium on Security and Privacy (Euro S&P), Genoa, Italy. – 2022. – P. 188-205. DOI: 10.1109/EuroSP53844.2022.00020.

16. Encrypted DNS -> Privacy? A Traffic Analysis Perspective [Text] / S. Siby, M. Juárez, C. Díaz, N. Vallina-Rodriguez, C. Troncoso // ArXiv. – 2020. – Vol. abs/1906.09682. – P. 1-21. DOI: 10.14722/ndss.2020.24301.

17. Liu, Y. Computer Method Research on Risk Control Identification System Based on Deep Learning [Text] / Y. Liu // Advances in Electrical Engineering and Computer Applications (AEECA): 2021 IEEE International Conference, Dalian, China. – 2021. – P. 561-565. DOI: 10.1109/AEECA52519.2021.9574442.

18. Model and Algorithms for User Identification by Network Traffic [Text] / V. Gai, I. Ephode, R. Barinov, I. Polyakov, V. Golubenko, O. Andreeva // Computer Graphics and Vision: Proceedings of the 31st International Conference. – 2021. – Vol. 2. – P. 1-11. DOI: 10.20948/graphicon-2021-3027-1017-1027.

19. The implementation of depersonalization algorithm of digital images [Text] / A. Zadereyko, A. Troyanskiy, N. Loginova, E. Trofimenko // Advanced information and communication technologies-2017 (AICT-2017): Proceedings of the 2nd International IEEE Conference. (July 4–7, 2017). – Lviv, 2017. – P. 56-61.

20. Development of an algorithm to protect user communication devices against data leaks [Text] / O. Zadereyko, Y. Prokop, O. Trofymenko, N. Loginova, O. Plachinda // Eastern-European Journal of Enterprise Technologies. – 2021. – № 1/2 (109). – P. 24-34. DOI: 10.15587/1729-4061.2021.225339.

21. Zadereyko, O. Algorithm of user's personal data protection against data leaks in Windows 10 OS [Text] / O. Zadereyko, O. Trofymenko, N. Loginova // Informatyka, Automatyka, Pomiary w Gospodarcei Ochronie Środowiska. – Lublin, 2019. – Vol. 9. – P. 41-44.

22. Detection of Data Leaks from Android Applications [Text] / S. Natesan, M. Gupta, L. Iyer, D. Sharma // II International Conference on Inventive Research in Computing Applications (ICIRCA): Proceedings of the II International Conference. – 2020. – P. 326-332. DOI: 10.1109/ICIRCA48905.2020.9183066.

23. Accessi Leaks: Investigating Privacy Leaks Exposed by the Android Accessibility Service [Text] / M. Naseri, N. Borges, A. Zeller, R. Rouvoy // Proceedings on Privacy Enhancing Technologies. – 2019. – Vol. 2. – P. 291-305. DOI: 10.2478/popets-2019-0031.

24. NoRoot Firewall [Electronic resource]. – Access mode: https://play.google.com/store/apps/details?hl=en&id=app.greyshirts.firewall. – 4.08.2022.

25. Windows Firewall Control [Electronic resource]. – Access mode: https://binisoft.org/wfc. – 4.08.2022.

26. Wireshark [Electronic resource]. – Access mode:https://www.wireshark.org/. – 4.08.2022.

27. Kewate, N. A Review on AWS – Cloud Computing Technology [Text] / N. Kewate // International Journal for Research in Applied Science and Engineering Technology. – 2022. – Vol. 10. – P. 258-263. DOI: 10.22214/ijraset.2022.39802.

28. Transparent Forwarders: An Unnoticed Component of the Open DNS Infrastructure / M. Nawrocki, M. Koch, T. Schmidt, M. Wählisch // The International Conference on emerging Networking Experiments and Technologies: Proceedings of CoNEXT'21, ACM, New York, NY, USA. – 2021. – P. 454-462. DOI: 10.1145/3485983.3494872.

29. A Client Based Anomaly Traffic Detection and Blocking Mechanism by Monitoring DNS Name Resolution with User Alerting Feature [Text] / Y. Jin, K. Kakoi, N. Yamai, N. Kitagawa, M. Tomoishi // 2018 International Conference on Cyberworlds (CW). – 2018. – P. 351-356. DOI: 10.1109/CW.2018.00070.

30. Kizza, J. M. Cloud Computing Technology and Security [Text] / J. Kizza // Guide to Computer Network Security. – 2020. – P. 477-502. DOI: 10.1007/978-3-030-38141-7_22.

31. The impact of cloud computing on network security and the risk for organization behaviors [Text] / A. Ouda, A. Yousif, A. Hasan et al. // Webology. – 2022. – Vol. 19, No 1. – P. 195-206. DOI: 10.14704/web/v19i1/web19015.

32. Shahana, P. N. Impact and Implications of Big Data Analytics in Cloud Computing Platforms [Text] / P. Shahana // International Journal for Research in Applied Science and Engineering Technology. – 2022. –

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Vol. 10, Issue 5. – P. 4661-4666. DOI: 10.22214/ijraset.2022.43407.

33. Acquah, A. Managing Digitalization Challenges with Amazon Web Services [Electronic resource] / A. Acquah. – Access mode: https://www.theseus.fi/bitstream/handle/10024/750296/Acquah_masters_thesis_2022.pdf. – 4.08.2022.

34. Мониторинг безопасности облаков [Electronic resource]. – Access mode: https://habr.com/ru/company/cisco/blog/466103. – 4.08.2022.

35. Karagiannis, Ch. Digital Evidence and Cloud Forensics: Contemporary Legal Challenges and the Power of Disposal [Text] / Ch. Karagiannis, K. Vergidis // Information. – 2021. – Vol. 12(5), No. 181. – P. 1-18. DOI: 10.3390/info12050181.

36. Liang, J. Evolution of Firewalls: Toward Securer Network Using Next Generation Firewall [Text] / J. Liang, Y. Kim // IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC-2022). – P. 752-759. DOI: 10.1109/CCWC54503.2022.9720435.

37. Tudosi, A. D. Secure network architecture based on distributed firewalls [Text] / A. Tudosi, D. Balan, A. Potorac // International Conference on Development and Application Systems (DAS-2022). – P. 85-90. DOI: 10.1109/DAS54948.2022.9786092.

38. Tupsamudre, H. New Methods to Hide Windows Scheduled Tasks [Electronic resource]. – Access mode: https://www.researchgate.net/publication/361444433_New_Methods_to_Hide_Windows_Scheduled_Tasks. – 4.08.2022.

References (BSI)

1. *General Data Protection Regulation (EU GDPR)*. Available at: https://gdpr-text.com/en/ (accessed 4.08.2022).

2. Toscano, J. *Data Privacy Issues Are the Root of Our Big Tech Monopoly Dilemma*. Available at: https://www.forbes.com/sites/joetoscano1/2021/12/01/data-privacy-issues-are-the-root-of-our-big-tech-monopoly-dilemma/?sh=cd05cdb3cfd7 (accessed 4.08.2022).

3. Gugelmann, D. On Data and Privacy Leakage in Web Traffic. *Doctoral Thesis*. 2015. 186 p. DOI: 10.3929/ethz-a-010615756

4. Trofymenko, O., Dubovoy, Y. Evolution of glances at information changes in the era of information support. *Comparative analytical law*, 2017, vol. 1, pp. 189-192.

5. Zimmerman, S., Thorpe, A., Fox, C., Kruschwitz, U. Privacy Nudging in Search: Investigating Potential Impacts. *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 2019, pp. 283-287. DOI: 10.1145/3295750.3298952.

6. *Google: a sinister trait*. Available at: https://eurasia.film/2019/08/google-v-tvoej-golove/ (accessed 4.08.2022).

7. Esteve, A. The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. *International Data Privacy Law*, 2017, vol. 7(1), pp. 36-47. DOI: 10.1093/idpl/ipw026.

8. *ISO/IEC 2382:2015 Information technology – Vocabulary*. Available at: https://iso.org/standard/63598.html (accessed 4.08.2022).

9. Kolisnyk, M.Vulnerability analysis and method of selection of communication protocols for information transfer in Internet of Things systems. *Radioelectronic and Computer Systems*, 2021, vol. 1(97), pp. 133-149. DOI: 10.32620/reks.2021.1.12.

10. Bobrovnikova, K., Lysenko, S., Savenko, B., Gaj, P., Savenko, O. Technique for IoT malware detection based on control flow graph analysis. *Radioelectronic and Computer Systems*, 2022, vol. 1(101), pp. 141-153. DOI: 10.32620/reks.2022.1.11.

11. Rivera, S., Gurbani, V.K., Lagraa, S., Iannillo, A.K., State, R. Leveraging eBPF to preserve user privacy for DNS, DoT, and DoH queries. *Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES'20)*, 2020, vol. 78, pp. 1-10. DOI: 10.1145/3407023.3407041.

12. Bumanglag, K., Kettani, H. On the Impact of DNS Over HTTPS Paradigm on Cyber Systems.*3rd International Conference on Information and Computer Technologies (ICICT)*, San Jose, CA, USA, 2020, pp. 494-499. DOI: 10.1109/ICICT50521.2020.00085.

13. Charanjeet, S. How to Enable DNS Over HTTPS in Chrome, Firefox, Edge, Brave & More? *Fossbytes*, 2020. Available at: https://fossbytes.com/how-to-enable-dns-over-https-on-chrome-firefox-edge-brave/ (accessed 4.08.2022).

14. Imana, B., Korolova, A., Heidemann, J.S. Institutional privacy risks in sharing DNS data. *Proceedings of the Applied Networking Research Workshop (ANRW '21),* Virtual Event, USA, 2021, pp. 69-75. DOI: 10.1145/3472305.3472324.

15. Chang, D., Chen, J.Q., Li, Z., Li, X. Hide and Seek: Revisiting DNS-based User Tracking. *2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P)*, Genoa, Italy, 2022, pp. 188-205. DOI: 10.1109/EuroSP53844.2022.00020.

16. Siby, S.D., Juárez, M., Díaz, C., Vallina-Rodriguez, N., Troncoso, C. Encrypted DNS -> Privacy? A Traffic Analysis Perspective. *ArXiv*, 2020, vol. abs/1906.09682, pp. 1-21. DOI: 10.14722/ndss.2020.24301.

17. Liu, Y. Computer Method Research on Risk Control Identification System Based on Deep Learning. *2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, Dalian, China, 2021, pp. 561-565. DOI: 10.1109/AEECA52519.2021.9574442.

18. Gai, V., Ephode, I., Barinov, R., Polyakov, I., Golubenko, V., Andreeva, O. Model and Algorithms for User Identification by Network Traffic. *Proceedings of the 31st International Conference on Computer Graphics and Vision*, 2021, vol. 2, pp. 1-11. DOI: 10.20948/graphicon-2021-3027-1017-1027.

19. Zadereyko, A., Troyanskiy, A., Loginova, N., Trofimenko, E. The implementation of depersonalization algorithm of digital images. *2nd International IEEE Conference Advanced information and communication technologies-2017 (AICT-2017)*, 2017, pp. 56-61.

20. Zadereyko, O., Prokop, Y., Trofymenko, O., Loginova, N., Plachinda, O. Development of an algorithm to protect user communication devices against data leaks. *Eastern-European Journal of Enterprise Technologies*, 2021, vol. 1/2 (109), pp. 24-34. DOI: 10.15587/1729-4061.2021.225339.

21. Zadereyko, O., Trofymenko, O., Loginova, N. Algorithm of user's personal data protection against data leaks in Windows 10 OS. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, Lublin, 2019, vol. 9, pp. 41-44.

22. Natesan, S., Gupta, M., Iyer, L., Sharma, D. Detection of Data Leaks from Android Applications. *Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2020, pp. 326-332. DOI: 10.1109/ICIRCA48905.2020.9183066.

23. Naseri, M., Borges, N., Zeller, A., Rouvoy, R. Accessi Leaks: Investigating Privacy Leaks Exposed by the Android Accessibility Service. *Proceedings on Privacy Enhancing Technologies*, 2019, vol. 2, pp. 291-305. DOI: 10.2478/popets-2019-0031.

24. *NoRoot Firewall*. Available at: https://play.google.com/store/apps/details?hl=en&id=app.greyshirts.firewall (accessed 4.08.2022).

25. *Windows Firewall Control*. Available at: https://binisoft.org/wfc(accessed 4.08.2022).

26. *Wireshark*. Available at: https://www.wireshark.org/(accessed 4.08.2022).

27. Kewate, N. A Review on AWS - Cloud Computing Technology. *International Journal for Research in Applied Science and Engineering Technology*, 2022, vol. 10, pp. 258-263. DOI: 10.22214/ijraset.2022.39802.

28. Nawrocki, M., Koch, M., Schmidt, T., Wählisch, M. Transparent Forwarders: An Unnoticed Component of the Open DNS Infrastructure. *Proceedings of the International Conference on emerging Networking Experiments and Technologies: CoNEXT'21*, ACM, New York, NY, USA, 2021, pp. 454-462. DOI: 10.1145/3485983.3494872.

29. Jin, Y., Kakoi, K., Yamai, N., Kitagawa, N., Tomoishi, M. A Client Based Anomaly Traffic Detection and Blocking Mechanism by Monitoring DNS Name Resolution with User Alerting Feature. *2018 International Conference on Cyberworlds (CW)*, 2018, pp. 351-356. DOI: 10.1109/CW.2018.00070.

30. Kizza, J.M. Cloud Computing Technology and Security. *Guide to Computer Network Security*, 2020, pp. 477-502. DOI: 10.1007/978-3-030-38141-7_22.

31. Ouda, A., Yousif, A., Hasan, A., Hassan M., Shyaa M. The impact of cloud computing on network security and the risk for organization behaviors. *Webology*, 2022, vol. 19, no. 1, pp. 195-206. DOI: 10.14704/web/v19i1/web19015.

32. Shahana, P. Impact and Implications of Big Data Analytics in Cloud Computing Platforms. *International Journal for Research in Applied Science and Engineering Technology*, 2022, vol. 10, no. 5, pp. 4661-4666. DOI: 10.22214/ijraset.2022.43407.

33. Acquah, A. *Managing Digitalization Challenges with Amazon Web Services*. Available at: https://www.theseus.fi/bitstream/handle/10024/750296/Acquah_masters_thesis_2022.pdf (accessed 4.08.2022).

34. *Cloud security monitoring*. Available at: https://habr.com/ru/company/cisco/blog/466103/ (accessed 4.08.2022).

35. Karagiannis, Ch., Vergidis, K. Digital Evidence and Cloud Forensics: Contemporary Legal Challenges and the Power of Disposal. *Information*, 2021, vol. 12(5), no. 181, pp. 1-18. DOI: 10.3390/info12050181.

36. Liang, J., Kim, Y. Evolution of Firewalls: Toward Securer Network Using Next Generation Firewall. *IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC-2022*), pp. 0752-0759. DOI: 10.1109/CCWC54503.2022.9720435.

37. Tudosi, A., Balan, D., Potorac, A. Secure network architecture based on distributed firewalls. *International Conference on Development and Application Systems (DAS-2022)*, pp. 85-90. DOI: 10.1109/DAS54948.2022.9786092.

38. Tupsamudre, H. *New Methods to Hide Windows Scheduled Tasks*. Available at: https://www.researchgate.net/publication/361444433_New_Methods_to_Hide_Windows_Scheduled_Tasks (accessed 4.08.2022).

Надійшла до редакції 25.08.2022, розглянута на редколегії 20.11.2022

ДОСЛІДЖЕННЯ ПОТЕНЦІЙНИХ ВИТОКІВ ДАНИХ В ІНФОРМАЦІЙНИХ І КОМУНІКАЦІЙНИХ СИСТЕМАХ

Олександр Задерейко, Олена Трофименко, Юлія Прокоп, Наталія Логінова, Анастасія Дика, Сергій Кухаренко

Розглянуто питання забезпечення захисту користувачів інформаційних систем, представлених стаціонарними і мобільними пристроями комунікації та орієнтованих на взаємодію з цифровим простором Інтер-

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

нет. Встановлено, що провідні технологічні IT-корпорації збір збирають дані із пристроїв комунікації користувачів. Показано, що організація збору даних здійснюється шляхом перенаправлення DNS-трафіку пристрою комунікації на DNS-сервери IT-корпорацій, з подальшим його шифруванням по DoH протоколу. Це унеможливлює контроль за DNS-трафіком користувачів з боку уповноважених служб та відомств держав і забезпечує монопольне становище IT-корпорацій на глобальному цифровому ринку щодо збору та аналізу даних користувачів. Показано, що збір даних користувачів здійснюється з метою їхньої подальшої монетизації та впливу на рішення, які приймаються користувачами. Здійснено фіксацію DNS-трафіку пристроїв комунікації з цифровим простором Інтернет. Виконано аудит зафіксованого DNS-трафіку, в результаті якого виявлено спеціалізовані інтернет-ресурси, які відповідають за збір та оброблення даних користувачів. Доведено, що виявлені спеціалізовані інтернет-ресурси належать IT-корпораціям. Розглянуто методи ідентифікації пристроїв комунікації у цифровому просторі. Показано, що ідентифікація пристроїв комунікації базується на зборі унікального набору даних з кожного окремо взятого пристрою комунікації. На основі кожного унікального набору даних формується цифровий відбиток пристрою комунікації, який використовується для подальшої ідентифікації в цифровому просторі. Визначено підходи, що дозволяють організувати захист від збору даних користувачів в інформаційних системах. Запропоновано програмну та апаратну реалізацію захисту від збору даних з пристроїв комунікації. Експериментальним шляхом встановлено, що спільне використання запропонованого програмного та апаратного захисту є найбільш ефективним захистом від збору даних із пристроїв комунікації і не впливає на функціональність інформаційних систем.

Ключові слова: витоки даних; цифровий простір; DNS-запити; DNS-сервери; комунікаційний пристрій; захист даних.

Задерейко Олександр Владиславович – канд. техн. наук, доц., доц. каф. інформаційних технологій, Національний університет «Одеська юридична академія», Одеса, Україна.

Трофименко Олена Григорівна – канд. техн. наук, доц., доц. каф. інформаційних технологій, Національний університет «Одеська юридична академія», Одеса, Україна.

Прокоп Юлія Віталіївна – канд. іст. наук, доц., викл. каф. комп'ютерних наук, Чеський технічний університет у Празі, Прага, Чеська Республіка.

Логінова Наталія Іванівна – канд. пед. наук, зав. каф. інформаційних технологій, Національний університет «Одеська юридична академія», Одеса, Україна.

Дика Анастасія Іванівна – асист. каф. інформаційних технологій, Національний університет «Одеська юридична академія», Одеса, Україна.

Кухаренко Сергій Вікторович – канд. техн. наук, доц. каф. кібербезпеки, Національний університет «Одеська юридична академія», Одеса, Україна.

Olexander Zadereyko – PhD, Associate professor of Information Technology Department, National University "Odessa Law Academy", Odessa, Ukraine,
e-mail: zadereyko@onua.edu.ua, ORCID: 0000-0003-0497-9861.

Olena Trofymenko – PhD, Associate professor of Information Technology Department, National University "Odessa Law Academy", Odessa, Ukraine,
e-mail: egt@ukr.net, ORCID: 0000-0001-7626-0886.

Yuliia Prokop – PhD, Lecturer of Computer Science Department, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic,
e-mail: prokoyul@fel.cvut.cz, ORCID: 0000-0002-6608-3668.

Nataliia Loginova – PhD, Head of Information Technology Department, National University "Odessa Law Academy", Odessa, Ukraine,
e-mail: loginova@onua.edu.ua, ORCID: 0000-0002-9475-6188.

Anastasiia Dyka – Assistant of Information Technologies Department, National University "Odessa Academy of Law", Odessa, Ukraine,
e-mail: dyka.anastasiia@gmail.com, ORCID: 0000-0002-4196-8734.

Serhii Kukharenko – PhD, Associate professor of Cyber Security Department, National University "Odessa Law Academy", Odesa, Ukraine,
e-mail: skuharenko@ukr.net, ORCID: 0000-0001-7100-6408.