UDC 004.738.2
DOI: 10.15587/1729-4061.2021.225339

DEVELOPMENT OF AN ALGORITHM TO PROTECT USER COMMUNICATION DEVICES AGAINST DATA LEAKS

A. Zadereyko
PhD, Associate Professor*
E-mail: zadereyko@onua.edu.ua

Y. Prokop
Senior Lecturer
Department of Information Technology
O. S. Popov Odessa National Academy of Telecommunications
Kuznechna str., 1, Odessa, Ukraine, 65029
E-mail: yulia13.prokop@gmail.com

O. Trofymenko
PhD, Associate Professor*
E-mail: egt@ukr.net

N. Loginova
PhD, Associate Professor*
E-mail: loginova@onua.edu.ua

O. Plachinda
PhD, Associate Professor
Department of Oil and Gas and Chemical Engineering
Odessa National Polytechnic University
Shevchenko ave., 1, Odessa, Ukraine, 65044
E-mail: olga_plach2017@ukr.net

*Department of Information Technology
National University "Odessa Law Academy"
Fontanska doroha str., 23, Odessa, Ukraine, 65009

In order to identify ways used to collect data from user communication devices, an analysis of the interaction between DNS customers and the Internet name domain space has been carried out. It has been established that the communication device's DNS traffic is logged by the DNS servers of the provider, which poses a threat to the privacy of users. A comprehensive algorithm of protection against the collection of user data, consisting of two modules, has been developed and tested. The first module makes it possible to redirect the communication device's DNS traffic through DNS proxy servers with a predefined anonymity class based on the proposed multitest. To ensure a smooth and sustainable connection, the module automatically connects to a DNS proxy server that has minimal response time from those available in the compiled list. The second module blocks the acquisition of data collected by the developers of the software installed on the user's communication device, as well as by specialized Internet services owned by IT companies. The proposed algorithm makes it possible for users to choose their preferred level of privacy when communicating with the Internet space, thereby providing them with a choice of privacy level and, as a result, limiting the possibility of information manipulation over their owners. The DNS traffic of various fixed and mobile communication devices has been audited. The analysis of DNS traffic has enabled to identify and structure the DNS requests responsible for collecting data from users by the Internet services owned by IT companies. The identified DNS queries have been blocked; it has been experimentally confirmed that the performance of the basic and application software on communication devices was not compromised

Keywords: DNS query, DNS server, DNS leaks, DNS traffic, DNS proxy server, data collection

Received date 17.12.2020
Accepted date 08.02.2021
Published date 26.02.2021

1. Introduction

In today's Internet space, huge amounts of information are circulating, most of which is user-sharing data as a result of their interaction with various Internet services. Structuring and analyzing these data make it possible to identify seemingly hidden patterns, to predict, and with a system approach to form behavioral trends of the Internet audience.

This situation is exacerbated by the energetic efforts of high-tech IT companies to introduce user digital data collection and analysis systems, which leads to the unspoken monopolization of the market of users' digital data. At the same time, the regulatory role of various state institutions in respecting the rights to privacy of users, namely, the secrecy of correspondence and activities in the Internet space, is steadily decreasing. The growing trend is of increasing concern to Internet users, IT companies' employees, and non-governmental organizations. They draw attention to the inadmissibility of unauthorized collection and monetization of data without any consent of users of the Internet space [1]. Requirements for the implementation of measures to increase user privacy are regulated by IISO/IEC 24760-1:2019 (E) IT Security and Privacy.

In this regard, it is a relevant task to undertake research aimed at developing new approaches and developing tools to protect users' data on the Internet. Users have the right not only to know what information about them can be collected by Internet services but also to have the opportunity to choose the level of privacy in the Internet space. Moreover, the development of tools to control and manage privacy in order to prevent unwanted processing of personal identification information is predetermined by the standard ISO/IEC 29100:2011.

2. Literature review and problem statement

Work [2] reports the results of a study showing that both government and IT companies are interested in collecting user data in the Internet space. Data collected while tracking Domain Name System (DNS) requests by the MoreCowBell subsystem as part of the PRISM project were involved in the management of various public processes [3]. IT companies are very interested in collecting and analyzing user data, trying to monetize their ads as efficiently as possible [4]. In addition, various IT companies collect statistics about the use of Internet resources and process them automatically [5]. That confirms the hypothesis that information about their actions is tracked and collected from communication devices connected to the Internet without the knowledge of users. However, the cited works do not offer ways to protect users' data that could exclude monitoring by IT companies.

The most promising direction in terms of ensuring the maximum possible accuracy of collecting information about user actions in the Internet space is the analysis of DNS traffic of DNS clients installed on communication devices [6]. Paper [7] shows that the DNS traffic analysis can identify software installed on communication devices. One can also obtain data on the history of geolocation, accounting records in Internet services, interests, religious preferences, financial status, medical needs, etc. The result of this analysis is the creation of a database of unique digital profiles of communication devices [8, 9], and, as a result, accurate prediction of the behavioral response of the Internet audience and the development of possible scenarios for influencing its behavior [10]. That causes users to be insecure from monitoring their network traffic and makes it impossible to choose their level of privacy when communicating with the Internet space.

Researchers studied ways to prevent user data leaks. For example, work [11] identified and generalized the data collected by the Windows family operating systems on the user's communication devices, sent to Microsoft servers. This trend naturally leads to the search for solutions that give users the choice of what data they can access in the Internet space. In particular, article [12] presents the interface URetail in the form of radar, allowing the user to choose which of his/her personal data can be disclosed. However, the implementation of this approach is narrowly focused on the data collected in retail when shopping in online stores.

Many scientists have been involved in the development of methods for analyzing DNS traffic, issues of its encryption in order to protect users' DNS requests from monitoring and censorship. For example, the authors of paper [13] have concluded that the existing standard DNS traffic schemes are ineffective. Works [14–17] emphasize the relevance of DNS traffic protection and point to the need for a thorough analysis of possible leaks. For example, article [14] explores the principles of DNS operation and analyzes Namecoin, GNU, and RAINS systems. Work [15] looks at the vulnerabilities of the DNS protocol and how malicious software exploits these vulnerabilities. Study [16] identified the problem of DNS privacy leakage and analyzed the use of HTTPS/TLS (DoH/DoT) and SNI (ESNI) encryption technologies. The DNS traffic leaks were evaluated in [17]. Papers [18–20] analyze the pros and cons of DNS traffic encryption using DNS over TLS (DoT) protocols, DNS over HTTPS (DoH). Study [18] found that even when encryption is enabled, users' data outflow through their DNS queries. In addition, it was found in [19] that doT and DoH protocols are supported by only a small number of DNS servers. Significantly, encryption requires additional computing resources and slows down the processing of DNS queries [20]. Work [21] analyzes the vulnerabilities of the DoT protocol. Article [22] explores the performance of the DoH protocol and the impact of DNS traffic encryption protocols on Internet space participants. However, the issue of implementing measures to increase the privacy of users when communicating on the Internet remains unresolved.

Ways to increase user privacy are discussed in works [23, 24], which propose the introduction of filtering network traffic of communication devices. However, packet filtering, due to its specificity and the peculiarities of individual protocols to which filters are applied, is not a sufficient means to ensure the protection of user data. Network traffic filtering can be used as one of the means of blocking incoming and outgoing IP packets.

Redirecting traffic through an additional intermediate DNS server, implemented between the DNS client and the remote DNS server, is proposed in [25, 26]. Thus, the idea of using Smart DNS Proxy Server is considered in [25] to gain access to Internet resources to sites that are not available due to geographical constraints. Study [26] considers building the architecture of the network service that functions as UDP Proxy. However, filtering and cryptographic transformation mechanisms are not used to protect DNS requests from monitoring by Internet providers.

The systematic results of the above papers suggest that there is an insufficient study of how data are collected from user communication devices when DNS clients interact with the domain namespace. All this allows us to argue that it is appropriate to conduct a study on the development of tools that can simultaneously localize DNS traffic leaks, hide the actual IP address of the communication device, and block the collection of user data.

3. The aim and objectives of the study

The aim of this study is to develop an algorithm to protect communication devices from unauthorized collection and leakage of user data on the Internet. The practical application of the developed algorithm would give users the opportunity to determine the level of their privacy.

To accomplish the aim, the following tasks have been set:

– to analyze the process of data sharing between DNS customers and the Internet services they interact with to identify leaks and ways to collect data from users' communication devices;

– to develop an algorithm to block data leaks collected by developers of the software installed on a communication device to enable users to choose their privacy when interacting with various Internet services;

– to audit the TCP/UDP traffic of various communication devices in order to identify services that send requests for user data collection;

– to check the proposed algorithm for the absence of DNS traffic leaks from the communication device.

4. Exploring the process of data exchange between DNS customers and Internet services

The physical connection of the user's communication device to the Internet space and its subsequent access to Internet resources begins with DNS sending requests to various Internet services. At the same time, any software installed on users' communication devices, such as web browsers, file managers, email clients, messengers, etc., which execute DNS requests, can act as a client. DNS customers interact with the Internet space and process DNS queries in the domain space in a strictly defined order [27].

In practice, to reduce response times to a DNS query and reduce the load on the server's root DNS, providers create their own DNS server cache [28]. If the DNS request previously recorded in the DNS server cache is met, an IP address is issued (Fig. 1).

Thus, all queries from DNS customers are accumulated in the DNS logs of the provider's server. Structuring and analyzing DNS query data can provide comprehensive information about a user's online activities. Various state security structures, advertising and analytical units of IT companies, as well as representatives of organized cybercrime, are becoming increasingly interested in their collection, storage, and analysis. That is why user data is increasingly referred to as "digital gold".

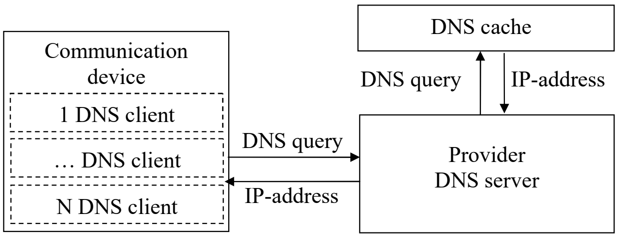

Fig. 1. DNS query caching scheme on provider's DNS server

Given the above, a scheme for data exchange of a communication device with the Internet space is proposed (Fig. 2). Its analysis leads to the conclusion that the ultimate beneficiaries of user data, one way or another, are IT-companies.

The organization of mass scale and continuity of the process of data collection from communication devices is achieved by IT companies' introduction of free access to internet statistics collection and analysis services: Google Analytics, Yandex Metrika, Liveinternet, Rambler, etc. This approach allows IT companies, introducing systems for automated processing of collected data, to carry out not only digital profiling of communication devices but also to create unique digital profiles for each of their real users [8, 9].

Not surprisingly, this trend is a concern for the leadership of a number of democratic countries. For example, EU countries at the legislative level have tightened control and responsibility for infringements on the personal data of EU citizens on its territory and beyond, adopting the GDPR (General Data Protection Regulation) Act [29]. However, even these strict measures do not, in fact, solve the main problem. They do not give users the ability to determine their own level of privacy by managing the collection of their data while doing any actions in the Internet space in real-time.

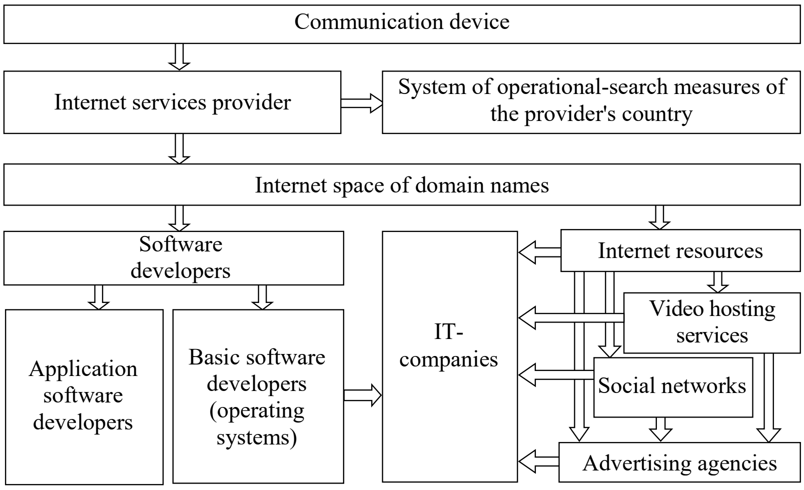
Communication device
Internet services provider
System of operational-search measures of the provider's country
Internet space of domain names
Software developers
Application software developers
Basic software developers (operating systems)
IT-companies
Internet resources
Video hosting services
Social networks
Advertising agencies

Fig. 2. The scheme of communication between a communication device and the Internet space

5. Developing an algorithm to block data leaks from the user's communication device

The set of measures to prevent data collection from communication devices, and therefore reduce the likelihood of their digital profiling, includes two modules:

1. DNS traffic leakage protection module by:

– sending DNS queries under the DoH protocol;

– redirecting DNS traffic to a DNS proxy server with a predefined level of privacy.

2. The data collection lock module by:

– locking dataset plugins integrated into the Content Management System (CMS) of online resources;

– blocking the DNS traffic of system-wide and application software.

The first DNS traffic leak protection module is key. This is due to the fact that Internet providers connecting users to the Internet domain space perform it through DNS servers controlled by them, keeping mandatory logs of records of DNS requests of each user. It is obvious that ISPs can:

– link each user's IP address to all the domain names they've been asked for;

– store the accumulated data indefinitely;

– to provide the accumulated data to authorized government agencies.

Thus, the users cannot be sure of their privacy by conducting Internet communication through the provider's DNS server.

In addition, ISPs by default set their users a mode of forced connection to their DNS server if the user changes the settings to use a third-party DNS server. If such DNS settings are found in a communication device, ISPs use a transparent DNS proxy that redirects user traffic to DNS. Thus, the provider is masking the real route of the user's DNS traffic. This technique makes it possible to secure the

DNS user requests to the DNS provider's server and continue to log its DNS traffic.

Another important factor in controlling user traffic for DNS is that the default DNS protocol does not encrypt DNS queries. Attempts to implement DNS traffic cryptographic encryption have been reflected in the development and implementation of DNScrypt, DoT, and DoH protocols. These protocols encrypt DNS traffic, creating a cryptographically secure channel between DNS customers and servers. It was this circumstance that prompted IT companies to declare support for implemented DNS traffic encryption technologies and to create the same public DNS servers controlled by them with the support of DNSCrypt, DoT, DoH protocols (Table 1).

The number of IT companies supporting DNS traffic encryption using these cryptographic protocols continues to increase, which unequivocally allows the following:

– to counter DNS substitution of responses at DNS transit hubs;

– to bypass the blocking (censorship) of DNS traffic by providers;

– to make it impossible to log and then inspect DNS traffic;

– to reduce the role of providers connecting communication devices to the Internet space;

– to reduce the role of root DNS servers;

– to redistribute data collection on DNS users' traffic in favor of large IT companies.

In addition, most web browser developers have not only implemented the DoH protocol in their software products but have also implemented the ability to connect to the public DNS servers of leading IT companies [30]. The tendency to monopolize DNS traffic by IT companies significantly stresses the urgency of the issue of ensuring the real privacy of users, as it is these IT companies that own the services of collecting and analyzing Internet statistics. Examples of such services are Google Analytics, Yandex Metrika, Liveinternet, Rambler TOP, etc. In addition, it cannot be ruled out that IT companies may provide third parties or authorized government agencies with access to the DNS traffic history of users who have used the services of public DNS servers.

To ensure the privacy of Internet users, along with the use of the specified DNS traffic encryption technologies, it is suggested that DNS requests be redirected through DNS proxy servers of a different class of anonymity. These DNS proxies must have a fixed lifespan and should not log DNS requests. The advantage of this approach makes it possible to exclude the possibility of accumulating data on DNS user requests not only from providers and authorized government agencies but also from IT companies.

However, the use of DNS proxy servers that support the DoH protocol is a prerequisite to ensure that DNS requests are secure. And to ensure the highest user privacy, DNS traffic from the communication device should be redirected through the HIA (High anonymous) proxy DNS. These proxy servers hide the actual IP address of the DNS client and prevent the requested DNS server from determining the use of DNS proxies [31, 32].

A scheme is proposed to redirect DNS communication device traffic through a DNS proxy server (Fig. 3).

The redirection of DNS queries from a communication device is executed by changing the route of DNS requests from the DNS client to the requested domain. Its distinctive features are:

– creating a local DNS server;

– redirecting DNS queries of DNS clients from a communication device to a local DNS server;

– redirecting DNS queries from a local DNS server to a pre-selected DNS proxy server using the proposed algorithm (Fig. 4).

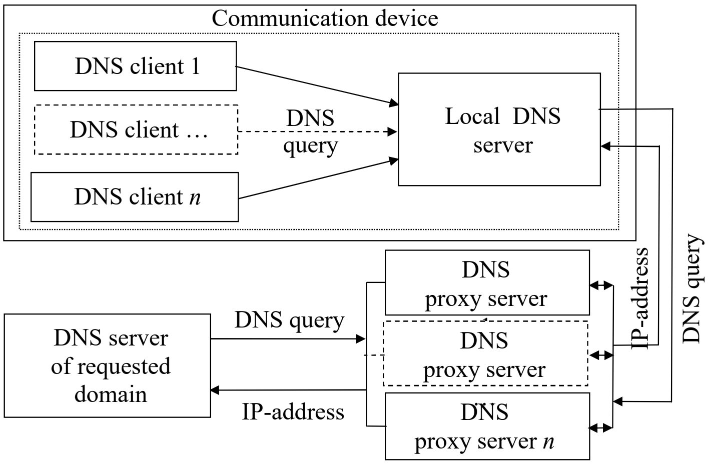

Fig. 3. Redirecting DNS traffic through a DNS proxy server

Table 1

Public DNS servers of IT-companies that support DNSCrypt, DoT, DoH

IT corporations' public DNS servers	IP address mask		Protocol support (+/–)		
	IPv4	IPv6	DNSCrypt	DoT	DoH
Cloudflare	1.1.1.1 1.0.0.1	2606:4700:4700::1111 2606:4700:4700::1001	–	+	+
Google Public DNS	8.8.8.8 8.8.4.4	2001:4860:4860::8888 2001:4860:4860::8844	–	+	+
Quad9	9.9.9.9 149.112.112.112	2620:fe::fe 2620:fe::9	+	+	+
CleanBrowsing	185.228.168.168 185.228.169.168	2a0d:2a00:1:: 2a0d:2a00:2::	+	+	+
Adguard	176.103.130.130 176.103.130.131	2a00:5a60::ad1:0ff 2a00:5a60::ad2:0ff	+	+	+
Cisco OpenDNS	208.67.222.222 208.67.220.220	2620:119:35::35 2620:119:53::53	+	+	+

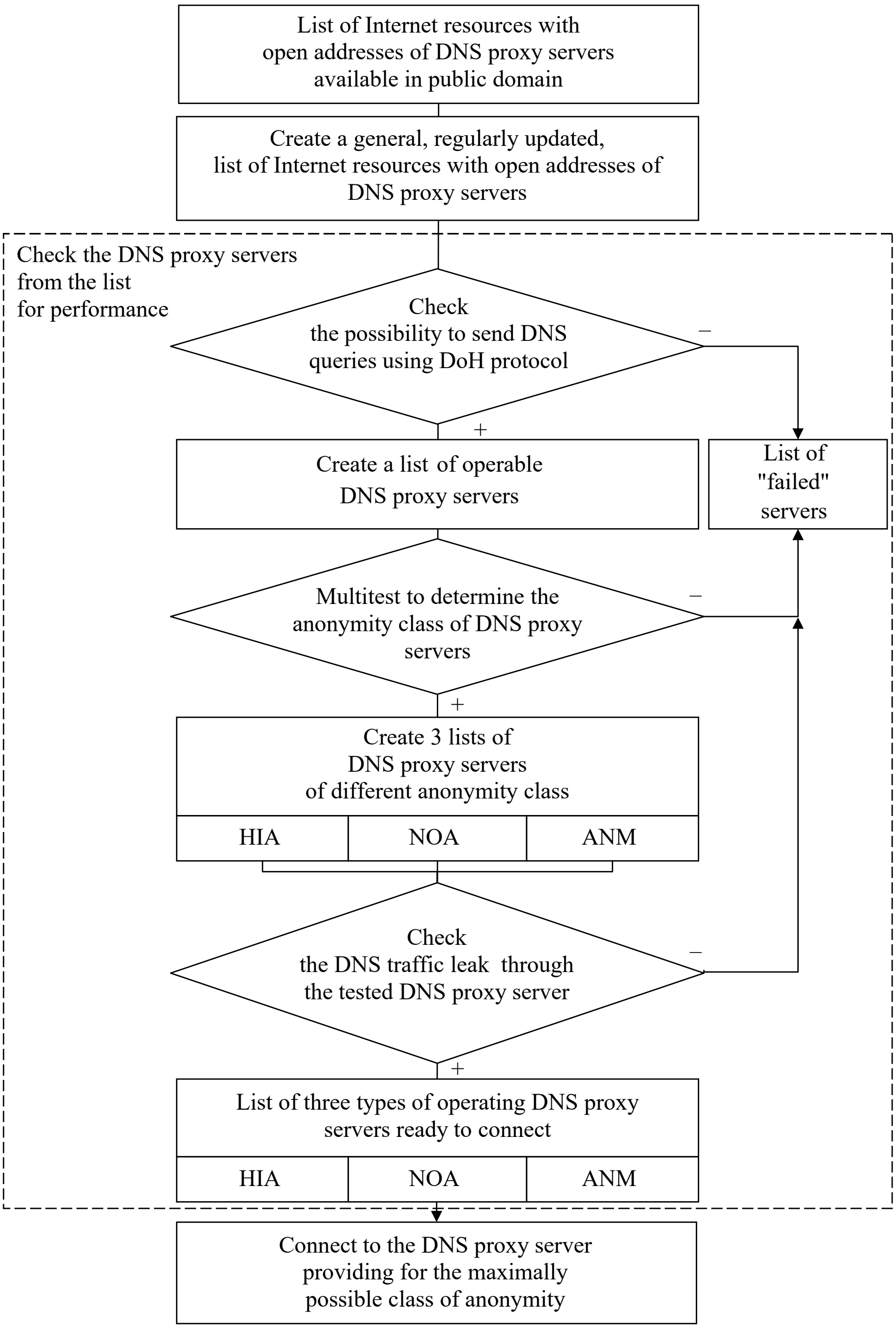
List of Internet resources with open addresses of DNS proxy servers available in public domain
Create a general, regularly updated, list of Internet resources with open addresses of DNS proxy servers
Check the DNS proxy servers from the list for performance
Check the possibility to send DNS queries using DoH protocol
–
+
Create a list of operable DNS proxy servers
List of "failed" servers
Multitest to determine the anonymity class of DNS proxy servers
–
+
Create 3 lists of DNS proxy servers of different anonymity class
HIA
NOA
ANM
Check the DNS traffic leak through the tested DNS proxy server
–
+
List of three types of operating DNS proxy servers ready to connect
HIA
NOA
ANM
Connect to the DNS proxy server providing for the maximally possible class of anonymity

Fig. 4. DNS proxy server testing algorithm

To implement the proposed DNS proxy server selection algorithm, the following is required:

1) search for online resources with open lists of DNS proxy servers that are publicly available;

2) form a common list of Internet resources with open lists of DNS proxy servers;

3) create a regularly updated list of DNS proxy servers for later testing:

3. 1) check if DNS queries can be sent under the HTTPS protocol. To create a list of operating DNS proxy servers that meet the specified requirements of anonymity, one needs to implement the process of checking them (testing) for performance. To this end, one needs to consistently execute DNS requests through each DNS proxy server of the following form:

– https://ajax.googleapis.com/ajax/libs/jquerymobile/1.4.5/jquery.mobile.min.css;

– https://cdnjs.cloudflare.com/ajax/libs/twitter-bootstrap/3.3.7/css/bootstrap-theme.css;

– https://maxcdn.bootstrapcdn.com/font-awesome/4.3.0/css/font-awesome.min.css?ver=4.9.8;

3. 2) move to the list of operating DNS proxy servers those that have executed DNS requests from p. 3. 1;

3. 3) conduct a multi-test on determining a DNS proxy server anonymity class:

3. 3. 1) check the presence of HTTP-headers: HTTP_VIA, HTTP_X_FORWARDED_FOR, HTTP_FORWARDED_FOR, HTTP_X_FORWARDED, HTTP_FORWARDED, HTTP_CLIENT_IP, HTTP_FORWARDED_FOR_IP, VIA, X_FORWARDED_FOR, FORWARDED_FOR, X_FORWARDED, FORWARDED, CLIENT_IP, FORWARDED_FOR_IP, HTTP_PROXY_CONNECTION;

3. 3. 2) check the presence of open ports HTTP proxy: 3128, 1080, 8123, 8000, 1080;

3. 3. 3) check the presence of open web proxy ports: 80, 8080;

3. 3. 4) check a DNS proxy server for the presence of names: vpn, hide, hidden, proxy (suspicious host name);

3. 3. 5) check the difference in time zones between the IP address of the communication device and the DNS proxy IP address;

3. 3. 6) check the DNS proxy server IP address for the Tor network;

3. 3. 7) check the use of traffic compression services from Google, Yandex, and Opera by comparing the IP address pool of these companies' services with the DNS proxy IP address (Turbo mode);

3. 3. 8) check the DNS proxy server for the presence of a redirection of the communication device by comparing the host content received from window.location.hostname with the host content of the requested Internet resource (JavaScript method);

3. 3. 9) check the sending of IP address of the communication device bypassing the DNS proxy server (leaked IP address via Flash);

3. 3. 10) determine the duration of DNS requests routing in milliseconds (routing duration of more than 30 milliseconds is considered as the presence of a DNS proxy server (bilateral ping);

3. 3. 11) check the leaks of IP address of the communication device via WebRTC;

3. 3. 12) check the DNS proxy server for the VPN technology: analysis is performed on the size of the intercepted MTU packet and the maximum volume of MSS data in the packet transmitted (VPN fingerprint);

3. 4) distribute working DNS proxy servers by the anonymity class based on the results of the multitest in p. 3. 3:

a) Not anonymous, not hiding the real IP address of a DNS client;

b) Anonymous ANM (Anonymous) that hides the real IP address of a DNS client but allows the requested DNS server to determine the use of DNS proxy servers;

c) High anonymity HIA (High anonymous) that hides the IP address of a DNS client and does not make it possible for the requested DNS server to determine the use of DNS proxy servers;

3. 5) connect to a DNS proxy server that provides the highest possible class of anonymity.

The criteria for distributing DNS proxy servers based on a multitest's results (p. 3. 3) for the assignment of an anonymity class are listed in Table 2.

Table 2

Criteria for assigning an anonymity class to DNS proxy servers

DNS proxy server testing criterion	DNS proxy server anonymity class		
	HIA	ANM	NOA
HTTPS connection	yes	yes	yes
HTTP proxy headers	no	no	yes
HTTP proxy open ports	no	yes	yes
Web proxy open ports	no	no	yes
VPN open ports	no	no	yes
Suspicious host name	no	no	yes
Time zone difference (between a communication device and the IP address of the DNS proxy server)	no	no	yes
IP belonging to the Tor net	no	no	no
Turbo browser mode	no	no	yes
IP belonging to the hosting provider	no	no	no
Check a web proxy by the Java Script method	no	yes	yes
IP leak thru Flash	no	no	yes
Channel identification (two-way ping)	no	yes	yes
VPN fingerprint	no	yes	yes
IP leak thru WebRTC	no	yes	yes

The second module of the user's data collection lock algorithm blocks connections between DNS customers of the communication device and specialized Internet data collection services. In addition, it blocks connections to third-party services and services of system and application software developers (Fig. 5).

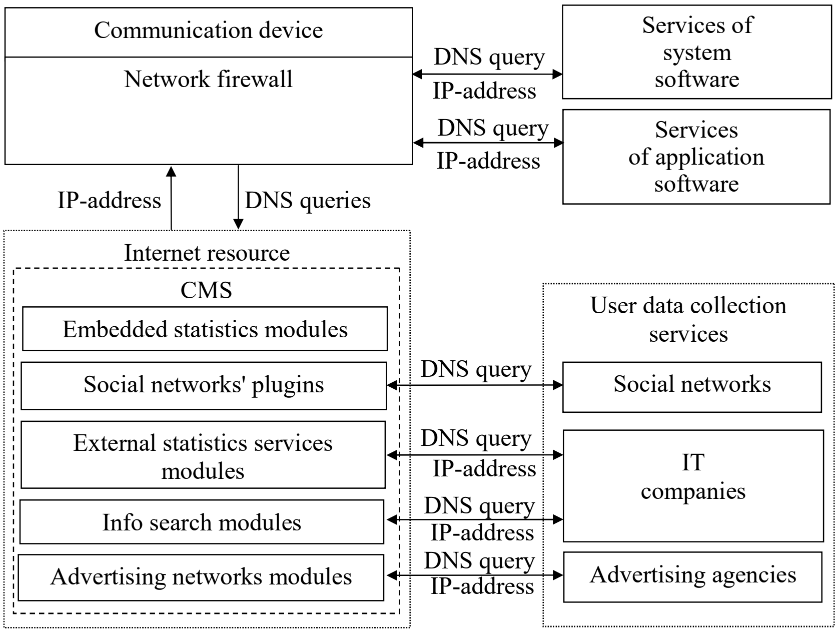

Fig. 5. Sharing data between a communication device and Internet space

This is executed by organizing the TCP/UDP traffic filtration process, which is responsible for communicating with Internet services:

– collecting user data;

– system software;

– application software.

In practice, firewalls are used for filtering, capable of working at the network packet level and ensuring that all incoming and outgoing DNS communications device requests that match the following are blocked:

– IP addresses of user data collection services;

– IP addresses of service and third-party traffic system and application software.

6. TCP/UDP traffic audit results

Table 3 gives the results of comprehensive monitoring of stationary and mobile TCP/UDP traffic from communication devices over a long time.

Table 3

User data collection, analysis, and monetization services

IT-companies	Software	Internet connection			
		Domain name	IP-address	Protocol	
				TCP	UDP
1	2	3	4	5	6
Yandex	Monetization service «Yandex Direct»	an.yandex.ru	93.158.134.90		+
			77.88.21.90		
			213.180.204.90		
			87.250.250.90		
			213.180.193.90		
	Data collection service «Yandex Metrika»	ya.ru	87.250.250.242		+
		yastatic.net	178.154.131.215		+
			178.154.131.216		
			178.154.131.217		
Google	Monetization service «AdSense»	googletagservices.com	216.239.38.10		+
		googletagmanager.com	216.58.208.200		+
		partner.googleadservices.com	172.217.8.2		+
		googlesyndication.com	216.58.215.100		+
		pagead2.googlesyndication.com	172.217.13.66		+
		www.googletagservices.com	216.239.36.10		+
	Data collection service «Analytics»	www-google-analytics.l.google.com	172.217.2.110		+
		google-analytics.com	216.58.208.196		+
		tpc.googlesyndication.com	172.217.13.65		+
		googleads.g.doubleclick.net	142.250.73.194		+
		adservice.google.com	172.217.20.2		+
	OS Android	android.clients.google.com	172.217.19.110		+
			172.217.20.14		+
			172.217.16.110		+
			172.217.18.78		+
Liveinternet	Data collection service	counter.yadro.ru	88.212.201.210		+
			88.212.201.216		+
			88.212.201.198		+
			88.212.201.204		+
			88.212.202.52		+
Microsoft	OC Windows	teredo.ipv6.microsoft.com	40.90.4.4		+
Application software for communication devices					
Telegram Messenger LLP	Telegram	1e100.net	216.239.32.10		+
			216.239.36.10		+
			216.239.38.10		+
			216.239.34.10		+
		dns.google	8.8.8.8		+
		cloudflare.com	104.16.248.249		+

Continuation of Table 3

1	2	3	4	5	6
Rakuten	Viber	1e100.net & googleusercontent.com	216.239.32.10		+
			216.239.36.10		+
			216.239.38.10		+
			216.239.34.10		+
		cloudfront.net	205.251.197.26		+
			205.251.198.61		+
			205.251.193.162		+
			205.251.194.154		+
		eu-central-1.amazonaws.com	205.251.192.27		+
			205.251.195.199		+
			156.154.64.10		+
			156.154.65.10		+
Facebook	Facebook	1e100.net	216.239.32.10		+
			216.239.36.10		+
			216.239.38.10		+
		cloudfront.net	205.251.197.26		+
			205.251.198.61		+
			205.251.193.162		+
			205.251.194.154		+
		te.net.net	199.59.242.153	+	
		host.hit.gemius.pl	91.221.127.226	+	
			178.33.54.6	+	
			81.0.212.193	+	
	Instagram	1e100.net	216.239.32.10		+
			216.239.36.10		+
			216.239.38.10		+
	WhatsApp	1e100.net	216.239.32.10		+
			216.239.36.10		+
			216.239.38.10		+
		static.sl-reverse.com	67.228.254.4	+	

Table 3 gives the identified domain names and IP addresses of system and application software, Internet services of data collection, analysis, and monetization, which establish a connection to the communication device. They are arranged in accordance with the affiliation of IT companies.

Our analysis of DNS traffic related to the system and application software has made it possible to establish those domains among the Internet resources that are accessed by system and application software (Table 2). Domain data were obtained from open sources:

– the sl-reverse.com domain is owned by CSC Digital Brand Services, an IT company specializing in digital brand management and digital marketing;

– the cloudfront.net domain is owned by Amazon, an IT company that specializes in providing a wide range of services in cloud services based on DNS traffic analysis;

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

– the domain host.hit.gemius.pl is owned by Gemius, an IT company that does media research and develops tools used to optimize advertising campaigns;

– the 1e100.net domain is owned by Google's IT company;

– the compute-1.amazonaws.com and eu-central-1.amazonaws.com domains are owned by the Amazon IT company.

The data related to domain owners (Table 3) suggest that mobile application software such as Facebook, Instagram, Viber, and Telegram establishes connections to Internet services owned by the IT companies Google, Amazon, and Cloudflare.

To ensure user privacy, all connections to IP addresses listed in Table 3 should be blocked, which is determined by the functionality of the second module of the proposed algorithm.

7. Discussion of results of applying the algorithm that determines the absence of DNS traffic leaks from a communication device

We have proposed a data-sharing scheme between communication devices and Internet space (Fig. 2), which helped establish that DNS customer requests are accumulating in the DNS logs of the provider's server. After structuring and analyzing DNS queries, DNS logs can be used by various government security agencies, advertising and analytics units at IT companies, as well as organized cybercrime, to obtain private information about users.

The proposed algorithm for blocking data leaks from the user's communication device consists of two modules – the DNS traffic leakage protection module and the data collection lock module. The first module sends DNS requests using the DoH protocol and redirects DNS traffic to a DNS proxy server with a predefined anonymity class. The second module blocks data collection plugins integrated into the Content Management System (CMS) of Internet resources and blocks third-party TCP/UDP traffic from system and application software. Our analysis of the public DNS servers of IT-companies that supported the implementation of DNScrypt, DoT, and DoH protocols (Table 1) revealed that IT companies can counteract the substitution of DNS responses at DNS transit nodes and bypass DNS traffic blocking by providers. In addition, the inability to log and then inspect DNS traffic reduced the role of providers connecting communication devices to the Internet space. A significant feature in the redistribution of DNS user traffic is the decreased role of root DNS servers. As a result of the verification of the developed algorithm, it is proposed to redirect DNS traffic through DNS proxy servers of different classes of anonymity (Fig. 3). That has made it possible to exclude the possibility of accumulating DNS user requests from providers. The advantage of the proposed algorithm is to change the route of DNS queries from a DNS client to the pre-selected DNS proxy server with the highest possible class of anonymity (Fig. 4). The DNS proxy server class of anonymity is determined by applying a devised multi-test to meet the testing criteria (Table 2). The second module of the developed algorithm blocks connections between DNS communication device customers and specialized Internet data collection services. Connections to third-party services and services of system and application software developers (Fig. 5) are also blocked. The combination of the two modules of the proposed algorithm has allowed users to choose the level of their privacy when interacting with the Internet space.

Our comprehensive TCP/UDP audit of the traffic from various communication devices has revealed the IT companies' services involved in user's data collection (Table 3).

The proposed algorithm has been checked for the absence of DNS traffic leaks from a communication device. Its results showed no DNS traffic leaks when using an arbitrarily selected HIA class DNS proxy server (Table 4).

Thus, the task formulated for this study was solved with the help of the developed algorithm to protect communication devices from unauthorized collection and leakage of user data on the Internet. The combination of DNS redirection of communication devices' traffic through DNS proxy servers and the simultaneous filtering of TCP/UDP traffic in this algorithm is an advantage of the current research over the papers reviewed above [23–26]. At the same time, the application of the algorithm to block data leaks from communication devices showed no loss of operability of the system and application software. Users were able to choose their own level of privacy, managing the collection of their data while doing any actions in the Internet space in real-time.

The disadvantages of the proposed algorithm include the implementation of the process of sequential scanning of each of the DNS proxy servers, which leads to a temporal delay before its operation, which is defined experimentally and is from 300 to 900 seconds depending on the number of DNS proxy servers derived from open Internet resources. That, in turn, makes it impossible to instantly provide the required level of user privacy due to the actual lack of tested and sorted NOA, ANM, HIA DNS proxy servers.

In addition, the DNS proxy testing process increases the total amount of DNS traffic generated by a communication device, which may not be acceptable to users paying for a fixed amount of Internet traffic.

Reducing the total testing time of DNS proxy servers can be achieved by organizing the multi-threading (parallel) process of their scanning. Moreover, the reduction in the total testing time of DNS proxy servers would decrease in direct proportion to the increase in the number of testing threads.

Further prospects for improving the proposed algorithm may include:

– introducing a User-Agent ID for DNS customers who communicate under the HTTP protocol;

– introducing a check time installation feature for a DNS proxy server tested;

– introducing the DNS proxy recognition feature AnchorFree, CoDeen, TinyProxy, owned by IT companies providing private surfing services;

– introducing the anchorFree, CoDeen, TinyProxy proxy servers excluding function from the work server list.

Implementing these features could reduce the time to test DNS proxy servers and improve user privacy.

8. Conclusions

1. We have analyzed the process of data exchange between DNS clients and the Internet services with which they interact. The study of the scheme of data exchange between a communication device and the Internet space has revealed the ways of data leakage from communication devices. Because all DNS customer requests are accumulated in the provider's DNS logs, DNS query analysis makes it possible to form a digital profile of the communication device.

2. An algorithm has been developed to block data leaks collected by developers of the software installed on a communication device, in order to give users the ability to choose their privacy level. The practical application of the developed algorithm has made it possible to exclude the logging of DNS traffic by Internet providers and thus block the collection of user data from communication devices. The proposed algorithm could significantly reduce the accuracy of digital profiling of the user's communication devices. A significant advantage is the ability to give the user the choice of the desired level of privacy in the Internet space.

3. TCP/UDP traffic from various communication devices has been audited over a long time. The analysis revealed the domains and IP addresses of Internet resources that the system and application software of communication devices refers to. Internet data collection and monetization services that perform requests for user data are organized in accordance with the affiliation of IT companies.

4. Checking the proposed algorithm for the absence of DNS traffic leaks from a communication device showed no loss of operability of the system and application software. The selective blocking of Internet traffic was carried out by setting up a list of prohibited IP addresses of the network firewall in accordance with the experimentally obtained data.

References

1. García-Dorado, J. L., Ramos, J., Rodríguez, M., Aracil, J. (2018). DNS weighted footprints for web browsing analytics. Journal of Network and Computer Applications, 111, 35–48. doi: http://doi.org/10.1016/j.jnca.2018.03.008
2. Guelke, J. (2020). Leaking. International Encyclopedia of Ethics, 6, 1–7. doi: http://doi.org/10.1002/9781444367072.wbiee898
3. Trish, B. (2018). Big Data under Obama and Trump: The Data-Fueled U.S. Presidency. Politics and Governance, 6 (4), 29–39. doi: http://doi.org/10.17645/pag.v6i4.1565
4. Esteve, A. (2017). The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. International Data Privacy Law, 7 (1), 36–47. doi: http://doi.org/10.1093/idpl/ipw026
5. Google: зловещая черта (2019). Available at: https://eurasia.film/2019/08/google-v-tvoej-golove/
6. Saeli, S., Bisio, F., Lombardo, P., Massa, D. (2020). DNS Covert Channel Detection via Behavioral Analysis: a Machine Learning Approach. International Conference on Malicious and Unwanted Software (MALWARE), 46–55. Available at: https://www.researchgate.net/publication/344485984_DNS_Covert_Channel_Detection_via_Behavioral_Analysis_a_Machine_Learning_Approach
7. Chen, X., Navidi, T., Rajagopal, R. (2020). Generating private data with user customization. Available at: https://www.researchgate.net/publication/346614406_Generating_private_data_with_user_customization
8. Liu, X., Li, H., Lu, X., Xie, T., Mei, Q., Feng, F., Mei, H. (2018). Understanding Diverse Usage Patterns from Large-Scale Appstore-Service Profiles. IEEE Transactions on Software Engineering, 44 (4), 384–411. doi: http://doi.org/10.1109/tse.2017.2685387
9. Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D. et. al. (2020). Predicting personality from patterns of behavior collected with smartphones. Proceedings of the National Academy of Sciences, 117 (30), 17680–17687. doi: http://doi.org/10.1073/pnas.1920484117
10. Waheed, H., Anjum, M., Rehman, M., Khawaja, A. (2017). Investigation of user behavior on social networking sites. PLOS ONE, 12 (2), e0169693. doi: http://doi.org/10.1371/journal.pone.0169693
11. Zadereyko, O., Trofymenko, O., Loginova, N. (2019). Algorithm of user's personal data protection against data leaks in Windows 10 OS. Informatyka Automatyka Pomiary w Gospodarce i Ochronie Środowiska, 9 (1), 41–44. doi: http://doi.org/10.5604/01.3001.0013.0905
12. Raber, F., Vossebein, N. (2017). URetail: Privacy User Interfaces for Intelligent Retail Stores. Human-Computer Interaction INTERACT 2017. Lecture Notes in Computer Science. Cham: Springer, 10516, 473–477. doi: http://doi.org/10.1007/978-3-319-68059-0_54
13. Siby, S., Juarez, M., Diaz, C., Narseo, V., Troncoso, C. (2019). Encrypted DNS – Privacy? A Traffic Analysis Perspective. Cryptography and Security, 1–19. Available at: https://arxiv.org/abs/1906.09682
14. Grothoff, C., Wachs, M., Ermert, M., Appelbaum, J. (2018). Toward secure name resolution on the internet. Computers & Security, 77, 694–708. doi: http://doi.org/10.1016/j.cose.2018.01.018
15. Bumanglag, K., Kettani, H. (2020). On the Impact of DNS Over HTTPS Paradigm on Cyber Systems. 3rd International Conference on Information and Computer Technologies (ICICT). San Jose, 494–499. doi: http://doi.org/10.1109/icict50521.2020.00085
16. Yan, Z., Lee, J.-H. (2020). The road to DNS privacy. Future Generation Computer Systems, 112, 604–611. doi: http://doi.org/10.1016/j.future.2020.06.012
17. Imana, B., Korolova, A., Heidemann, J. (2018). Enumerating Privacy Leaks in DNS Data Collected Above the Recursive. Proceedings of the ISOC NDSS Workshop on DNS Privacy. San Diego, 1–7. Available at: https://www.isi.edu/~johnh/PAPERS/Imana18a.pdf
18. Hoang, N., Niaki, A., Borisov, N., Gill, P., Polychronakis, M. (2020). Assessing the Privacy Benefits of Domain Name Encryption. Proceedings of the 15th ACM Asia Conference on Computer and Communications Security (ASIA CCS '20). New York, 290–304. doi: http://doi.org/10.1145/3320269.3384728
19. Deccio, C., Davis, J. (2019). DNS privacy in practice and preparation. Proceedings of the 15th International Conference on Emerging Networking Experiments and Technologies (CoNEXT'19), 138–143. doi: http://doi.org/10.1145/3359989.3365435
20. Beliavskii, D. (2015). DNS: kto ne spriatalsia, tot i vinovat. Internet v tsifrakh, 1 (21), 74–77. Available at: http://37.230.117.45/upload/iblock/690/6900620c7bef412cfa870a549817b4fd.pdf
21. Houser, R., Li, Zh., Cotton, Ch., Wang, H. (2019). An investigation on information leakage of DNS over TLS. Proceedings of the 15th International Conference on Emerging Networking Experiments and Technologies (CoNEXT '19) New York, 123–137. doi: http://doi.org/10.1145/3359989.3365429
22. Borgolte, K., Chattopadhyay, T., Feamster, N., Kshirsagar, M., Holland, J., Hounsel, A., Schmitt, P. (2019). How DNS over HTTPS is Reshaping Privacy, Performance, and Policy in the Internet Ecosystem. SSRN Electronic Journal. doi: http://doi.org/10.2139/ssrn.3427563
23. Rai, T., Verma, R. (2015). Packet Filtering Technique for Network Security. International Journal of Engineering Research & Technology (IJERT), 3 (20), 1–3. Available at: https://www.ijert.org/research/packet-filtering-technique-for-network-security-IJERTCONV3IS20047.pdf
24. Sheluhin, O. I., Smychek, M. A., Simonyan, A. G. (2018). Filtering unwanted applications of Internet resources for information security purposes. H&ES Research, 10 (2), 87–98. Available at: https://www.elibrary.ru/item.asp?id=34939631

25. Smart DNS Proxy Servers. Available at: https://www.smartdnsproxy.com/Servers
26. Podkorytov, D., Floka, A., Kuleshov S. (2019). Arkhitektura krossplatformennogo DNS Proxy servisa. T-Comm: Telekommunikatsii i transport, 13 (5), 35–40. Available at: https://www.researchgate.net/publication/333844552_Podkorytov_DA_Floka_AB_Kulesov_SV_Arhitektura_krossplatformennogo_DNS_Proxy_servisa_T-Comm_Telekommunikacii_i_transport_2019_Tom_13_No5_S_35-40
27. Dooley, M., Rooney, T. (2020). Navigating the Internet with DNS. IP Address Management, 75–92. doi: http://doi.org/10.1002/9781119692263.ch4
28. Fujiwara, K., Sato, A., Yoshida, K. (2019). Cache Effect of Shared DNS Resolver. IEICE Transactions on Communications, E102.B (6), 1170–1179. doi: http://doi.org/10.1587/transcom.2018ebp3184
29. General Data Protection Regulation (EU GDPR). Available at: https://gdpr-text.com/
30. Charanjeet, S. (2020). How to Enable DNS Over HTTPS in Chrome, Firefox, Edge, Brave & More? Fossbytes. Available at: https://fossbytes.com/how-to-enable-dns-over-https-on-chrome-firefox-edge-brave/
31. Ashok, A., John, A., Joy, P., Vijayan, R., Amrutha, V., Deepa, K., Jooby, E. (2016). Proxy Server Protection for Web Search. International Journal of Computer Science and Technology, 7 (1), 165–169. Available at: http://www.ijcst.com/vol71/2/34-amrutha-ashok.pdf
32. Shima, K., Nakamura, R., Okada, K., Ishihara, T., Miyamoto, D., Sekiya, Y. (2019). Classifying DNS Servers Based on Response Message Matrix Using Machine Learning. International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, 1550–1551. doi: http://doi.org/10.1109/csci49370.2019.00291