Eastern-European Journal of Enterprise Technologies
jet.com.ua

EASTERN-EUROPEAN JOURNAL OF ENTERPRISE TECHNOLOGIES

ISSN 1729-3774

1/2 (109)
2021

information technologies
інформаційні технології
информационные технологии

new economy
нова економіка
новая экономика

industrial applications
промислові технології
промышленные технологии

УДК 004.738.2

РАЗРАБОТКА АЛГОРИТМА ЗАЩИТЫ УСТРОЙСТВ КОММУНИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ ОТ УТЕЧЕК ДАННЫХ

А. В. Задерейко, Ю. В. Прокоп, Е. Г. Трофименко, Н. И. Логинова, О. Е. Плачинда

Для выявления путей сбора данных с устройств коммуникации пользователей проведен анализ взаимодействия клиентов DNS с доменным пространством имен Интернет. Установлено, что DNS трафик устройства коммуникации журналируется DNS серверами провайдера, что само по себе несет угрозу приватности пользователей. Разработан и апробирован комплексный алгоритм защиты от сбора данных пользователей, состоящий из двух модулей. Первый модуль позволяет перенаправить DNS трафик устройства коммуникации через DNS прокси-сервера с заданным классом анонимности, определяемым на основе предложенного мультитеста. Для обеспечения бесперебойного и устойчивого соединения модуль осуществляет автоматическое подключение к DNS прокси-серверу, имеющему минимальное время отклика из доступных в сформированном списке. Второй модуль блокирует сбор данных, собираемых разработчиками программного обеспечения, установленного на устройстве коммуникации пользователя, и специализированными интернет-сервисами, принадлежащими IT-компаниям. Предложенный алгоритм позволяет пользователям выбирать предпочитаемый уровень приватности при осуществлении коммуникации с интернет-пространством, тем самым обеспечивая им выбор уровня приватности и, как следствие, ограничить возможность информационных манипуляций над их владельцами. Проведен аудит DNS трафика различных стационарных и мобильных устройств коммуникации. Анализ DNS трафика позволил идентифицировать и структурировать DNS запросы, отвечающие за сбор данных пользователей интернет-сервисами, принадлежащими IT-компаниям. Выполнена блокировка идентифицированных DNS запросов и экспериментально подтверждено отсутствие потерь работоспособности базового и прикладного программного обеспечения на устройствах коммуникации.

Ключевые слова: DNS запрос, DNS сервер, утечки DNS, DNS трафик, DNS прокси-сервер, сбор данных.

1. Введение

В современном интернет-пространстве циркулируют гигантские объемы информации, большую часть которой составляют данные обмена пользователей в результате их взаимодействия с различными интернет-сервисами. Структуризация и анализ этих данных позволяет выявлять скрытые на первый взгляд закономерности, прогнозировать, а при системном подходе и формировать поведенческие тенденции интернет-аудитории.

Данная ситуация усугубляется благодаря энергичным усилиям высокотехнологичных IT-компаний, по внедрению систем сбора и анализа данных, что ведет к негласной монополизации рынка цифровых данных пользователей. При этом регулирующая роль различных государственных институтов в вопросах соблюдения прав на приватность пользователей, а именно: тайну переписки и действиях в интернет-пространстве, неуклонно снижается. Складывающаяся тенденция вызывает все большую обеспокоенность у пользователей интернет-пространства, сотрудников IT-компаний и общественных неправительственных организаций. Они обращают внимание на недопустимость несанкционированного сбора и монетизации данных без какого-либо согласия пользователей интернет-пространства [1]. Требования к реализации и мер по повышению приватности пользователей регулируются IISO/IEC 24760-1:2019(E) IT Security and Privacy.

В связи с этим актуальными являются исследования, направленные на формирование новых подходов и разработку инструментов защиты данных пользователей в сети Интернет. Пользователи имеют право не только знать, какую именно информацию о них могут собирать интернет-сервисы, но и обладать возможностью самостоятельно выбирать уровень приватности в интернет-пространстве. Тем более, что разработка средств контроля и управления приватностью для предотвращения нежелательной обработки персональной идентификационной информации определена стандартом ISO/IEC 29100:2011.

2. Анализ литературных данных и постановка проблемы

В работе [2] приведены результаты исследований, показывающие, что к сбору данных пользователей в интернет-пространстве проявляют интерес как государственные, так и IT-компании. Данные, собранные при отслеживании Domain Name System (DNS) запросов подсистемой MoreCowBell в рамках проекта PRISM, были задействованы в управлении различными общественными процессами [3]. Большой интерес к сбору и анализу данных пользователей проявляют IT-компании, старающиеся как можно эффективнее монетизировать свои объявления [4]. Кроме этого, различные IT-компании собирают статистику об использовании интернет-ресурсов и обрабатывают ее в автоматическом режиме [5]. Это подтверждает гипотезу о том, что с устройств коммуникации, подключенных к сети Интернет, без ведома пользователей отслеживается и собирается информация об их действиях. Но при этом в этих работах не предложены пути защиты данных пользователей, исключающие мониторинг со стороны IT-компаний.

Наиболее перспективным направлением с точки зрения обеспечения максимально возможной точности сбора информации о действиях пользователей в интернет-пространстве является анализ DNS трафика клиентов DNS, установленных на устройствах коммуникации [6]. В работе [7] показано, что анализ DNS трафика позволяет определить программное обеспечение, установленное на устройствах коммуникации. Также можно получить данные об истории геолокации, учетных записях в интернет-сервисах, интересах,

религиозных предпочтениях, финансовом состоянии, медицинских потребностях и др. Результатом такого анализа является создание базы уникальных цифровых профилей устройств коммуникации [8, 9], и как следствие, точное прогнозирование поведенческой реакции интернет-аудитории и разработка возможных сценариев влияния на её поведение [10]. Это обуславливает проблему незащищенности пользователей от мониторинга их сетевого трафика и делает невозможным выбор уровня их приватности при осуществлении коммуникации с интернет-пространством.

Исследователи изучали возможности предотвращения утечек данных пользователей. Так, в работе [11] определены и систематизированы данные, собираемые операционными системами семейства Windows на устройствах коммуникации пользователя и отправляемые на серверы компании Microsoft. Такая тенденция естественным образом обуславливает поиск решений, предоставляющих пользователям возможность выбора, к каким своим данным они могут давать доступ в интернет-пространстве. В частности, в статье [12] представлен интерфейс URetail в виде радара, позволяющий пользователю выбирать, какие из его личных данных могут быть раскрыты. Но реализация этого подхода носит узкую направленность относительно данных, собираемых в розничной торговле при покупках в интернет-магазинах.

Разработкой методов анализа DNS трафика, вопросами его шифрования с целью защиты DNS запросов пользователей от мониторинга и цензуры занимались многие ученые. Например, авторы работы [13] пришли к выводу о неэффективности существующих стандартных схем организации DNS трафика. В работах [14–17] подчёркивается актуальность защиты DNS трафика, и указывается на необходимость тщательного анализа возможных его утечек. Так, в статье [14] рассматривается принципы работы DNS и анализируются системы Namecoin, GNU и RAINS. В работе [15] рассматриваются уязвимости протокола DNS и то, как вредоносное программное обеспечение использует эти уязвимости. В исследовании [16] выявлена проблема утечки конфиденциальности DNS и анализируется использование технологий шифрования сетевого трафика HTTPS/TLS (DoH/DoT) и SNI (ESNI). В работе [17] была выполнена оценка утечек DNS трафика. В работах [18–20] анализируются преимущества и недостатки шифрования DNS трафика с применением протоколов DNS over TLS (DoT), DNS over HTTPS (DoH). Исследование [18] показало, что даже когда шифрование включено, данные пользователей утекают через их DNS-запросы. Кроме того, было выявлено [19], что протоколы DoT и DoH поддерживаются лишь небольшим количеством DNS серверов. Существенным является то, что шифрование требует дополнительных вычислительных ресурсов и замедляет обработку DNS-запросов [20]. В работе [21] проанализированы уязвимости протокола DoT. В статье [22] исследованы производительность протокола DoH и последствия внедрения протоколов шифрования DNS трафика для участников интернет-пространства. Однако нерешенным остается вопрос о реализации мер по повышению приватности пользователей при осуществлении коммуникации в сети Интернет.

Способы, позволяющие повысить приватность пользователей, рассмотрены в работах [23, 24], где предлагается внедрение фильтрации сетевого трафика устройств коммуникации. Однако пакетная фильтрация в силу своей специфики и особенностей отдельных протоколов, к которым применяются фильтры, не является достаточным средством для обеспечения защиты данных пользователей. Фильтрация сетевого трафика может использоваться как одно из средств блокировки входящих и исходящих IP-пакетов.

Перенаправление трафика через дополнительный промежуточный DNS сервер, внедряемый между клиентом DNS и удаленным DNS сервером, предлагается в работах [25, 26]. Так, в работе [25] идея использования Smart DNS Proxy Server рассматривается для получения доступа к интернет-ресурсам сайтам, недоступным из-за географических ограничений. Исследование [26] посвящено построению архитектуры сетевого сервиса, выполняющего функции UDP Proxy. Но при этом для защиты DNS запросов от мониторинга со стороны интернет-провайдеров не используются механизмы фильтрации и криптографических преобразований.

Систематизация результатов исследований позволяет утверждать о недостаточном изучении путей сбора данных с устройств коммуникации пользователей при взаимодействии клиентов DNS с доменным пространством имен. Все это позволяет утверждать, что целесообразным является проведение исследования, посвященного разработке средств, позволяющих одновременно локализовать утечки DNS трафика, скрыть реальный IP-адрес устройства коммуникации и заблокировать сбор данных пользователя.

3. Цель и задачи исследования

Цель исследования – разработка алгоритма защиты устройств коммуникации от несанкционированных сбора и утечек данных пользователей в сети Интернет. Практическое применение разработанного алгоритма предоставит пользователям возможность самим определять уровень приватности.

Для достижения поставленной цели сформированы такие задачи:

– проанализировать процесс обмена данными между клиентами DNS и интернет-сервисами, с которыми они взаимодействуют, для выявления утечки и способов сбора данных с устройств коммуникации пользователей;

– разработать алгоритм блокировки утечек данных, собираемых разработчиками программного обеспечения, установленного на устройстве коммуникации, для предоставления пользователям возможности выбора уровня своей приватности при взаимодействии с различными интернет-сервисами;

– выполнить аудит TCP/UDP трафика различных устройств коммуникации для выявления сервисов, посылающих запросы на сбор данных пользователя;

– выполнить проверку предложенного алгоритма на отсутствие утечек DNS трафика с устройства коммуникации.

4. Исследование процесса обмена данными между клиентами DNS и интернет-сервисами

Физическое подключение устройства коммуникации пользователя к интернет-пространству и его последующее обращение к ресурсам сети Интернет начинается с отсылки DNS запросов на различные интернет-сервисы. При этом, клиентом DNS может выступать любое программное обеспечение, установленное на устройствах коммуникации пользователей, например: веб-браузеры, файловые менеджеры, почтовые клиенты, мессенджеры и т. п., выполняющие DNS запросы. Взаимодействие клиентов DNS с интернет-пространством и обработка DNS запросов в доменном пространстве выполняется в строго определенном порядке [27].

На практике, чтобы уменьшить время ответа на DNS запрос и снизить нагрузку на корневые DNS сервера, провайдеры создают собственный кэш DNS сервера [28]. В случае соответствия DNS запроса, ранее зафиксированному в кэше DNS сервера, выдается соответствующий IP-адрес (рис. 1).

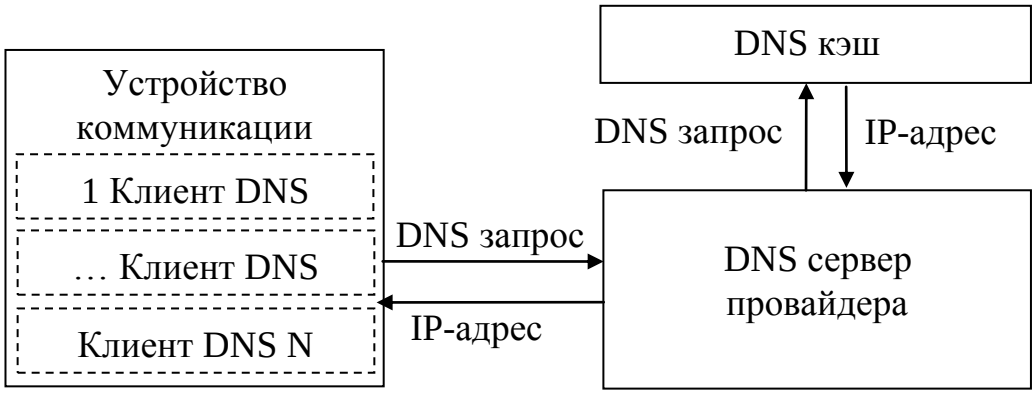

Рис. 1. Схема кэширования DNS запросов на DNS сервере провайдера

Таким образом, все запросы клиентов DNS накапливаются в журналах DNS сервера провайдера. Структурирование и анализ данных DNS запросов может дать исчерпывающую информацию о действиях пользователя в интернет-пространстве. К их сбору, хранению и анализу которых с каждым годом проявляют все больший интерес различные структуры государственной безопасности, рекламные и аналитические подразделения IT-компаний, а также представители организованной киберпреступности. Именно поэтому данные о пользователях все чаще называют «цифровым золотом».

С учетом выше рассмотренного, предложена схема обмена данными устройства коммуникации с интернет-пространством (рис. 2). Ее анализ позволяет сделать вывод, что конечными бенефициарами данных пользователей, так или иначе, становятся IT-компании.

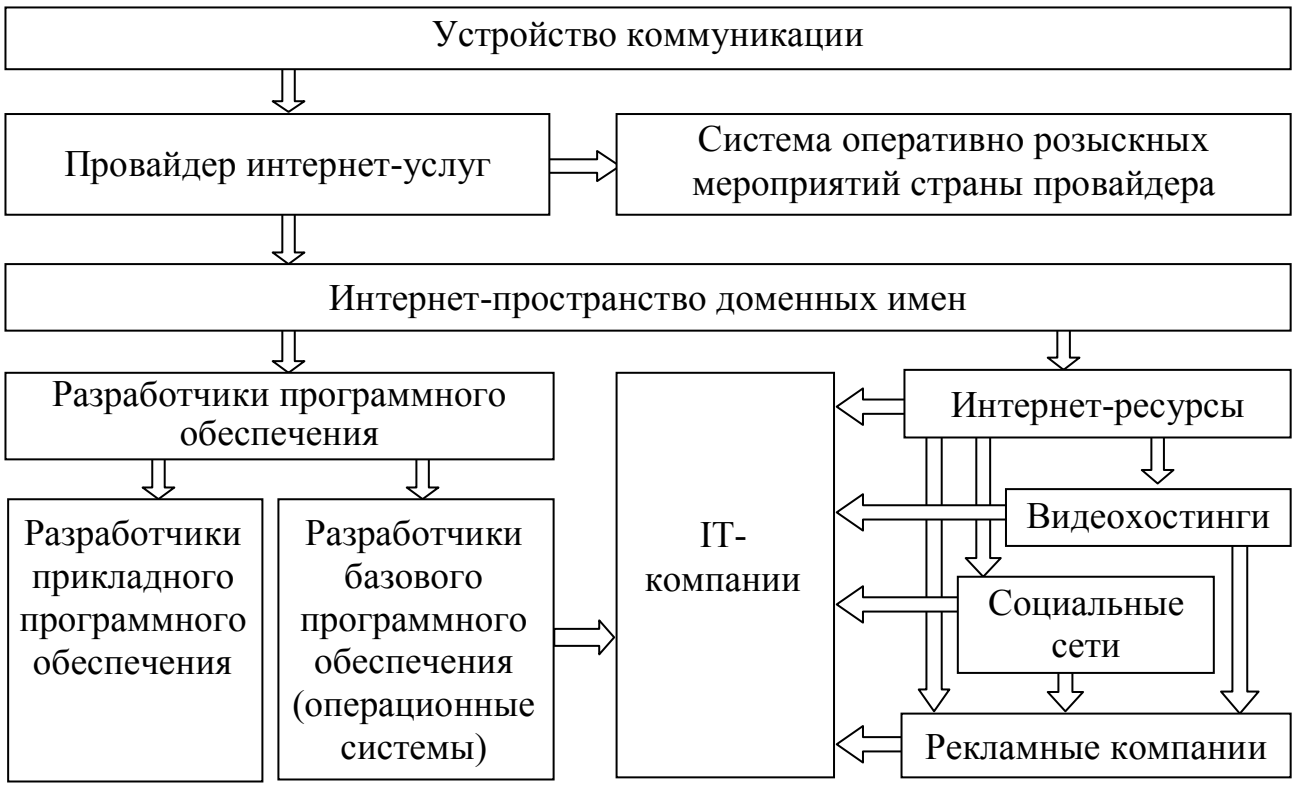

Рис. 2. Схема обмена данными между устройством коммуникации и интернет-пространством

Организация массовости и непрерывности процесса сбора данных с устройств коммуникации достигается за счет внедрения IT-компаниями бесплатного доступа к сервисам сбора и анализа интернет-статистики: Google Analytics, Yandex Metrika, Liveinternet, Rambler и т.д. Такой подход позволяет IT-компаниям, внедряя системы автоматизированной обработки собираемых данных, осуществлять не только цифровое профилирование устройств коммуникации, но и создавать уникальные цифровых профили для каждого из их реальных пользователей [8, 9].

Не удивительно, что такая тенденция вызывает обеспокоенность руководства ряда демократических стран. Так, например, страны Евросоюза на законодательном уровне ужесточили контроль и ответственность за посягательства на персональные данные граждан ЕС на его территории и за его пределами, приняв Закон GDPR (General Data Protection Regulation) [29]. Однако, даже эти строгие меры, по сути, не решают главной проблемы. Они не наделяют пользователей возможностью самим определять уровень своей приватности, управляя сбором своих данных при выполнении каких-либо действий в интернет-пространстве в режиме реального времени.

5. Разработка алгоритма блокировки утечек данных с устройства коммуникации пользователя

Комплекс мер для предотвращения сбора данных с устройств коммуникации, и, следовательно, снижению вероятности их цифрового профилирования, включает два модуля:

1. модуль защиты от утечек DNS трафика путем:

– передачи DNS запросов по протоколу DoH;

– перенаправления DNS трафика на DNS прокси-сервер с заданным уровнем приватности;

2. модуль блокировки сбора данных путем:

– блокировки плагинов сбора данных, интегрированных в Content Management System (CMS) интернет-ресурсов;

– блокировки DNS трафика системного и прикладного программного обеспечения.

Первый модуль защиты от утечек DNS трафика является ключевым. Это обусловлено тем, что интернет-провайдеры, осуществляющие подключение пользователей к доменному пространству Интернет, выполняют его через подконтрольные им DNS сервера, ведущие обязательные журналы записей DNS запросов каждого пользователя. Очевидно, что провайдеры имеют возможность:

– связать IP-адрес каждого пользователя со всеми доменными именами, которые были им запрошены;

– хранить накопленные данные неограниченно долго;

– предоставить накопленные данные уполномоченным государственным структурам.

Таким образом, пользователь не может быть уверен в своей приватности, осуществляя интернет-коммуникацию через DNS сервера провайдера.

Кроме этого, провайдеры по умолчанию устанавливают для своих пользователей режим принудительного подключения к своему DNS серверу, если пользователь изменяет свои настройки для использования стороннего DNS сервера. В случае обнаружения таких настроек DNS в устройстве коммуникации, провайдеры используют прозрачный DNS прокси-сервер, который осуществляет перенаправление DNS трафика пользователя. Таким образом, провайдер осуществляет маскировку реального маршрута DNS трафика пользователя. Этот прием позволяет гарантировано отправлять DNS запросы пользователя на DNS сервер провайдера и продолжать выполнять журналирование его DNS трафика.

Еще одним немаловажным обстоятельством, определяющим контроль за DNS трафиком пользователя, является то, что протокол DNS по умолчанию не выполняет шифрование DNS запросов. Попытки осуществления криптографического шифрования DNS трафика отразились в разработке и внедрении протоколов DNScrypt, DoT и DoH. Эти протоколы выполняют шифрование DNS трафика, создавая криптографически защищенный канал между клиентами и серверами DNS. Именно это обстоятельство подтолкнуло IT-компании заявить о поддержке реализованных технологий шифрования DNS трафика и создать контролируемые ими же публичные DNS сервера с поддержкой протоколов DNSCrypt, DoT, DoH (табл. 1).

Таблица 1

Публичные DNS сервера IT-компаний с поддержкой DNSCrypt, DoT, DoH

Публичные DNS сервера IT-корпораций	Маски IP-адресов		Поддержка протоколов (+/–)		
	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	+	+	+

Количество IT-компаний, поддерживающих шифрование DNS трафика с использованием приведенных криптографических протоколов, продолжает увеличиваться, что однозначно позволяет:

– противодействовать подмене DNS ответов на транзитных узлах системы DNS;

– обойти блокировку (цензурирование) DNS трафика провайдерами;

– сделать невозможным журналирование и последующее инспектирование DNS трафика;

– снизить роль провайдеров, осуществляющих подключение устройств коммуникации к интернет-пространству;

– снизить роль корневых DNS серверов;

– перераспределить сбор данных о DNS трафике пользователей в пользу крупных IT-компаний.

В дополнение к этому большинство разработчиков веб-браузеров уже не только реализовали возможность использования протокола DoH в свои программные продукты, но и реализовали возможность подключения к публичным DNS серверам ведущих IT-компаний [30]. Тенденция к монополизации IT-компаниями DNS трафика значительно усиливает актуальность вопроса обеспечения реальной приватности пользователей, поскольку именно этим IT-компаниям принадлежат сервисы сбора и анализа интернет-статистики. Примерами таких сервисов являются Google Analytics, Yandex Metrika, Liveinternet, Rambler TOR и др. Кроме того, нельзя исключать того, что IT-компании могут предоставить доступ сторонним лицам или уполномоченным государственным структурам к истории DNS трафика пользователей, воспользовавшихся услугами предоставляемых публичных DNS серверов.

Для обеспечения приватности пользователей интернет-пространства, наряду с применением упомянутых технологий шифрования DNS трафика, предложено перенаправлять DNS запросы через DNS прокси-сервера различного класса

анонимности. Эти DNS прокси-сервера должны иметь фиксированный во времени срок функционирования и не должны выполнять журналирование DNS запросов. Преимущество применения такого подхода позволяет исключить возможность накопления данных о DNS запросах пользователей не только у провайдеров и уполномоченных государственных структур, но и у IT-компаний.

При этом для обеспечения защищенной передачи DNS запросов необходимым условием является использование DNS прокси-серверов, поддерживающих протокол DoH. А для обеспечения наивысшей приватности пользователя следует перенаправлять DNS трафик устройства коммуникации через DNS прокси-сервера класса HIA (High anonymous). Такие прокси-сервера скрывают реальный IP-адрес клиента DNS и не позволяют запрашиваемому DNS серверу определить использование DNS прокси-серверов [31, 32].

Предложена схема перенаправления DNS трафика устройства коммуникации через DNS прокси-сервер (рис. 3).

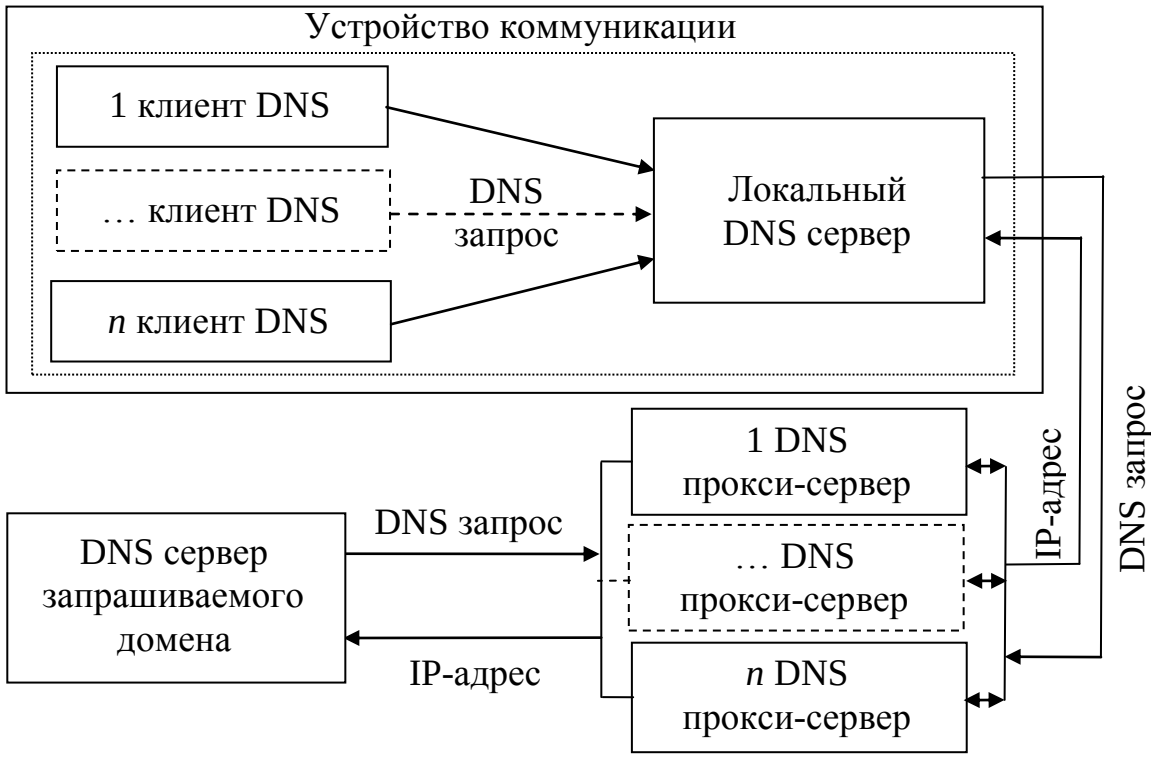

Рис. 3. Перенаправление DNS трафика через DNS прокси-сервер

Перенаправление DNS запросов устройства коммуникации реализовано за счет изменения маршрута DNS запросов от клиента DNS к запрашиваемому домену. Его отличительными особенностями являются:

– создание локального DNS сервера;

– перенаправление DNS запросов клиентов DNS с устройства коммуникации на локальный DNS сервер;

– перенаправление DNS запросов с локального DNS сервера на заранее выбранный DNS прокси-сервер по предлагаемому алгоритму (рис. 4).

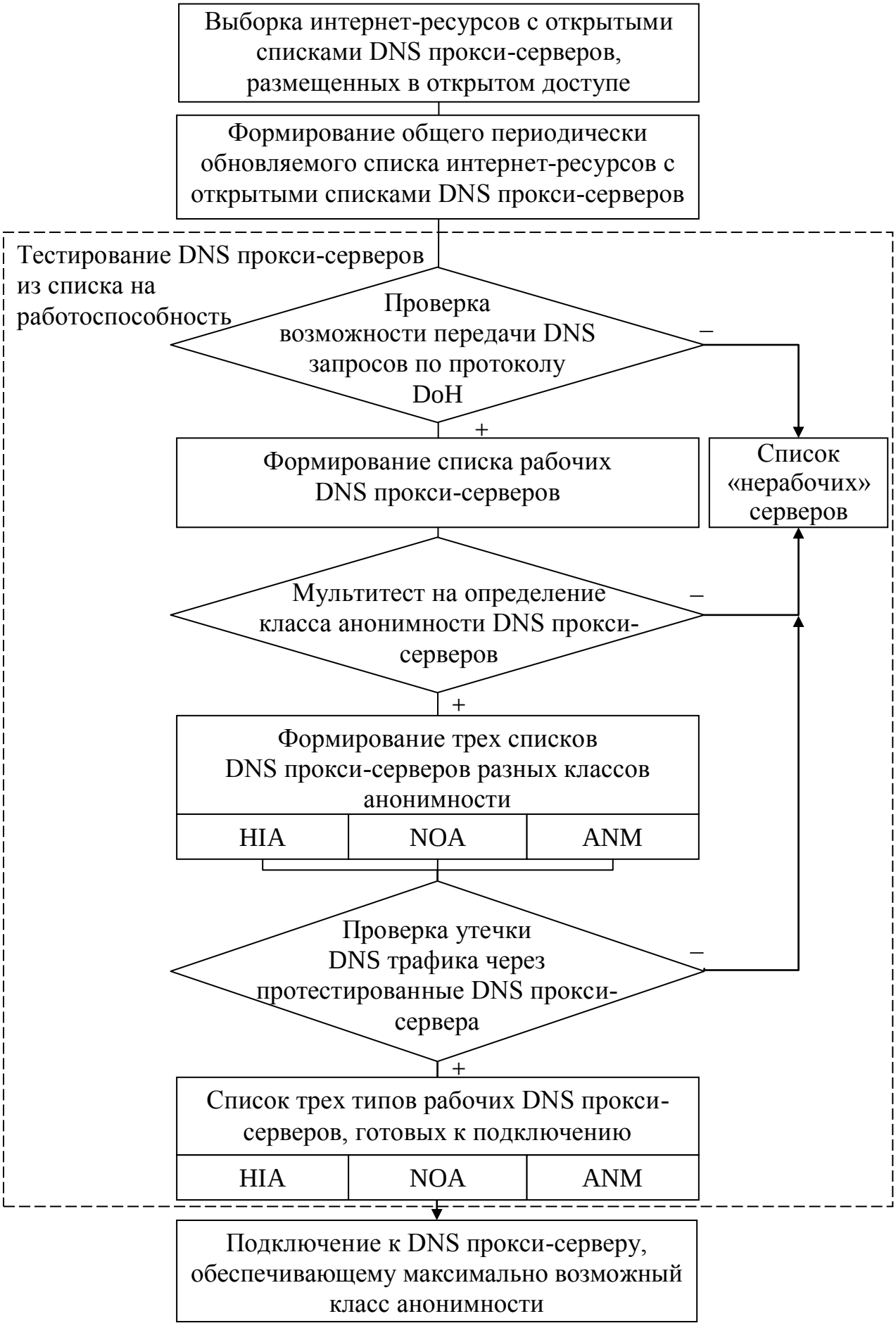

Рис. 4. Алгоритм тестирования DNS прокси-серверов

Для реализации предлагаемого алгоритма выбора DNS прокси-серверов необходимо выполнить следующее:

1. осуществить поиск интернет-ресурсов с открытыми списками DNS прокси-серверов, размещенных в открытом доступе;

2. сформировать общий список интернет-ресурсов с открытыми списками DNS прокси-серверов;

3. создать периодически обновляемый список DNS прокси-серверов для последующего тестирования:

3.1. проверить наличие возможности передачи DNS запросов по протоколу HTTPS. Для формирования списка рабочих DNS прокси-серверов, удовлетворяющих заданным требованиям анонимности, необходимо реализовать процесс их проверки (тестирования) на работоспособность. Для этого необходимо последовательно выполнить DNS запросы через каждый DNS прокси-сервер вида:

– 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. переместить в список рабочих DNS прокси-серверов, которые выполнили DNS запросы п.3.1;

3.3. провести мультитест на определение класса анонимности DNS прокси-сервера:

3.3.1. проверить наличие HTTP-заголовков: 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. проверить наличие открытых портов HTTP proxy: 3128, 1080, 8123, 8000, 1080;

3.3.3. проверить наличие открытых портов web proxy: 80, 8080;

3.3.4. проверить DNS прокси-сервер на наличие названий: vpn, hide, hidden, proxy (подозрительное название хоста).

3.3.5. проверить разницу во временных зонах между IP-адресом устройства коммуникации и IP-адресом DNS прокси-сервера;

3.3.6. проверить принадлежность IP-адреса DNS прокси-сервера к сети Tor;

3.3.7. проверить использование сервисов сжатия трафика от компаний Google, Yandex и Opera путем сравнения пула IP-адресов сервисов этих компаний с IP-адресом DNS прокси-сервера (режим Turbo);

3.3.8. проверить DNS прокси-сервер на наличие переадресации устройства коммуникации путем сравнения содержимого host, полученного с window.location.hostname с содержимым host запрошенного интернет-ресурса (JavaScript метод);

3.3.9. проверить отправку IP-адреса устройства коммуникации минуя DNS прокси-сервер (утечка IP-адреса через Flash).

3.3.10. определить длительность маршрутизации DNS запросов в миллисекундах (длительность маршрутизации, составляющая более 30 милли-секунд, расценивается, как наличие DNS прокси-сервера (двусторонний пинг);

3.3.11. проверить утечки IP-адреса устройства коммуникации через WebRTC;

3.3.12. проверить DNS прокси-сервер на использование технологии VPN: выполняется анализ размера перехваченного пакета MTU и максимального размера данных MSS в передаваемом в пакете (VPN fingerprint);

3.4. распределить рабочие DNS прокси-сервера по классам анонимности, исходя из результатов мультитеста п. 3.3:

1) не анонимные NOA (Not anonymous), не скрывающие реальный IP-адрес клиента DNS;

2) анонимные ANM (Anonymous), скрывающие реальный IP-адрес клиента DNS, но позволяют запрашиваемому DNS серверу определить использование DNS прокси-серверов;

3) высокой анонимности HIA (High anonymous), скрывающие IP-адрес клиента DNS и не позволяющие запрашиваемому DNS серверу определить использование DNS прокси-серверов;

3.5. подключиться к DNS прокси-серверу, обеспечивающему максимально возможный класс анонимности пользователя.

Критерии распределения DNS прокси-серверов на основе результатов мультитеста (п. 3.3) на присвоение класса анонимности приведены в табл. 2.

Таблица 2

Критерии для присвоения класса анонимности DNS прокси-серверов

Критерии тестирования DNS прокси-серверов	Классы анонимности DNS прокси-серверов		
	HIA	ANM	NOA
Соединение HTTPS	да	да	да
Заголовки HTTP proxy	нет	нет	да
Открытые порты HTTP proxy	нет	да	да
Открытые порты web proxy	нет	нет	да
Открытые порты VPN	нет	нет	да
Подозрительное название хоста	нет	нет	да
Разница во временных зонах (устройства коммуникации и IP-адреса DNS прокси-сервера)	нет	нет	да
Принадлежность IP к сети Tor	нет	нет	нет
Режим браузера Turbo	нет	нет	да
Принадлежность IP хостинг провайдеру	нет	нет	нет
Проверка web proxy по Java Script методу	нет	да	да
Утечка IP через Flash	нет	нет	да
Определение туннеля (двусторонний пинг)	нет	да	да
VPN fingerprint	нет	да	да
Утечка IP через WebRTC	нет	да	да

Второй модуль алгоритма блокировки сбора данных пользователя осуществляет блокировку соединений между клиентами DNS устройства коммуникации и специализированными интернет-сервисами сбора данных. Кроме того, он блокирует соединения со сторонними сервисами и сервисами разработчиков системного и прикладного программного обеспечения (рис. 5).

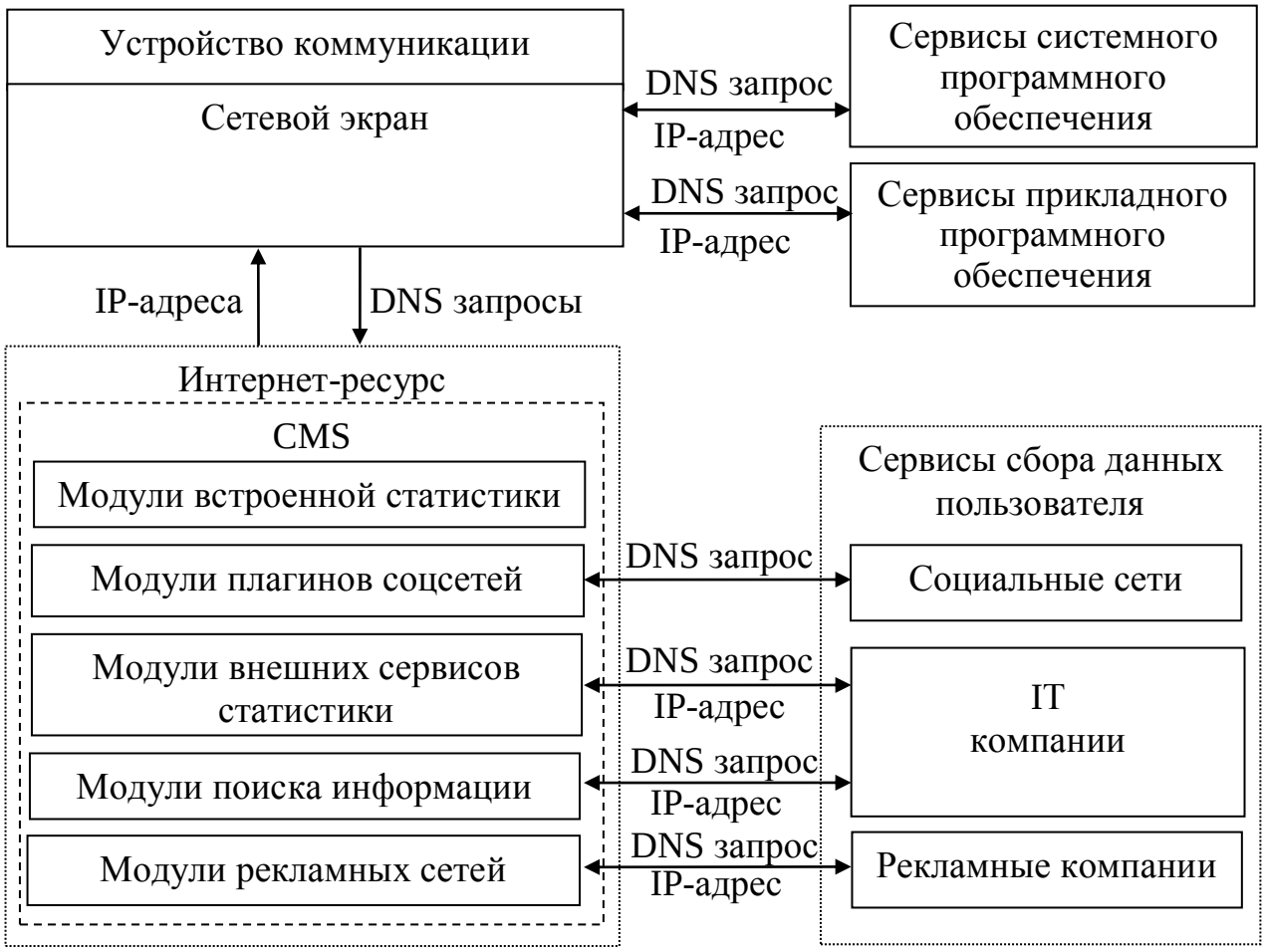

Рис. 5. Обмен данными между устройством коммуникации и интернет-пространством

Это реализуется посредством организации процесса фильтрации TCP/UDP-трафика, отвечающего за коммуникацию с интернет-сервисами:

– сбора данных пользователя;

– системного программного обеспечения;

– прикладного программного обеспечения.

На практике для фильтрации применяются межсетевые экраны, способные работать на сетевом пакетном уровне и обеспечивающие блокировку всех входящих и исходящих DNS запросов устройства коммуникации, соответствующих:

– IP-адресам сервисов сбора данных пользователя;

– IP-адресам служебного и стороннего трафика системного и прикладного программного обеспечения.

6. Результаты аудита TCP/UDP трафика

В табл. 3 представлены результаты комплексного мониторинга стационарного и мобильного TCP/UDP трафика устройств коммуникации в течение длительного времени.

Таблица 3

Сервисы сбора, анализа и монетизации данных пользователей

IT-компании	Програм-мное обеспечение	Интернет-соединения			
		Доменное имя	IP-адрес	Протоколы	
				TCP	UDP
Yandex	Сервис монети-зации “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		+
	Сервис сбора данных “Yandex Metrika”	ya.ru	87.250.250.242		+
		yastatic.net	178.154.131.215 178.154.131.216 178.154.131.217		+
Google	Сервис монетизации “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.googlesyndica-tion.com	172.217.13.66		+
		www.googletagservices.com	216.239.36.10		+
	Сервис сбора данных “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		+
Liveinter-net	Сервис сбора данных	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		+

Прикладное программное обеспечение для устройств коммуникации					
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		+
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	+	

В табл. 3 определены доменные имена и IP-адреса системного и прикладного программного обеспечения, интернет-сервисов сбора, анализа и монетизации, устанавливающие соединения с устройством коммуникации. Они систематизированы в соответствии принадлежности IT-компаниям.

Анализ DNS трафика системного и прикладного программного обеспечения позволил установить домены интернет-ресурсов, к которым

обращаются системное и прикладное программное обеспечение (табл. 2). Данные о доменах, полученные из открытых источников:

– домен sl-reverse.com принадлежит IT-компании CSC Digital Brand Services, специализирующейся на цифровом управлении брендами и цифровом маркетинге;

– домен cloudfront.net принадлежит IT-компании Amazon, специализирующейся на предоставлении широкого спектра услуг в облачных сервисах, основанных на результатах анализа DNS трафика;

– домен te.net.net принадлежит IT-компании Bodis, LLC, оказывающей услуги по монетизации и управлению доменным трафиком;

– домен host.hit.gemius.pl принадлежит IT-компании Gemius выполняющей исследования в области потребления средств массовой информации, и разрабатывающей инструменты, используемые для оптимизации рекламных кампаний;

– домен 1e100.net принадлежит IT-компании Google;

– домены compute-1.amazonaws.com и eu-central-1.amazonaws.com принадлежат IT-компании Amazon.

Полученные данные о владельцах доменов (табл. 3), позволяют утверждать, что мобильное прикладное программное обеспечение: Facebook, Instagram, Viber и Telegram, устанавливает соединения с интернет-сервисами, принадлежащими IT-компаниям Google, Amazon и Cloudflare.

Для обеспечения приватности пользователей все соединения с IP-адресами, указанными в табл. 3, должны быть заблокированы, что определено функционалом второго модуля предложенного алгоритма.

7. Обсуждение результатов работы алгоритма на отсутствие утечек DNS трафика с устройства коммуникации

Предложена схема обмена данными между устройствами коммуникации и интернет-пространством (рис. 2), которая позволила установить, что запросы клиентов DNS накапливаются в журналах DNS сервера провайдера. Журналы после структурирования и анализа DNS запросов могут быть использованы различными структурами государственной безопасности, рекламными и аналитическими подразделениями IT-компаний, а также представителями организованной киберпреступности для получения приватной информации о пользователях.

Предложенный алгоритм блокировки утечек данных с устройства коммуникации пользователя состоит из двух модулей – модуль защиты от утечек DNS трафика и модуль блокировки сбора данных. Первый модуль передает DNS запросы по протоколу DoH и перенаправляет DNS трафик на DNS прокси-сервер с заданным классом анонимности. Второй модуль блокирует плагины сбора данных, интегрированные в Content Management System (CMS) интернет-ресурсов, и блокирует сторонний TCP/UDP трафик системного и прикладного программного обеспечения. Проведенный анализ публичных DNS серверов IT-компаний, поддержавших внедрение протоколов DNScrypt, DoT и DoH (табл. 1), позволил сделать выводы о том, что IT-

компании могут противодействовать подмене DNS ответов на транзитных узлах системы DNS и обходить блокировку DNS трафика провайдерами. Кроме этого, невозможность журналирования и последующего инспектирования DNS трафика снизила роль провайдеров, осуществляющих подключение устройств коммуникации к интернет-пространству. Существенной особенностью при перераспределении DNS трафика пользователей является снижение роли корневых DNS серверов. В результате проверки разработанного алгоритма предложено перенаправлять DNS трафик через DNS прокси-сервера различного класса анонимности (рис. 3). Это позволило исключить возможность накопления DNS запросов пользователей у провайдеров. Преимуществом предложенного алгоритма является изменение маршрута DNS запросов от клиента DNS на заранее выбранный DNS прокси-сервер с максимально возможным классом анонимности (рис. 4). Класс анонимности DNS прокси-серверов определяется посредством разработанного мультитеста на соответствие критериям тестирования (табл. 2). Второй модуль разработанного алгоритма осуществляет блокировку соединений между клиентами DNS устройства коммуникации и специализированными интернет-сервисами сбора данных. Также блокируются соединения со сторонними сервисами и сервисами разработчиков системного и прикладного программного обеспечения (рис. 5). Сочетание двух модулей предложенного алгоритма позволило пользователям выбирать уровень своей приватности при взаимодействии с интернет-пространством.

Выполненный комплексный аудит TCP/UDP трафика различных устройств коммуникации позволил выявить сервисы IT-компаний, выполняющие сбор данных пользователя (табл. 3).

Выполнена проверка предложенного алгоритма на отсутствие утечек DNS трафика с устройства коммуникации. Ее результаты показали отсутствие утечек DNS трафика при использовании произвольно выбранного HIA класса DNS прокси-сервера (табл. 4).

Таким образом, сформулированная в исследовании проблема была решена с помощью разработанного алгоритма защиты устройств коммуникации от несанкционированных сбора и утечек данных пользователей в сети Интернет. Сочетание перенаправления DNS трафика устройств коммуникации через DNS прокси-сервера и одновременная фильтрация TCP/UDP трафика в этом алгоритме является преимуществом выполненного исследования по сравнению с рассмотренными работами по данной тематике [23–26]. При этом применение алгоритма блокировки утечек данных с устройств коммуникации показало отсутствие потерь работоспособности системного и прикладного программного обеспечения. Пользователи получили возможность самостоятельно выбирать уровень приватности, управляя сбором своих данных при выполнении каких-либо действий в интернет-пространстве в режиме реального времени.

К недостаткам предложенного алгоритма следует отнести реализацию процесса последовательного сканирования каждого из DNS прокси-серверов, что приводит к временной задержке перед началом его эксплуатации, которая определена экспериментально и составляет от 300 до 900 секунд и зависит от

количества DNS прокси-серверов полученных из открытых интернет-ресурсов. В свою очередь это обуславливает невозможность мгновенного обеспечения требуемого уровня приватности пользователя из-за фактического отсутствия протестированных и отсортированных NOA, ANM, HIA DNS прокси-серверов.

Кроме этого, процесс тестирования DNS прокси-серверов увеличивает суммарный объем DNS трафика, генерируемого устройством коммуникации, что может быть неприемлемым для пользователей, оплачивающих фиксированный объем интернет-трафика.

Сокращение общего времени тестирования DNS прокси-серверов может быть достигнуто за счет организации процесса многопоточного (параллельного) их сканирования. Причем сокращение общего времени тестирования DNS прокси-серверов будет уменьшаться прямо пропорционально увеличению количества потоков тестирования.

Дальнейшие перспективы совершенствования работы предложенного алгоритма могут включать:

– внедрение функции скрытия или замены идентификатора User-Agent для клиентов DNS, выполняющих коммуникацию по HTTP протоколу;

– внедрение функции установки времени проверки для тестируемого DNS прокси-сервера;

– внедрение функции распознавания DNS прокси-серверов AnchorFree, CoDeen, TinyProxy, принадлежащих IT-компаниям, оказывающим услуги приватного серфинга;

– внедрение функции исключения обнаруженных DNS прокси-серверов AnchorFree, CoDeen, TinyProxy из списка рабочих серверов.

Реализация этих функций позволит сократить время тестирования DNS прокси-серверов и повысить приватность пользователей.

8. Выводы

1. Проанализирован процесс обмена данными между клиентами DNS и интернет-сервисами, с которыми они взаимодействуют. Исследование схемы обмена данными между устройством коммуникации и интернет-пространством позволило выявить пути утечек данных с устройств коммуникации. Поскольку все запросы клиентов DNS накапливаются в журналах DNS сервера провайдера, то анализ DNS запросов позволяет сформировать цифровой профиль устройства коммуникации.

2. Разработан алгоритм блокировки утечек данных, собираемых разработчиками программного обеспечения, установленного на устройстве коммуникации, для предоставления пользователям возможности выбора уровня своей приватности. Практическое применение разработанного алгоритма позволило исключить журналирование DNS трафика интернет-провайдерами и тем самым заблокировать сбор данных пользователей с устройств коммуникации. Предложенный алгоритм позволяет существенно снизить точность цифрового профилирования устройств коммуникации пользователя. Существенным преимуществом является возможность предоставления пользователю выбора желаемого уровня приватности в интернет-пространстве.

3. Выполнен аудит TCP/UDP трафика с различных устройств коммуникации в течение длительного времени. Анализ позволил определить домены и IP-адреса интернет-ресурсов, к которым обращается системное и прикладное программное обеспечение устройств коммуникации. Интернет-сервисы сбора и монетизации данных, выполняющие запросы на передачу данных пользователей, систематизированы в соответствии принадлежности IT-компаниям.

4. Проверка предложенного алгоритма на отсутствие утечек DNS трафика с устройства коммуникации показала отсутствие потерь работоспособности системного и прикладного программного обеспечения. Выборочная блокировка интернет-трафика выполнялась посредством настройки списка запрещенных IP-адресов сетевого экрана в соответствии с экспериментально полученными данными.

Литература

1. García-Dorado J., Ramos J., Rodríguez M., Aracil J. (2018). DNS weighted footprints for web browsing analytics. Journal of Network and Computer Applications, 111, 35–48. doi: 10.1016/j.jnca.2018.03.008. Available at: http://www.sciencedirect.com/science/article/pii/S1084804518300894.

2. Guelke, J. (2020). Leaking. International Encyclopedia of Ethics, 6, 1–7. doi: 10.1002/9781444367072.wbiee898. Available at: https://www.researchgate.net/publication/342893197_Leaking.

3. Trish, B. (2018). Big Data under Obama and Trump: The Data-Fueled U.S. Presidency. *Politics and Governance*, 6 (4), 29–38. doi: 10.17645/pag.v6i4.1565. Available at: https://www.researchgate.net/publication/329111791_Big_Data_under_Obama_and_Trump_The_Data-Fueled_US_Presidency.

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: 10.1093/idpl/ipw026. Available at: https://academic.oup.com/idpl/article-abstract/7/1/36/3097625?redirectedFrom=PDF.

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. Generating private data with user customization. (2020). 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:

10.1109/TSE.2017.2685387. Available at: https://www.researchgate.net/publication/313796770_Understanding_Diverse_Smarpthone_Usage_Patterns_from_Large-Scale_Appstore-Service_Profiles.

9. Stachl, C., Quay, A., Schoedel, R., Gosling, S., Harari, G., Buschek, D., Völkel, S., Schuwerk, T., Oldemeier, M., Ullmann, Th., Hussmann, H., Bischl, B., Buehner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (30), 17680–17687. doi: 10.1073/pnas.1920484117. Available at: https://www.pnas.org/content/117/30/17680.

10. Waheed, H., Anjum, M., Rehman, M., Khawaja, A. (2017). Investigation of user behavior on social networking sites. *PloS one*, 12(2): e0169693. doi: 10.1371/journal.pone.0169693. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5289435.

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: 10.5604/01.3001.0013.0905. Available at: https://www.researchgate.net/publication/331868427_ALGORITHM_OF_USER%27S_PERSONAL_DATA_PROTECTION_AGAINST_DATA_LEAKS_IN_WINDOWS_10_OS.

12. Raber, F., Vossebein, N. (2017). URetail: Privacy User Interfaces for Intelligent Retail Stores. In: Bernhaupt, R., Dalvi, G., Joshi, A., Balkrishan, D., O'Neill, J., Winckler, M. Human-Computer Interaction INTERACT 2017. Lecture Notes in Computer Science (Springer, Cham), 10516, 473–477. doi: 10.1007/978-3-319-68059-0_54. Available at: https://www.researchgate.net/publication/319949798_URetail_Privacy_User_Interfaces_for_Intelligent_Retail_Stores.

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: 10.1016/j.cose.2018.01.018. Available at: https://www.researchgate.net/publication/318115818_Towards_Secure_Name_Resolution_on_the_Internet.

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, CA, USA), 494–499, doi: 10.1109/ICICT50521.2020.00085. Available at: https://ieeexplore.ieee.org/document/9092077/authors#authors.

16. Zhiwei, Y., Jong-Hyouk, L. (2020). The road to DNS privacy. *Future Generation Computer Systems*, 112, 604–611. doi: 10.1016/j.future.2020.06.012. Available at: https://www.researchgate.net/publication/342099404_The_road_to_DNS_privacy/stats.

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, California, USA), 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, USA), 290–304. doi: 10.1145/3320269.3384728. Available at: https://www.researchgate.net/publication/337019716_Assessing_the_Privacy_Benefits_of_Domain_Name_Encryption.

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)*, pp. 138–143. doi: 10.1145/3359989.3365435. Available at: https://www.researchgate.net/publication/337780186_DNS_privacy_in_practice_and_preparation.

20. Белявский, Д. (2015). DNS: кто не спрятался, тот и виноват. *Интернет в цифрах*, 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, USA), 123–137. doi: 10.1145/3359989.3365429. Available at: https://www.researchgate.net/publication/337777770_An_investigation_on_information_leakage_of_DNS_over_TLS.

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*, 1–9. doi: 10.2139/ssrn.3427563. Available at: https://www.researchgate.net/publication/335016692_How_DNS_over_HTTPS_is_Reshaping_Privacy_Performance_and_Policy_in_the_Internet_Ecosystem.

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. Шелухин, О., Смычёк, М., Симонян, А. (2018). Фильтрация нежелательных приложений интернет-ресурсов в целях информационной безопасности. *Наукоемкие технологии в космических исследованиях Земли*, 10 (2), 87–98. doi: 10.24411/2409-5419-2018-10044. Available at: https://www.elibrary.ru/item.asp?id=34939631.

25. Smart DNS Proxy Servers. Available at: https://www.smartdnsproxy.com/Servers.

26. Подкорытов, Д., Флока, А., Кулешов С. (2019). Архитектура кроссплатформенного DNS Proxy сервиса. *T-Comm: Телекоммуникации и транспорт*, 13 (5), 35–40. doi: 10.24411/2072-8735-2018-10269. Available at: https://www.researchgate.net/publication/333844552_Podkorytov_DA_Floka_AB_K

ulesov_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: 10.1002/9781119692263.ch4. Available at: https://www.researchgate.net/publication/347877088_Navigating_the_Internet_with_DNS.

28. Fujiwara, K., Sato, A., Yoshida, K. (2018). Cache Effect of Shared DNS Resolver. *IEICE Transactions on Communications*, E102.B (6). doi: 10.1587/transcom.2018EBP3184. Available at: https://www.researchgate.net/publication/329361930_Cache_Effect_of_Shared_DNS_Resolver.

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, NV, USA, pp. 1550–1551, doi: 10.1109/CSCI49370.2019.00291. Available at: https://ieeexplore.ieee.org/document/9071252.