СПОСОБЫ ЗАЩИТЫ ОТ СОВРЕМЕННЫХ МЕТОДОВ СБОРА ДАННЫХ ПОЛЬЗОВАТЕЛЕЙ ИНТЕРНЕТ-САЙТОВ БЕЗ ИХ СОГЛАСИЯ

METHODS OF PROTECTION AGAINST MODERN METHODS OF DATA COLLECTION OF USERS OF INTERNET SITES WITHOUT THEIR CONSENT

O. Rubin
A. Kharisov

Summary. In recent years issues connected with security of personal data of information systems` users have become one of the most important in the internet technologies area. The outcome of this article is an observation of technical and law aspects of unauthorized collection of internet sites users` personal data. The paper also provided with several examples of how web sites can collect user's data without any notification or agreement and ways how this data collection can be blocked.

Keywords: cybersecurity, internet, anonymity, proxy server, data collection, personal data.

Рубин Олег Ильич
Уральский Федеральный Университет им. Б. Н. Ельцина, ИРИТ РФ
olegrubin97@gmail.com
Харисов Азамат Робертович
К.т.н., доцент, Уральский Федеральный Университет им. Б. Н. Ельцина, ИРИТ РФ
t2201111@yandex.ru

Аннотация. В наше время вопрос конфиденциальности персональных данных пользователей тех или иных информационных систем, в том числе и веб-сайтов, является одним из наиболее важных и актуальных для сферы информационных технологий во всем мире. В настоящей статье рассмотрены технические и правовые аспекты несанкционированного сбора личных данных пользователей интернет-сайтов. Приведены примеры способов сбора данных без уведомления и согласия со стороны пользователя и методы их блокировки и защиты от такого сбора.

Ключевые слова: информационная безопасность, интернет, анонимность, прокси-сервер, сбор данных, персональные данные.

Введение

В настоящий момент на большинстве интернет-сайтов производится сбор тех или иных данных о посетителях этих сайтов. Данные собираются с совершенно различными целями, среди которых:

- Формирование статистики о половозрастных и иных характеристиках аудитории сайта для оценки ее релевантности относительно содержимого сайта;
- Сбор данных о посещении конкретными пользователями ресурса для дальнейшей продажи или передачи этих данных третьим лицам;
- Осуществление спам-рассылок по собранным базам данных посетителей сайта, в том числе без их согласия;
- Иные цели;

При этом методы сбора данных можно разделить на два типа по наличию явного согласия на осуществления сбора этих данных со стороны посетителя сайта — санкционированный и несанкционированный методы сбора данных. Наиболее опасным является несанкционированный сбор данных, поскольку он может быть осуществлен без оповещения и согласия на такой сбор посетителя сайта.

Несанкционированный сбор данных осуществляется как посредством сторонних сервисов, работающих для разработчика сайта по принципу черного ящика (например, сервисы «Яндекс.Метрика» и «Google Analytics»), так и с помощью собственных решений.

Сторонние сервисы, как правило, предоставляют очень ограниченный набор данных по отдельным посетителям сайта, но широкий спектр данных для статистического анализа аудитории сайта. Собственные же решения позволяют разработчикам без каких-либо санкций со стороны пользователя сохранять, копировать и распространять те данные о нем, которые они смогли получить. Однако собственные решения для несанкционированного сбора данных имеют множество ограничений. Некоторые из таких ограничений автоматически накладываются браузерами, другие пользователь может настроить сам.

В настоящей статье рассмотрены некоторые уязвимости современных браузеров и не только, позволяющие

осуществлять несанкционированный сбор данных, а также способы блокировки такого сбора пользователями.

Понятия личных данных и анонимности

Под личными данными пользователя сайта в настоящей статье следует понимать любые данные, прямо или косвенно относящиеся к конкретному пользователю сайта, который идентифицируется по посещению сайта с использованием определенного IP-адреса (подразумевается именно IP адрес, переданный в качестве идентификатора отправителя запроса на сетевом уровне). Примеры личных данных пользователя:

- Данные об используемом пользователем браузере — тип браузера, браузерное время и так далее;
- Данные об операционной системе и оборудовании пользователя — размер монитора, используемая операционная система и прочее;
- Персональные данные пользователя (возможно, обезличенные) — фамилия, имя, отчество, дата рождения;
- Контактные данные пользователя — телефон, адрес электронной почты;
- Иные данные, прямо или косвенно относящиеся к конкретному пользователю сайта.

Соответственно, основываясь на таком определении личных можно дать определение анонимности пользователя интернет-сайта — это степень сокрытия его личных данных от возможного сбора со стороны интернет-сайта. В то же время пользователь является анонимным только тогда, когда его нельзя идентифицировать по ранее собранным личным данным. Например, сайт может получить следующие данные: наименование браузера пользователя, ширина и высота монитора, используемая операционная система, временная зона пользователя и другие данные; в дальнейшем такая совокупность информации может быть использована для идентификации пользователя даже при условии смены им IP-адреса для доступа к сайту.

Кроме того, наборы личных данных различной степени точности могут быть сохранены в базе данных и в последствии переданы его разработчиками третьим лицам без какого-либо согласия пользователя для использования в любых целях.

С точки зрения законодательства Российской Федерации однозначно запрещенными к несанкционированному сбору и использованию без согласия со стороны субъекта данных являются персональные данные граждан РФ. При этом, согласно федеральному закону 152-ФЗ «О персональных данных», согласие на сбор, обработку, хранение и передачу третьим лицам требуется только для тех наборов данных, которых достаточно для однозначной идентификации субъекта персональных данных. Например, паспортные данные — серия и номер — являются уникальным идентификатором гражданина РФ и позволяют однозначно его определить. Но, предположим, что в системе есть следующий набор данных:

- IP-адрес;
- Фамилия, имя, отчество;

Такой набор данных не позволяет однозначно идентифицировать гражданина Российской Федерации, поскольку существуют полные тезки, ФИО которых совпадают, а IP-адрес может быть зарегистрирован у провайдера на любое физическое лицо, не обязательно то, которое его использует для доступа к сайту. Следовательно, такой набор данных с точки федерального закона № 52 «О персональных данных» является обезличенным, то есть не позволяющим установить личность гражданина РФ. Соответственно, даже несанкционированный сбор и использование в любых целях таких данных не запрещены законодательством Российской Федерации.

Из всего вышеперечисленного следует, что по крайней мере в Российской Федерации забота об анонимности в интернете — прерогатива пользователя и если какие-то данные о нем были собраны, то нести ответственность за это нести может только сам пользователь.

Методы защиты от несанкционированного сбора данных

Утечка локального IP адреса через протокол WebRTC

WebRTC — протокол, предназначенный для организации передачи потоковых данных между браузерами или другим поддерживающим его программным обеспечением по технологии «peer to peer». Как правило, данный протокол используется для захвата и потоковой передачи видео- и аудиоинформации во время звонков с использованием IP-телефонии. Аббревиатура «RTC» в названии протокола происходит от английского словосочетания «Real Time Communications», что дословно означает «коммуникация в режиме реального времени».

Данные протокол используется как в общедоступном программном обеспечении, так и в корпоративных сетях, построенных по принципу «Интранет», для коммуникации по видео- и аудиосвязи между сотрудниками. Среди примеров программного обеспечения, использующего данный протокол, можно выделить: «Skype», «Google Hangouts», а также большинство современных браузеров и мессенджеров с возможностью осуществлять аудиосвязь.

Однако протокол содержит одну особенность — программное обеспечение, использующее его, обязано узнать IP-адрес пользователя в локальной сети. В том числе и в браузерах получение локальных IP-адресов как версии 4, так и версии 6 возможно посредством выполнения определенных сценариев на языке Javascript в том случае, если на стороне пользователя не заблокировано использование протокола WebRTC. Для большинства пользователей глобальной сети «Интернет» утечка локальных IP-адресов не является проблемой — ведь их доступ к сайтам осуществляется в подавляющем большинстве случаев по протоколам HTTP или HTTPS, при использовании которых виден IP-адрес маршрутизатора, а не конечного устройства.

Тем не менее, в корпоративных сетях вместе с другими данными утечка локальных адресов может быть критичной и, в том числе, противоречащей политике безопасности компании. К примеру, данные о локальном IP-адресе пользователя внутри корпоративной сети банка могут дать злоумышленнику возможность понять, из какого подразделения банка осуществлялся доступ к сайту, а также в дальнейшем возможность осуществить атаку на конкретную подсеть подразделения банка по имеющимся данным.

Заблокировать доступ к WebRTC в браузере «Google Chrome» можно, например, посредством установки официального бесплатного и свободно распространяемого плагина «Web RTC Control». При активации данного плагина несанкционированное подключение браузера к компьютеру и подключенному к нему видео- и аудио-оборудованию по протоколу WebRTC будет невозможным.

Утечка реального IP-адреса при использовании прозрачных прокси

Основной идентификатор пользователя интернет-сайта — это его IP-адрес. Данный факт обусловлен тем, что IP-адрес сайт получит в любом случае — ведь на сетевом уровне в сети «Интернет» устройства коммуницируют исключительно по протоколам «IPv4» и «IPv6». Для сокрытия IP-адреса, как правило, используются прокси-серверы, которые принимают запросы пользователей и пересылают их на сайт, к которому осуществляется доступ. При этом на сетевом уровне в качестве IP-адреса отправителя запроса сайт определяет именно IP-адрес прокси, реальный IP-адрес пользователя недоступен. Используемые прокси можно разделить на три группы:

- Транспарентные или прозрачные прокси. Такие прокси-серверы, хоть и позволяют скрыть IP-адрес на сетевом уровне, на уровне приложения и заголовков HTTP-сообщения сообщают конечному узлу — например, интернет-сайту — о том, кто изначально сделал запрос. Как правило, исходный IP-адрес отправителя записывается прокси-сервером в заголовок «X-Forwarded-For» согласно стандарту IETF (Инженерного Совета Интернета);
- Анонимные прокси. Такие прокси-серверы не добавляют заголовок «X-Forwarded-For» в HTTP-запрос, однако добавляют другие заголовки, по которым сайт может автоматически принять решение о том, что пользователь использует прокси и, например, отказать в доступе, в связи с тем, что анонимные пользователи на сайте запрещены его политикой. К примеру, анонимный прокси может добавить заголовок «Via», в котором указать собственный IP-адрес и программное обеспечение, используемое для пересылки запросов (например, это автоматически происходит в ПО «MikTeX Proxy»).
- «Элитные» прокси. Данный вид прокси-серверов никак не модифицирует HTTP-запросу, отправляя их со своего IP-адреса так, как эти запросы были бы выполнены с исходного IP-адреса. Результат — конечный узел, в том числе любой интернет-сайт, не будет иметь возможности определить ни реальный IP-адрес исходного запроса, ни факт использования прокси-сервера. Как правило, услуги таких прокси предоставляются на платной основе.

Ниже приведены примеры запросов и полученных сайтом заголовков при использовании транспарентного, анонимного и элитного прокси. Доступ осуществлялся по HTTPS.

Тем не менее, при отправке HTTPS-запросов даже прозрачные прокси не имеют возможности добавить заголовки в запрос, поскольку они все шифруются по алгоритмам, определенным в ходе SSL-рукопожатия между клиентом и конечным сайтом и неизвестным для прокси-сервера. Это означает, что при посещении сайтов, работающих по HTTPS, реальный IP-адрес будет сокрыт даже при использовании транспарентного прокси.

Однако есть одна уязвимость, позволяющая разработчику сайта обойти это ограничение даже на сайте, доступном по HTTPS. Формально в браузерах запрещена отправка незашифрованных HTTP-запросов из пользовательских скриптов на сайте, доступ к которому осуществляется по HTTPS. Содержимое, запрашиваемое таким способом, называется смешанным (от англ. «Mixed Content»). Однако это ограничение касается только запросов активного смешанного содержимого — запросов других страниц сайта, скриптов и так далее. Существует также пассивное смешанное содержимое — изображе-

Рис. 1. Отправка запроса без использования прокси-сервера, определен реальный ip-адрес

ния, аудио, видео и другие файлы, которые могут быть указаны как источники данных для HTML-элементов и его запрос в незашифрованном виде сайтов, доступных по HTTPS не запрещен, хоть и не рекомендуется в официальной документации.

Например, когда в теге «img» на веб-странице в атрибуте «src» указывается ссылка на изображение, при загрузке этой страницы оно загружается отдельным HTTP-запросом и, если указать адрес этого изображения именно через «http://», то запрошен он будет без шифрования, причем со стороны браузера, а, следовательно, с использованием пользовательских настроек прокси-сервера. Таким образом, при использовании транспарентного или анонимного прокси-сервера пользователем сайтом может быть осуществлен несанкционированный доступ к реальному IP-адресу пользователя или может быть обнаружен факт использования им прокси, что может послужить основанием для автоматического принятия решения об отказе в обслуживании.

Пассивная утечка данных из социальной сети «ВКонтакте» через программный интерфейс «VK Open API»

Социальная сеть «ВКонтакте» насчитывает более 536 миллионов пользователей по состоянию на 21 марта 2019 года. Разработчики социальной сети, помимо ее пользовательского интерфейса, активно занимаются развитием программного интерфейса (API — Application Programming Interface), цели которого:

- Дать возможность разработчикам сторонних сайтов развивать программное обеспечение, предоставляющее после авторизации по протоколу «Oauth 2.0» пользователям социальной сети дополнительную функциональность;
- Дать возможность разработчикам другого программного обеспечения анализировать и использовать в своих целях информацию о пользователях и другую информацию из социальной сети «ВКонтакте», находящуюся в открытом доступе.

Один из инструментов API «ВКонтакте» — «VK Open API» позволяет получать данные о пользователе «ВКонтакте» в пассивном режиме, то есть без согласия получение этих данных со стороны пользователя. Для того, чтобы данные были получены сайтом, его разработчику необходимо:

- Создать приложение на портале разработчиков «ВКонтакте»;
- Настроить данное приложение для работы с доменом сайта;

Рис. 2. Отправка запроса через прозрачный прокси-сервер, определен реальный IP-адрес (из заголовка «X-Forwarded-For»), а также факт использования прокси

- Указать на сайте в HTML-коде веб-страницы ссылку на скрипт «VK Open API», доступную в документации API «ВКонтакте».

Затем, если пользователь авторизован в рамках текущей сессии браузера, при осуществлении доступа к сайту и выполнении пользовательских сценариев Javascript на запрашиваемой веб-странице может быть получен идентификатор пользователя в социальной сети «ВКонтакте», по которому, в свою очередь, посредством других методов API могут быть запрошены и другие данные, среди которых могут быть:

- Фотография пользователя;
- ФИО;
- Контактные данные;
- Иные сведения, находящиеся в публичном доступе, например, место работы, место учебы, интересы и так далее.

Причем подобное получение данных без согласия пользователя, исходя из описанных выше в статье норм права, не противоречит законодательству Российской Федерации, а также правилам использования API «ВКонтакте», опубликованным на официальном сайте. Для того, чтобы избежать подобной утечки, пользователь должен установить расширения для браузера, блокирующие загрузку скрипта «VK Open API», позволяющего извлекать идентификатор пользователя «ВКонтакте» из сессии браузера без согласия или установить браузер, встроенная функциональность которого включает в себя блокировку такого содержимого сайтов (например, «Brave» или «Tor Browser»).

Заключение

На сегодняшний день существует множество способов несанкционированного сбора сайтами различных

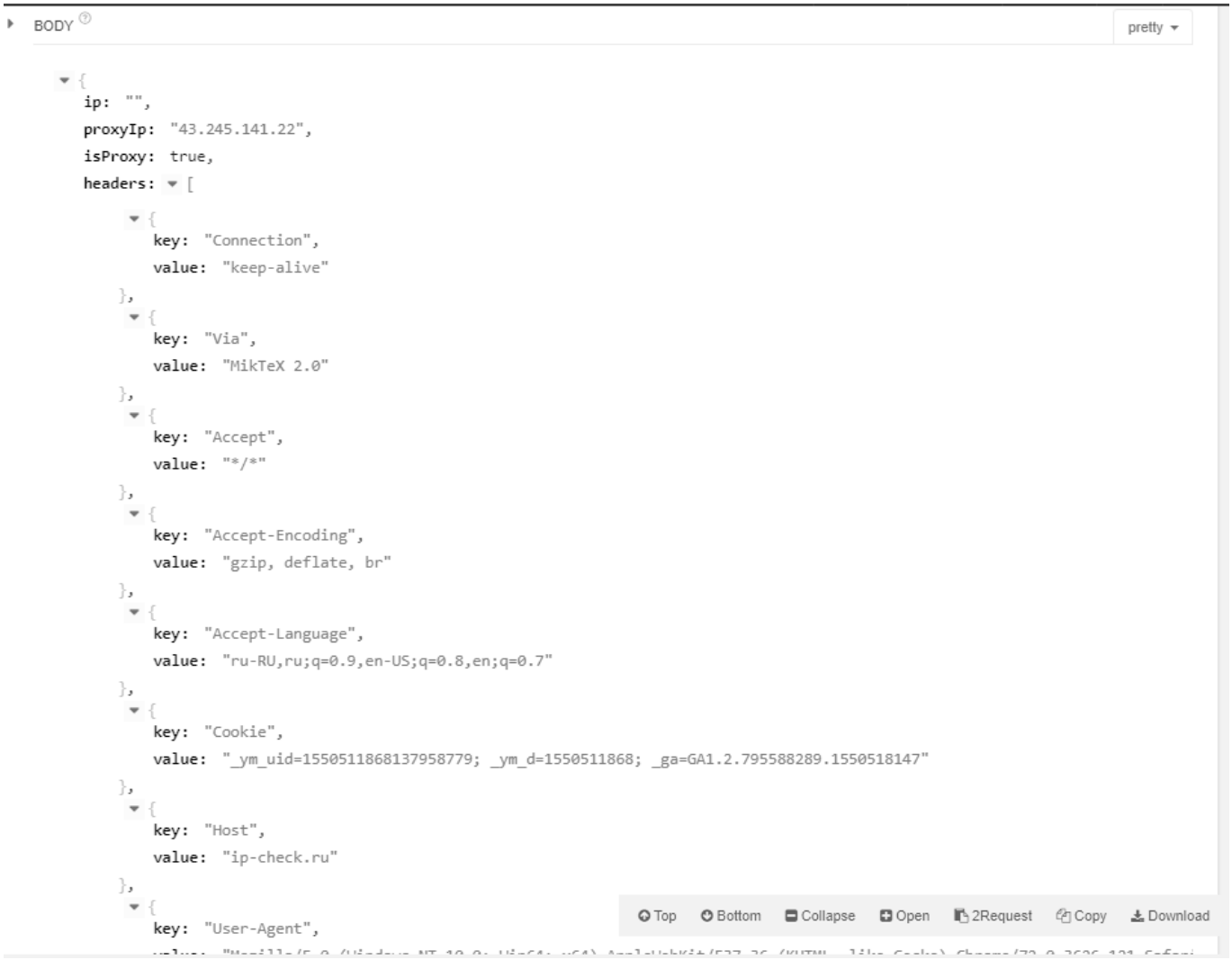

Рис. 3. Видимые заголовки запроса при отправке его через анонимный прокси — реальный IP-адрес пользователя не определен, однако по наличию заголовка «Via» определен факт использования прокси-сервера

данных о посетителях. Последствия такой утечки данных могут непредсказуемы, поскольку если данные собираются без уведомления и без явного согласия на их сбор пользователя, владельцы сайта также без ведома пользователя могут бессрочно хранить и как угодно распространять собранные данные, не нарушая при этом действующее законодательство Российской Федерации.

В настоящей статье было приведено определение личным данным и анонимности пользователей сайтов, было рассмотрено несколько использующихся на сегодняшний день способов несанкционированного сбора данных, а также методы их блокировки и защиты от такого сбора. Показано, что при условии осведомленности пользователя в вопросах собственной кибербезопасности вполне возможно не допускать сбора личных данных и оставаться анонимным при посещении сайтов в сети «Интернет», а в частности были сделаны следующие выводы:

- Для сокрытия IP-адреса следует использовать только категорию «элитных» прокси, поскольку только они не добавляют в HTTP-запрос свои специализированные заголовки, из которых можно узнать либо факт использования прокси, либо, что еще хуже, реальный IP-адрес отправителя запроса;
- Для того, чтобы не допустить утечки локального IP-адреса по протоколу WebRTC, следует установить и активировать расширение для используемого веб-браузера, позволяющее заблокировать подключения по WebRTC к компьютеру без явного согласия пользователя;
- Во избежание пассивного получение данных сайтом о пользователе из социальной сети «ВКон-

```
{
  ip: "90.63.12.204",
  proxyIp: "",
  isProxy: false,
  headers: [
    {
      key: "Connection",
      value: "keep-alive"
    },
    {
      key: "Accept",
      value: "*/*"
    },
    {
      key: "Accept-Encoding",
      value: "gzip, deflate, br"
    },
    {
      key: "Accept-Language",
      value: "ru-RU,ru;q=0.9,en-US;q=0.8,en;q=0.7"
    },
    {
      key: "Cookie",
      value: "_ym_uid=1550511868137958779; _ym_d=1550511868; _ga=GA1.2.795588289.1550518147"
    },
    {
      key: "Host",
      value: "ip-check.ru"
    },
    {
      key: "User-Agent",
      value: "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/72.0.3626.121 Safari
    }
  ]
}
```


Рис. 4. Результат отправки запроса через «элитный» прокси — по заголовкам не определен ни факт использования прокси-сервера, ни реальный IP-адрес

такте» следует либо выйти из нее в рамках текущей сессии браузера, либо скрыть свой профиль в ней, что позволяют сделать настройки безопасности профиля в данной социальной сети, либо использовать расширения для браузера, позволяющие заблокировать скрипт «VK Open API», предоставляющий разработчикам сайтов возможность пассивного получения данных.

ЛИТЕРАТУРА

1. Hojung Cha, Jongmin Lee. Replacing media caches in streaming proxy servers // Journal of Systems Architecture, Vol. 52, Issue 1, January 2006, Pages 25–40.
2. N. Gautam. Performance analysis and optimization of web proxy servers and mirror sites // European Journal of Operational Research, Volume 142, Issue 2, 16 October 2002, Pages 396–418.
3. Forwarded HTTP Extension // Internet Engineering Task Force URL: https://tools.ietf.org/html/rfc7239 (дата обращения: 10.02.2019).
4. Федеральный закон от 27.07.2006 N152-ФЗ «О персональных данных» // СПС КонсультантПлюс // В последней редакции опубликован 13.12.207 на официальном интернет-портале правовой информации http://www.pravo.gov.ru
5. Технология WebRTC // 3CX URL: https://www.3cx.ru/webrtc/ (дата обращения: 03.03.2019).
6. Open API // Документация API социальной сети «ВКонтакте» URL: https://vk.com/dev/openapi (дата обращения: 01.02.2019).
7. Google open source WebRTC for open video and audio chat // The H Open URL: http://www.h-online.com/open/news/item/Google-open-source-WebRTC-for-open-video-audio-chat-1253848.html (дата обращения: 25.02.2019).
8. Что такое WebRTC и как это отключить // The Safety URL: https://thesafety.us/ru/what-is-webrtc (дата обращения: 01.03.2019).

Журнал «Современная наука: актуальные проблемы теории и практики»