АНАЛИЗ СИСТЕМ АВТОНОМНОЙ ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ САЙТА

ANALYSIS OF SYSTEMSAUTONOMOUS SITE USER IDENTIFICATION

УДК 004.62

Алисултанова Э. Д., доктор педагогичеких наук, профессор, директор института прикладных информационных технологий, Грозненский государственный нефтяной технический университет

Исаева М. З., старший преподаватель кафедры «Информатика и вычислительная техника», Грозненский государственный нефтяной технический университет

Болтиев Д. У., студент, Грозненский государственный нефтяной технический университет

Alisultanova E.D., esmira59@mail.ru

Isaeva M.Z., isaeva.mareta@mail.ru

Baltiev D.U., ugodayconzass1995@gmail.com

Аннотация

В этой статье анализируются системы автономной идентификации, которые алгоритмически фиксируют уникального пользователя сети интернет во время посещения им определенного сетевого информационного ресурса. Представлены характерные особенности наиболее значимых для администрирования сетей технологий идентификации, таких, как IP, отпечаток браузера и cookie. Существуют и другие идентификаторы пользователя, каждый из этих идентификаторов отличается друг от друга тем, что один идентификатор уникален, но менее надежен, другой наоборот, более

надежен, но менее уникален, поэтому необходимо использовать особую логику конфигурации этих данных, о чем и говорится в данной статье. Рассмотрены ключевые особенности при идентификации пользователя сети. Обосновывается необходимость как унификации данных, так и повышение уровня надежности, чтобы минимизировать возможности нарушений этики информационной безопасности. Проведен общий анализ основных технологий, на базе которых функционирует данная система идентификации пользователя.

Annotation

This article analyzes autonomous identification systems that algorithmically record a unique Internet user during his visit to a certain network information resource. The characteristic features of the most significant identification technologies for network administration, such as IP, browser fingerprint and cookie, are presented. There are other user identifiers, each of these identifiers differs from each other in that one identifier is unique, but less reliable, the other, on the contrary, is more reliable, but less unique, therefore, it is necessary to use special logic for configuring this data, which is what is said in this article.

Key features of the identification of a network user are considered. The necessity of both data unification and increasing the level of reliability is substantiated in order to minimize the possibility of violations of information security ethics. A general analysis of the main technologies on the basis of which this user identification system functions.

Ключевые слова: идентификация пользователя, IP- адрес, отпечаток браузера, файлы cookie, технологии, web – программа, скрипты, интернет-протокол.

Keywords: user identification, IP address, browser fingerprint, cookies, technology, web program, scripts, internet protocol.

Идентификация посетителя сайта без ввода каких-либо данных от пользователя достаточно трудно, а если говорить об абсолютной идентификации, то и вовсе невозможна. Пользователя можно идентифицировать с какой-то вероятностью, система, предложенная в этой статье, достаточно точно определяет пользователя сайта.

Система автономной идентификации представляет собой веб программу в виде таблицы, к которой подвязывается целевой сайт для идентификации посетителей этого сайта и введения их отчета действий. Идентификация посетителя осуществляется посредством сочетания различных элементов идентификации и созданием на основе этой комбинации идентификатора. Для определения, является ли пользователь уникальным или это его повторный заход, будет производиться сравнение идентификатора пользователя с идентификаторами из базы, при истинности сравнения, пользователь будет определен как ранее зашедший в противном случае как «уникальный».

К этой системе можно подключать несколько сайтов, для анализа активности, и идентификации пользователей. Так, при подключении к этой системе несколько сайтов, можно будет определить общих пользователей этих сайтов.

В этой статье представлены лишь основные технологии идентификации, такие как IP, отпечаток браузера и *cookie*. Существуют и другие идентификаторы пользователя, каждый из этих идентификаторов отличается друг от друга тем, что один идентификатор уникален, но менее надежен, другой на оборот, более надежен, но менее уникален, поэтому нужно продумать логику конфигурации этих данных, о чем и говорится в этой статье.

Отпечатки браузера (также называемые отпечатками пальцев устройства или онлайн-отпечатками пальцев) относятся к методам отслеживания, которые веб-сайты используют для сбора информации. Современные функции веб-сайтов требуют использования скриптов - наборов инструкций, которые сообщают вашему браузеру, что делать. Работая незаметно в фоновом режиме, скрипты могут идентифицировать большой

объем информации устройстве и браузере, которая, будучи сшита вместе, образует уникальный онлайн-отпечаток. Затем этот отпечаток браузера можно отследить через Интернет и в различных сеансах просмотра.

Что именно могут обнаружить скрипты? Они могут определить об используемом устройстве, например, его операционную системе, браузер на устройстве, программное обеспечение, язык и местонахождение отслеживаемого субъекта, используется ли блокировщик рекламы, разрешение и глубина цвета экрана, все установленные расширения браузера и более подробные технические характеристики видеокарты, драйверов и т.п.

Аналогичным образом, отпечатки браузера предоставляют достаточно конкретных атрибутов вашего устройства и его настроек, чтобы вас можно было надежно идентифицировать из толпы, даже из чрезвычайно большой группы миллионов пользователей интернета и миллиардов устройств.

Снятие отпечатков пальцев браузера может использоваться для предотвращения попыток мошенников взломать, спамить или обмануть владельцев веб-сайтов путем точной идентификации пользователей сайта. Отпечатки браузера труднее обойти, чем файлы cookie, поскольку отпечаток пальца пользователя не меняется между сеансами просмотра в инкогнито или очисткой данных браузера. Чтобы сгенерировать отпечаток браузера с достаточной точностью (или энтропией), чтобы однозначно идентифицировать веб-посетителя, сценарий должен использовать различные методы отпечатка пальца браузера для сбора данных (называемых сигналами), которые будут различаться для разных посетителей. Хотя у многих посетителей веб-сайта может быть одна и та же модель iPhone, установленное программное обеспечение и драйверы, геолокация, версия браузера и ОС и даже незначительные отклонения в оборудовании могут отличаться.

Отпечатки браузера работают, потому что веб-сайты используют сценарии, которые выполняются в фоновом режиме вашего браузера. Современные веб-браузеры имеют встроенные программные функции, называемые API, которые могут использоваться скриптами веб-сайтов для

сбора информации. Как правило, скрипты предназначены для законных целей, таких как рендеринг видео или фотографий. Скрипты собирают атрибуты - спецификации устройства, ОС, настройки и плагины браузера, пользовательские агенты, возможности аудио и видео, часовой пояс и многое другое - которые могут быть скомпилированы в «хэш» или цифровой отпечаток.

Многие владельцы веб-сайтов и рекламные сети используют функцию снятия отпечатков в браузере для межсайтового отслеживания. Подавляющее большинство использует эти данные для рекламы и персонализации вашего в Интернете, чтобы на основе собранной информации, полученной в результате исследования, производить продукты или услуги, которые будут отвечать интересам потенциальных потребителей.

Файлы cookie - это текстовые файлы с небольшими фрагментами данных, такими как имя пользователя и пароль, которые используются для идентификации вашего компьютера, когда вы используете компьютерную сеть. Конкретные файлы cookie, известные как файлы cookie HTTP, используются для идентификации конкретных пользователей и улучшения вашего опыта просмотра веб-страниц.

Данные, хранящиеся в файле cookie, создаются сервером при подключении. Эти данные помечены идентификатором, который являются уникальным для пользователя и его устройства.

Файлы cookie созданы специально для веб-браузеров, чтобы отслеживать, персонализировать и сохранять информацию о сеансе каждого пользователя. «Сеанс» означает просто время, которое посетитель проводит на сайте. Они создаются для идентификации при посещении нового веб-сайта. Веб-сервер, на котором хранятся данные веб-сайта, отправляет короткий поток идентифицирующей информации веб-браузер пользователя.

Файлы cookie браузера идентифицируются и считываются парами «имя-значение». Они сообщают cookie-файлам, куда следует отправлять и какие данные следует вспомнить.

Сервер отправляет файл cookie только тогда, когда он хочет, чтобы веб-браузер сохранил его. Если пользователь вернется на этот сайт в будущем, веб-браузер вернет эти данные на веб-сервер в виде файла cookie. Это когда ваш браузер отправит его обратно на сервер, чтобы восстановить данные из ваших предыдущих сеансов.

Веб-сайты используют файлы cookie для оптимизации веб-опыта. Без файлов cookie пользователю придется снова войти в систему после того, как он покинет сайт или перестроит корзину покупок. Сделать файлы cookie важной частью работы в Интернете. Основное предназначение файлов cookie заключается в следующем:

- Управление сеансом. Например, файлы cookie позволяют веб-сайтам распознавать пользователей и вспоминать их индивидуальные данные и предпочтения.
- Персонализация. Индивидуальная реклама - это основной способ использования файлов cookie для персонализации ваших сеансов. Файлы cookie используют эти данные для создания целевой рекламы, которая может заинтересовать потенциального покупателя.
- Отслеживание. Сайты покупок используют файлы cookie для отслеживания ранее просмотренных пользователями товаров, позволяя сайтам предлагать другие товары, которые им могут понравиться, и хранить их в тележках для покупок, пока они продолжают делать покупки.

Файлы cookie хранятся на вашем устройстве локально, что позволяет быстро загружать страницу при повторном заходе. В свою очередь, веб-сайты можно персонализировать, сэкономив при этом на обслуживании серверов и затратах на хранение.

Файлы cookie бывают двух типов: сеансовые и постоянные.

Сессионные файлы cookie используются только при навигации по веб-сайту. Они хранятся в оперативной памяти и никогда не записываются на жесткий диск.

По окончании сеанса файлы cookie сеанса автоматически удаляются. Они также помогают работать кнопке «назад» или сторонним плагинам анонимайзера. Эти плагины предназначены для работы в определенных браузерах и помогают поддерживать конфиденциальность пользователей.

Постоянные файлы cookie остаются на компьютере на неопределенный срок, хотя многие из них содержат дату истечения срока действия и автоматически удаляются по достижении этой даты.

IP-адрес - это уникальный адрес, который идентифицирует устройство в Интернете или локальной сети. IP означает «Интернет-протокол», который представляет собой набор правил, регулирующих формат данных, отправляемых через Интернет или локальную сеть.

По сути, IP-адреса - это идентификатор, который позволяет передавать информацию между устройствами в сети: они содержат информацию о местоположении и делают устройства доступными для связи. Интернету нужен способ различать разные компьютеры, маршрутизаторы и веб-сайты. IP-адреса позволяют это делать и являются важной частью работы Интернета.

IP-адрес - это строка чисел, разделенных точками. IP-адреса выражаются в виде набора из четырех чисел - например, адрес может быть 192.158.1.38. Каждое число в наборе может находиться в диапазоне от 0 до 255. Таким образом, полный диапазон IP-адресации находится в диапазоне от 0.0.0.0 до 255.255.255.255.

IP-адреса не случайны. Они математически производятся и распределяются Internet Assigned Numbers Authority (Управлением по присвоению номеров в Интернете, IANA), подразделением Интернет-корпорации по присвоению имен и номеров (ICANN). ICANN - это некоммерческая организация, основанная в США в 1998 году с целью помочь поддерживать безопасность Интернета и сделать его доступным для всех. Каждый раз, когда кто-либо регистрирует домен в Интернете, он проходит

через регистратора доменных имен, который платит ICANN небольшую плату за регистрацию домена.

IP-адреса могут быть постоянными (статическими) или временными (динамическими). Разница между статическими и динамическими IP-адресами заключается в том, что первые никогда не меняются, а вторые могут и меняются.

Статические адреса в основном используются предприятиями, поскольку их веб-сайты и веб-приложения должны быть всегда надежно доступны. Но ваш домашний IP-адрес не обязательно должен оставаться прежним, поскольку он нужен только тогда, когда вы пользуетесь Интернетом.

Ваш интернет-провайдер обычно предоставляет вам динамический IP-адрес. Хотя ваш IP-адрес может меняться не часто, вы можете получать новый IP от вашего интернет-провайдера при каждой перезагрузке компьютера. То же самое и с локальными IP-адресами, которые домашний беспроводной маршрутизатор назначает вашему ноутбуку, планшету или смартфону. Эти устройства могут получать новый адрес при каждом перезапуске маршрутизатора.

Единственный реальный недостаток динамических адресов заключается в том, что данный компьютер невозможно надежно найти. Это затрудняет, скажем, запуск веб-сервера у себя дома, поскольку адрес может измениться, и никто не сможет вас найти. Многие интернет-провайдеры позволяют организовать бизнес-соединение со статическим адресом, если вы хотите запустить сервер.

Интернет-протокол работает так же, как и любой другой язык, при общении с использованием установленных правил для передачи информации. Все устройства находят, отправляют и обмениваются информацией с другими подключенными устройствами, используя этот протокол. Использование IP-адресов обычно происходит негласно. Процесс работает так:

1. Устройство косвенно подключается к Интернету, затем подключается к сети, подключенной к Интернету, которая затем предоставляет устройству доступ к Интернету.
2. IP-адрес назначается устройству интернет-провайдером.
3. Интернет-активность проходит через интернет-провайдера, а они возвращают ее обратно, используя IP-адрес. Поскольку они предоставляют доступ в Интернет, их роль заключается в назначении IP-адреса вашему устройству.
4. Однако IP-адрес может измениться. Например, включение или выключение модема или маршрутизатора может изменить это. Пользователь можете связаться со своим интернет-провайдером, если ему необходимо сменить адрес.

Одним из самых важных частей идентификации пользователя данной системы, является создание логики конфигурации полученных данных, для более уникальной и надежной идентификации пользователя. Опытный пользователь сможет подделать, изменить эти данные, что приведет к регистрации нового пользователя в обход этой системы. Данные должны быть не только уникальными, но и надежными, чтобы пользователь не смог бы их подделать, но, к сожалению, нет таких абсолютно уникальных и надежных данных. Почти все данные, которые получает сайт от пользователя можно подделать, поэтому и создаются конфигурации этих данных.

К примеру, если использовать только IP адрес для идентификации пользователя, то для обхода нашей системы пользователю понадобиться только включить VPN-сервис, и он будет зарегистрирован как новый пользователь, а если использовать cookie, то при отчистке данных браузера, или при смене его на другой, тоже произойдет обход системы. Этим и обусловлено использование конфигурация полученных данных.

Сначала необходимо продумать логику для конфигурации системы. Какие-то данные сделать приоритетнее, какие-то сравнивать только в группе с другими данными. К примеру, если сравнивать IP и данные cookie, то это

уже усложнит задачу целенаправленного обхода нашей системы, потому что даже при использовании VPN-сервисов он будет идентифицирован по файлам cookie, а если очистит cookie, то по IP. Это всего лишь пример, если использовать такую конфигурацию данных, то не составит труда обойти его, просто включив VPN и очистив cookie. Любую систему можно обойти, мы лишь собираемся усложнить обход нашей системы, чтобы его идентифицировать не как нового пользователя, а как новое посещение раннее зашедшего пользователя.

Как говорилось уже ранее, в этой статье рассказывается лишь об основных идентификаторах, существуют еще масса таких же, что делает идентификацию пользователя более уникальной и надежной. По той же причине, в этой статье не был приведен конкретный пример конфигурации данных, так как логику конфигурации данных нужно составлять в конце, когда собраны все нужные идентификаторы пользователя.

На основе этой системы можно создать собственную базу данных пользователей сайта, анализировать их действия, улучшать сайт, создавать более подходящую ленту интересов пользователя.

Многие из получаемых данных законом не запрещены, даже нет каких-то условий для использования этих данных, но, к примеру, для хранения и использования файлов cookie, согласно законам Европейского Союза, организации, которые сохраняют файлы cookie на компьютеры посетителей веб-сайтов, обязаны заручиться их согласием, предоставив им понятную и полную информацию о том, как файлы cookie используются на этих веб-сайтах. Так будут проверяться все данные, которые берет наша система, и в соответствии с законом, будет уведомлять пользователя об их использовании.

Система автономной идентификации - это уникальная система, позволяющая с большой вероятностью идентифицировать пользователя. Надежность системы составляет 90-99%. На основе данных системы можно собрать статистику посещений, оценить востребованность ресурса у пользователей, определить основную аудиторию, определить мошенников на

сайте, блокировать конкретных нежелательных пользователей. Данная система будет полезна любому сайту, так как каждый проект стремится повысить свою популярность среди пользователей интернета, а для этого необходима система идентификации пользователя, чтобы проанализировать его действия и интересы, и на основе их создавать более подходящий контент под конкретные запросы и потребности потребителя.

Литература

1. Емелин А. В., Перов Б. Г. Правовые проблемы идентификации при дистанционном финансовом обслуживании физических лиц // Финансовый журнал. 2020. №5.
2. Иванов В. В., Лубова Е. С., Черкасов Д. Ю. Аутентификация и авторизация // Проблемы Науки. 2017. №2 (84).
3. Сидорова К. С. IP-адрес как один из идентификаторов личности в расследовании преступлений // Психопедагогика в правоохранительных органах. 2018. №3 (74).
4. Клочек М.С., Парфенова А. С. Изменение статического ip -адреса компьютера//Инновационное развитие. 2018. № 1 (18). С. 9-10.
5. Наралиев Нишонали Анорматович, Самаль Дмитрий Иванович Обзор и анализ стандартов и протоколов в области интернет вещей. Современные методы тестирования и проблемы информационной безопасности IoT // International Journal of Open Information Technologies. 2019. №8.
6. Козлов, С. Н. Защита информации : устройства несанкционированного съема информации и борьба с ними : Учебно-практическое пособие / Козлов С. Н. - Москва : Академический Проект, 2020. - 286 с. (Gaudeamus) - ISBN 978-5-8291-2956-9. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : https://www.studentlibrary.ru/book/ISBN9785829129569.html (дата обращения: 19.11.2021). - Режим доступа : по подписке.

Literature

1. Emelin A. V., Perov B. G. Legal identification problems in remote financial services for individuals // Financial journal. 2020. No. 5.

2. Ivanov V. V., Lubova E. S., Cherkasov D. Yu. Authentication and authorization // Problems of Science. 2017. No. 2 (84).

3. Sidorova KS IP-address as one of the personality identifiers in the investigation of crimes // Psychopedagogy in law enforcement agencies. 2018. No. 3 (74).

4. Klochek MS, Parfenova AS Changing the static ip-address of the computer // Innovative development. 2018. No. 1 (18). S. 9-10.

5. Naraliev Nishonali Anormatovich, Samal Dmitry Ivanovich Review and analysis of standards and protocols in the field of the Internet of Things. Modern testing methods and problems of information security of IoT // International Journal of Open Information Technologies. 2019. No. 8.

6. Kozlov, SN Information security: devices for unauthorized information retrieval and the fight against them: Study guide SN Kozlov - Moscow: Academic Project, 2020. - 286 p. (Gaudeamus) - ISBN 978-5-8291-2956-9. - Text: electronic // EBS "Student Consultant": [site]. - URL:https://www.studentlibrary.ru/book/ISBN9785829129569.html (date accessed: 11/19/2021). - Access mode: by subscription.