

Конспект статьи “Hiding in the Crowd: an Analysis of the Effectiveness of Browser Fingerprinting at Large Scale”.

Замечание: в данном конспекте под цифровым отпечатком устройств (Фингерпринт) или же просто отпечатком я понимал следующее английское словосочетание - Browser fingerprinting.

Прошлые исследования показали, что современные устройства настолько разнообразны, что цифровые отпечатки устройств можно использовать для идентификации и отслеживания пользователей в Интернете. В своей работе авторы хотят оценить, насколько технология отслеживания цифровых отпечатков устройств по-прежнему эффективна для уникальной идентификации большой группы пользователей при анализе миллионов цифровых отпечатков устройств за несколько месяцев.

Основной вклад авторов статьи заключается в следующем: 1) анализ 2 067 942 отпечатков, состоящих из 17 различных атрибутов 2) доказательство того, что Фингерпринт не так эффективна, как сообщалось в литературе. В то время как предыдущие исследования сообщали о наличии более 80% уникальных отпечатков, авторы получили 33,6% 3) сравнили свой набор данных с данными из “Panopticlick” и “AmIUnique” 4) рассмотрели развитие в будущем техники цифрового отпечатка устройства.

Предыдущие исследования в 2010 и 2014-2015 годах отличаются от исследования авторов статьи тем, что исследования авторов статьи охватывают более широкий круг людей и данные собраны с сайтов о погоде и политике, которые не были посвящены сайту о цифровых отпечатках устройств. Таким образом собранные данные были более приближены к реальности.

Уникальность отпечатка определяется с помощью сравнения определённого набора атрибутов, представленных в работе. Особое внимание уделено следующему атрибуту - шрифтам. Один из способов получения списка шрифтов — это Flash-плагин. Однако большинство современных веб-браузеров блокируют его, поэтому есть другой способ определения списка шрифтов - с помощью “JavaScript”. Тем не менее такой метод может дать неточный результат и в статье представлено его улучшение. Стоит отметить, что в исследовании авторы различают два типа отпечатков - тип, принадлежащий мобильным устройствам и тип, принадлежащий настольным и портативным компьютерам, так как уникальные отпечатки различных типов устройств собираются по-разному. Разделяя отпечатки на два типа, авторы приводят распределение операционных систем среди пользователей, которое по мнению авторов более приближено к реальности по сравнению с прошлыми исследованиями, а также приводят наборы данных для мобильных устройств и для портативных компьютеров, а именно приводится количество уникальных и различных значений атрибутов, полученных в ходе исследований.

В ходе исследования было выявлено 46 459 цифровых отпечатков на мобильных устройствах и 647,741 на персональных компьютерах. Отпечаток является уникальным, если он имеет атрибут, значение которого присутствует только один раз во всем наборе данных или же комбинация всех его атрибутов уникальна во всем наборе данных. Однако в течение своего срока службы устройства может давать разные отпечатки. Это может быть связано с обновлением версии браузера, активацией специальных плагинов, перемещением в другой часовой пояс и т.д. В какой-то момент времени может случиться и так, что неуникальные отпечатки станут уникальными. Например, при смене языка в браузере. Возникает вопрос, а когда ещё это может случиться? Авторы приходят к тому, что если один-единственный признак цифрового отпечатка изменился, то очень вероятно, что отпечаток станет уникальным. Кроме

того, авторы приводят математическую обработку сравнения атрибутов с помощью введённой энтропии, что позволяет более точно ответить на предыдущие вопросы.

Данные, собранные с сайтов "Panopticlick" и "Amlunique", дают высокие значения уникальности. Однако данные авторов статьи дают результаты заметно ниже, что может говорить о неэффективности Фингерпринт при большой и различной аудитории. Это отчасти связано с тем, что процент уникальности, по мнению авторов, тем выше, чем меньшее число отпечатков собрано.

Существуют некоторые проблемы, связанные с эволюцией веб-технологий, которые возникают из-за снижения идентифицирующей ценности определенных функций. Помимо этого, существуют некоторые трудности, появляющиеся в процессе сбора данных, причиной которых может быть к примеру - неравномерное распределение доли рынка компьютеров и мобильных на планете. Кроме того результаты, полученные в статье, показывают, что эволюция веб-технологий может принести пользу конфиденциальности пользователей. А также, что приватность отпечатков кажется достижимой благодаря низкому проценту уникальных отпечатков, которые представлены в этом исследовании.

В итоге результаты статьи показывают, что современные методы Фингерпринт не обеспечивают эффективных механизмов уникальной идентификации пользователей, принадлежащих к определенному демографическому региону. Вторым важным моментом заключается в том, что персональные компьютеры и мобильные устройства имеют уникальные отпечатки пальцев, которые составляют по-разному.