

Корниенко С.В. Big Data: проблемы безопасности пользовательских данных // Академия педагогических идей «Новация». Серия: Студенческий научный вестник. – 2019. – №2 (февраль). – АРТ 145-эл. – 0,3 п.л. - URL: <http://akademnova.ru/page/875550>

РУБРИКА: ПЕДАГОГИКА И ПСИХОЛОГИЯ

УДК 004.006.5

Корниенко Сергей Вячеславович
студент 4 курса, факультет информационных систем
в экономике и юриспруденции
Научный руководитель: Паскова А.А., к.п.н., доцент
ФГБОУ ВО «Майкопский государственный технологический университет»
г. Майкоп, Российская Федерация
e-mail: passann@mail.ru

**BIG DATA: ПРОБЛЕМЫ БЕЗОПАСНОСТИ
ПОЛЬЗОВАТЕЛЬСКИХ ДАННЫХ**

Аннотация: Статья посвящена технологиям Big Data, и, в частности, проблемам безопасности пользовательских данных. В отличие от законодательно регулируемого сбора и обработки персональных данных, сбор и обработка пользовательских данных никак не защищены. Рассмотрены возможные способы несанкционированного сбора пользовательских данных, приведены рекомендации по снижению рисков, связанных с их обработкой.

Ключевые слова: Big Data, персональные данные, пользовательские данные, политика конфиденциальности, безопасность пользовательских данных.

Kornienko Sergey
4th year student, faculty of information systems
in economics and jurisprudence
Supervisor: A. Paskova, PhD, Associate Professor
FGBOU VO "Maikop State Technological University "
Maikop, Russian Federation

BIG DATA SECURITY PROBLEM OF USER DATA

Abstract: the article is devoted to Big Data technologies, and, in particular, to the problems of user data security. Unlike the legally regulated collection and processing of personal data, the collection and processing of user data is not protected in any way. The possible methods of unauthorized collection of user data are considered, recommendations for reducing the risks associated with their processing are given.

Keywords: Big Data, personal data, user data, privacy policy, user data security.

Сейчас в сети Интернет настолько большое количество информации, что последняя зачастую перестает иметь хоть какую-то структуру. А общий объем данных давно перевалил за 10 зеттабайт [1]. Подобного рода большие объёмы данных необходимо быстро обрабатывать и структурировать для возможного дальнейшего использования. Поэтому, сравнительно недавно, появился не в полную меру определенный термин – «Big Data».

Большие данные – обозначение структурированных и неструктурированных данных огромных объёмов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами, появившимися в конце

2000-х годов и альтернативных традиционным системам управления базами данных и решениям [2].

Технологии Big Data функционируют, руководствуясь тремя основными принципами (3V):

1. «Volume» – количественная характеристика, отражающая размер информации.

Количество – один из основных показателей в технологиях Big Data. Информация, имеющая огромные размеры, не может быть обработана ручным способом. Это говорит о том, что технологии Big Data свойственны автоматизированным системам обработки информации (АСОИ).

2. «Velocity» – скоростной показатель обрабатывания данных.

Информацию необходимо подвергать обработке максимально быстро, так как их значения могут меняться каждую единицу времени, а массив собираемой информации может расти как вертикально, так и в горизонтальном направлении.

3. «Variety» – показатель многообразия собираемой информации.

Данный показатель в основном раскрывает степень упорядоченности информации, то есть структурированная и неструктурированная.

Также технологии Big Data могут функционировать согласно дополнительных правил (2V): «veracity» – показатель истинности собираемых и «value» – значимости этих данных [3].

Big Data представляет один из инструментов Customer Relationship Management (CRM). Он основывается на потребностях каждого конкретного пользователя электронных ресурсов [4].

Технологии Big Data в настоящее время применяется во многих сферах человеческой деятельности: маркетинг, промышленность, здравоохранение, образование и т.д.

Но использование технологий Big Data привело к возникновению проблемы безопасности пользовательских данных. Big Data – это вся информация, попадающая в электронную среду. Big Data формируется из логов, начиная с пользователей и заканчивая датчиками каких-либо систем. Первоначально она имеет неструктурированный вид; она никак не индексируется, а более 80% её не представляют никакого интереса, точнее сказать, представляют не для всех. Такая информация образуется в зависимости от источника добычи информации. Помимо уже озвученных, это могут быть копии целых сайтов, причем сохранение происходит при любом изменении сайта; все имена, логины, пароли – всё, что вводил человек при авторизации (KeyLogger); координаты месторасположения; отпечаток пальца и многое другое. Всё, что попадает в сетевую среду – остается там. Вашу информацию могут удалить из индексируемых источников по обоснованной причине, тем самым скрыть от ClearNet, но есть и DeepWeb, DarkNet.

Методики добычи и анализа больших данных постоянно совершенствуются, а также появляются новые. Принято выделять 25 методов, используемых в Big Data, из которых можно отметить: A/B testing, Classification, Data mining, Machine learning, Network analysis и другие [6].

Любой из уже существующих методов Big Data в узких кругах модифицирован для конкретных задач предметной области и имеет обширный функционал, а вычислительные мощности в разы превосходят стандартно используемые, что вполне характерно для пиринговой сети (P2P/I2P). Но только инструментальными способами все не ограничивается. Есть и заинтересованные в максимальной точности искомого люди – NetStalkers. Они «копаются» там, где автоматизированный метод может «застрять», пропустить и т.д. какую-либо информацию. Именно от них

исходят многие инсайды (инсайдерская информация) по совершенно разным областям жизни на основе дедукции и логики; они используют найденную информацию так, как им будет удобно, в своих интересах. Совершенно очевидно, что речь о порядочности здесь не идет. Может быть, поэтому термин Big Data и носит такое «мягкое» определение.

На данный момент о конфиденциальности пользовательских данных говорить не приходится. Помимо основного сервиса услуг при взаимодействии (в особенности при регистрации учетной записи) с Интернет-ресурсом, Вы, соглашаясь с лицензионным использованием, разрешаете передачу третьим лицам свои персональные данные (ПД) с пометкой «доверенные соуслуги». И после этого Вам звонят непонятные менеджеры на телефон, на электронный ящик приходят письма и т.д. А все потому, что третья организация посчитала более выгодным продать (в лучшем случае) Ваши ПД, а не заниматься конфиденциальной их обработкой.

В такой ситуации большинство пользователей сети Интернет просто помещают отправителя в черный список «Спам». Но при условии «динамических электронных почт», Ваш черный ящик будет очень быстро заполняться, а Ваше терпение заканчиваться.

Совсем недавно усложнение структуры одной из крупнейших социальных сетей привело к многочисленным сбоям этого ресурса. В результате появлялась возможность получения прав администратора сторонних групп и многое другое. Скрытые возможности данного ресурса, которые администрация называет «фишками», помогают модифицировать контент на данном ресурсе. Определенные спецсимволы могут вызывать перебои в работе определенных веб-страниц и даже блокировку личной страницы. Имеется возможность без согласия пользователя возвращать его

в покинутые им группы. Вы можете поставить «лайк» под записью с «котятами», а администратор через какое-то время изменит содержимое записи (в лучшем случае на « меховые изделия»). После этого алгоритмы данного ресурса (и не только) будут предлагать Вам не «котиков», а рекламу «шуб». Также этот ресурс позволяет искать и просматривать файлы и документы, которые пользователь не пометил, как конфиденциальные.

Аналогично, недавно реконструкция алгоритмов поиска известного российского поискового гиганта привела к индексации огромного количества данных, которые считались конфиденциальными для другого поисковика.

Алгоритмы Big Data некоторых Интернет-ресурсов бесконтрольно (или целенаправленно) осуществляют сбор и последующую выдачу той информации, которая никак не должна попадать в открытый доступ согласно политике конфиденциальности.

Алгоритмы методов сбора данных пользователя способны понимать, кто пользуется электронным средством, тем самым персонализируя рекламные, новостные компании и так далее. Отслеживание по веб-камере (взгляда, эмоций и т.д.), биометрическим характеристикам (особенности использования пользователем электронных средств, сканирование отпечатков пальцев, сетчатки глаза и др.), геологам, запросам в сети Ethernet и т.д.

Почему так популярны сервисы по компрометации пользователей с использованием телефонных номеров? Всё просто – политика конфиденциальности не работает. Если Интернет-ресурс обещает, что пользователя не определят по номеру телефона, то никто не должен узнать его ФИО, email, место жительства и так далее. А все получается наоборот.

То есть, функционирование Big Data позволяет собирать не только персональные данные, но и любые другие пользовательские данные – большие пользовательские данные (БПД). В данном случае персональные данные являются лишь частью всей собираемой информации, то есть БПД. В свою очередь БПД могут включать абсолютно всю информацию о пользователе Интернет-ресурса, консолидируясь при этом с другими ресурсами, и тем самым предоставляя им о Вас данные [5].

Таким образом, практически невозможно говорить о конфиденциальности применительно к технологиям Big Data.

Для того, чтобы обезопасить себя от бесконтрольного сбора информации в сети Интернет, можно применять следующие средства, инструменты и методы: Virtual Private Network (VPN), Pretty Good Privacy (PGP), Proxy, Socks, браузеры с расширенным управлением конфигураций (Tor, Firefox, Puffin), виртуальные машины, отключение JavaScript, блокировщики рекламы (включая настройку блокировки на уровне роутера), очистка Cookie, Cash и рекламного идентификатора и метаданных, мониторинг файлов, процессов и реестра операционной системы (Task Killer, Spy Hunter, Hitman Pro, Ccleaner). Все это можно использовать, чтобы не отсылать в корпорации свои данные.

Конечно, использование сетевых протоколов, предоставляющих анонимность также накладывает вероятность компрометации. А в свете последних событий, когда 80% «теневых» ресурсов подверглись достаточно серьезной уязвимости, которая давала права администратора для доступа к базам данных, о безопасности не стоит и упоминать.

Самым действенным методом является осторожность в Интернете с самого начала использования любого ресурса. Два разных устройства, одно – для выхода в Интернет, другое – для работы, обеспечат дополнительный

рубеж в защите конфиденциальных данных и БПД (в том числе персональных данных).

Безусловно, технологии Big Data очень полезный инструмент, они открывают широкие возможности как для разработчиков, так и для пользователей. Но, к сожалению, на сегодняшний день любой электронный ресурс/услуга в сети Интернет (речь о ClearNet) не отвечает требованиям конфиденциальности для комфортного использования обычными пользователями (не имеющими достаточного опыта использования электронно-вычислительных средств).

Список использованной литературы:

1. А. Беркана Что такое Big data: собрали всё самое важное о больших данных [Электронный ресурс] // Rusbase – URL: <https://rb.ru/howto/chto-takoe-big-data/> (дата обращения: 21.10.2018)
2. Большие данные (Big Data) [Электронный ресурс] // Tadviser – URL: [http://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5_\(Big_Data\)](http://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5_(Big_Data)) (дата обращения: 21.10.2019)
3. Большие данные [Электронный ресурс] // Wikipedia – URL: https://ru.wikipedia.org/wiki/Большие_данные (дата обращения: 21.10.2018)
4. Е Чемякин BIG DATA. то, о чем все слышали, но мало кто знает что это [Электронный ресурс] // IN-SCALE – URL: <https://in-scale.ru/blog/big-data.html> (дата обращения: 15.11.2018)
5. Паскова А.А. Технологии Big Data в автоматизации технологических и бизнес-процессов [Электронный ресурс] // Научное обозрение. Технические науки. – 2018. – № 4. – С. 23-27.– URL: <https://science-engineering.ru/ru/article/view?id=1193> (дата обращения: 10.12.2018)
6. Я. Сосов Большие пользовательские данные [Электронный ресурс] // Лекториум – URL: http://project.lektorium.tv/big_personal_data (дата обращения: 12.01.2019)

Дата поступления в редакцию: 31.01.2019 г.

Опубликовано: 07.02.2019 г.

© Академия педагогических идей «Новация». Серия «Студенческий научный вестник», электронный журнал, 2019

© Корниенко С.В., 2019