

УДК 004.67

ИСТОЧНИКИ БОЛЬШИХ ДАННЫХ И СОВРЕМЕННЫЕ СПОСОБЫ ХРАНЕНИЯ ДАННЫХ



А. Г. Ясюкевич
Студент БГУИР,
инженер-
программист



Т. В. Славинский
Студент БГУИР,
инженер-
программист



С. Н. Нестеренков
Кандидат
технических наук,
доцент, декан
факультета
компьютерных
систем и сетей



А. Н. Марков
Старший
преподаватель,
магистр технических
наук, заместитель
начальника Центра
информатизации и
инновационных
разработок БГУИР

Центр информатизации и инновационных разработок Белорусского государственного университета информатики и радиоэлектроники, Республика Беларусь
Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
E-mail: 85100186@study.bsuir.by, s.nesterenkov@bsuir.by, Ilyashilov@mail.by

А. Г. Ясюкевич

Студент 4-ого курса БГУИР ФКСИС, по специальности ПОИТ

Т. В. Славинский

Студент 4-ого курса БГУИР ФКСИС, по специальности ПОИТ

С. Н. Нестеренков

Кандидат технических наук, декан факультета компьютерных систем и сетей Белорусского государственного университета информатики и радиоэлектроники, доцент кафедры Программного обеспечения информационных технологий. Автор публикаций на тему машинного обучения, алгоритмов принятия решений, искусственных нейронных сетей и автоматизации

А. Н. Марков

Магистр технических наук, старший преподаватель кафедры ПИКС, заместитель начальника Центра информатизации и инновационных разработок Белорусского государственного университета информатики и радиоэлектроники.

Аннотация. Большие Данные на сегодняшний день являются достаточно известным трендом в области информационных технологий. Многие говорят о работе с Большими Данными, но немногие приводят конкретные примеры работы с ними, которые полностью соответствовали бы определению самих Больших Данных. Несмотря на это, нельзя оставить без внимания, все больший рост и развитие технологий работы с данными.

Ключевые слова: Big Data, Большие Данные, мобильные технологии, Интернет вещей, способы хранения данных.

Введение.

Существует несколько важных направлений, которые активно развиваются в области информационных технологий и напрямую относятся к Большим Данным [1], но основным

нужно выделить с точки зрения увеличения объемов данных – мобильные технологии и Интернет вещей.

Мобильные технологии, Интернет вещей и Большие Данные.

Направления исследования в области носимых вычислительных систем не являются новыми, однако значительное продвижение данное направление получило с появлением смартфонов и развитием мобильных технологий. Если некоторое время назад ученые говорили о том, что технология Smart Dust для получения в режиме реального времени данных о том или ином объекте или территории будет внедряться десятки лет, то с появлением смартфонов и технологическим прогрессом в области мобильных технологий фактически данный подход уже внедряется. Каждый человек добровольно пользуется датчиком в виде мобильного устройства, которое снимает определенные данные и может их передавать для дальнейшей обработки.

Наличие мобильных устройств и развитие технологий передачи данных привело к тому, что на сегодняшний день много исследований ведется по теме Интернета вещей. В настоящее время количество устройств, подключенных к Интернету, постоянно растёт (Рис. 1), и уже в 2009 году их число превысило население планеты, то есть на каждого человека стало приходиться почти 2 подключенных устройства. Эти устройства являются не только смартфонами и ноутбуками, но и многими другими не столь традиционными предметами, подключенными к Интернету, – чайниками, холодильниками, автомобилями и другими. Если обратиться к истории, то еще в 1926 году Никола Тесла в интервью для журнала «Collier's» сказал, что в будущем радио будет преобразовано в «большой мозг», все вещи станут частью единого целого, а инструменты, благодаря которым это станет возможным, будут легко помещаться в кармане. А в 1990 году выпускник МИТ Джон Ромки создал первую в мире интернет-вещь. Это был его тостер, подключенный к сети.

Рост количества IoT-устройств, подключенных к общей сети (млрд)

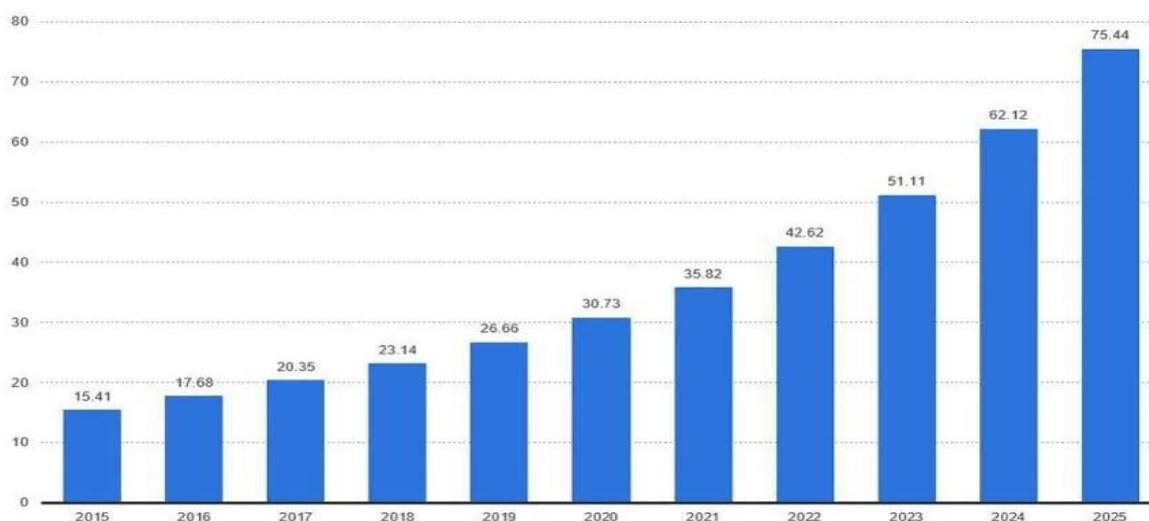


Рисунок 1 – Рост количества IoT-устройств, подключенных к общей сети (млрд)

Нужно отметить, что Интернет вещей (то есть устройства, которые подключены к сети) является источником одного типа данных – более ориентированных на устройства. Однако есть еще и мобильные технологии, в рамках которых пользователи в повседневной деятельности используют мобильные приложения, а мобильные приложения являются источниками данных, более ориентированных на пользователя.

В настоящее время Интернет вещей состоит из разрозненных и практически не связанных между собой сетей, каждая из которых создана для решения своих конкретных задач и собирает, накапливает и анализирует данные с определенных устройств. Для объединения в единый работающий механизм необходима стандартизация методов работы этих устройств и передачи информации. Внедряя Интернет вещей в повседневную жизнь, человечество сталкивается с его трехчастным построением (согласно Робу ван Краненбургу): Интернет вещей – это непрерывный поток данных, который начинается от нашего тела BAN (Body Area Network) [2], домашней и рабочей обстановки LAN (Local Area Network), городской инфраструктуры WAN (Wide Area Network) и растворяется в глобальной информационной системе VWAN (Very Wide Area Network). Монетизация такой глобальной системы происходит в результате того, что конечные пользователи оплачивают предоставление доступа к данным, собранным в результате работы этого непрерывного потока, либо к услугам, которые предлагаются на основе полученных данных и их проведенного анализа.

Благодаря современному стандарту взаимодействия 6LoWPAN [3], позволяющему подключаться к Интернету маломощным устройствам, в настоящее время установить микрокомпьютер в любой предмет, начиная от браслета или зубной щетки, не представляет особой сложности. Но на текущий момент целесообразно говорить о множестве различных несвязанных сетей, которые решают отдельные задачи и соединяют отдельные устройства. Один из немногих путей решения проблемы разрозненности – это серьезное экономическое стимулирование взаимодействия производителей мобильных устройств либо провайдеров услуг, которые могут оказывать новые услуги в рамках объединенной сети.

Независимо от того, что до сих пор человечество имеет раздробленные сети, на их основе уже выстроены различные бизнес-модели, по которым работают компании. Они занимаются мониторингом потребления ресурсов, экологической обстановки, слежением за здоровьем человека и животных, отслеживанием движения каждого конкретного потребительского товара для оптимизации поставок и др. Интернет вещей может быть использован в том числе и страховыми компаниями для отслеживания поведения (перемещение, скорость автомобиля) клиентов и предоставления персонализированных тарифов с учетом их наклонностей [4]. И если говорить о перспективах, то одним из основных трендов во многих сферах выделена повсеместная персонафикация, что позволит (совместно с развитием и распространением 3D-печати) изготавливать необходимые устройства прямо у себя дома и непосредственно под свои индивидуальные особенности. Данные об этих особенностях будут собираться с многочисленных датчиков, в том числе и мобильных телефонов. Получается, что сама информация, собираемая каким-либо продуктом по мере использования, становится активом наравне с физическими активами или трудом, пользование которым необходимо оплачивать. Это приводит к разрушению привычных бизнес-моделей, основанных на продаже продукта, и появлению новых, в которых монетизация идет через предоставление дополнительных услуг для приобретенного продукта и непосредственно для потребителя.

Также необходимо рассмотреть приоритетные проекты некоторых корпораций – лидеров в данной области: Cisco, HP и IBM. Проект Cisco в партнёрстве с NASA – Planetary Skin [5] предполагает объединение спутниковой сети, беспилотных самолетов, а также наземных средств и датчиков для сбора информации о Земле, ее процессах и явлениях последующего контроля земных ресурсов и предоставления этой информации людям для повышения уровня жизни. Это и подтверждается миссией, приведенной на официальном сайте проекта. Среди основных направлений Planetary Skin – прогнозирование и оптимизация потребления энергии, прогнозирование стихийных бедствий, исследование и принятие решений в сфере управления водными ресурсами, поддержка сельского хозяйства и анализ рисков с использованием сенсорных сетей, программа мониторинга состояния лесов, а также мониторинг взаимосвязей потребления различных ресурсов.

Central Nervous System for Earth от HP заключается в повсеместном внедрении сенсоров, считывающих такие показатели, как давление, температура, освещенность, вибрация, влажность и некоторые другие. Также будут использоваться другие датчики, похожие на популярные RFID-метки, однако являющиеся еще и мощными акселерометрами. Сферы применения так же обширны, как и у Planetary Skin от Cisco. Датчики могут устанавливаться на мосты и строения, вдоль дорог для мониторинга загрузки. В дальнейшем возможны вхождение датчиков в бытовую электронику и в конечном счете, переход к «Интернету вещей» от изначальной сети, отвечающей за мониторинг состояния природы и инфраструктурных сооружений.

Большой интерес представляет проект IBM – Smarter Planet [6]. В спецификации отмечается, что многие компании собирают данных куда больше, чем могут позволить себе обработать. Однако на «Умной планете» наиболее крупные организации смогут обработать и превратить эти данные в ценную информацию о клиентах, бизнесе и окружающем мире в целом, и таким образом откроются новые возможности для оптимизации принимаемых решений. Это также всеобъемлющая сеть датчиков, осуществляющих мониторинг важнейших показателей окружающей среды. Наибольший интерес в данном случае представляет документ, затрагивающий сферу электронных устройств на «Умной планете». В нем говорится об открывающихся возможностях мониторинга жизненного цикла продукта, начиная от производства, и заканчивая тем, как его использует владелец. На основе полученных данных можно осуществлять поиск идей для новых услуг и сервисов. Также говорится о смене ориентации деятельности некоторых компаний, когда они получают основную прибыль, продавая не продукт, а услугу, с ним связанную. Одним из интересных направлений IBM является процесс превращения обычных городов в «умные». Это будет происходить через создание товаров и услуг для городских управлений. Такое направление получило название IBM Smarter Cities. Большие данные и Интернет вещей, являющийся одним из источников Больших Данных, также имеют одну общую проблему — достоверность получаемых данных и защиту от несанкционированного доступа к данным. Таким образом, Большие Данные – это не только правило четырех V: Volume, Variety, Velocity и Value (объем, вариативность, скорость и ценность), но еще и Verification (подтвержденность, достоверность). В материале достаточно подробно указаны угрозы нарушения прав человека, которые могут возникать в процессе обработки данных, более ориентированных на пользователя. Также приведены мнения специалистов из различных областей – бизнеса, науки, политики. Отмечается, что необходимы новые методы обработки Больших Данных, которые позволили бы исключить возможность идентификации пользователя по имеющемуся набору, а также необходимо межгосударственное взаимодействие с точки зрения формирования базового законодательства из-за появления больших объемов данных о гражданах разных стран, которые хранятся в компаниях [7]. Данный вопрос также затрагивает понятие транснациональных компаний и впоследствии, как это было уже упомянуто, платежи компаний пользователям за использование их данных в деятельности компаний в рамках формирования их информационных активов.

Заключение.

В материале основное внимание уделено Интернету вещей из-за стремительного технологического развития, поскольку на сегодняшний день речь идет уже о создании и внедрении сетей на наноуровне. Это означает, что потоки генерируемых данных, передаваемых от устройств наноуровня, будут в десятки и сотни раз больше, чем потоки данных, которые генерируются в рамках сетей Интернета вещей. В других научных материалах отмечается, что устройства подобных размеров нуждаются в больших объемах встроенной памяти, для хранения и обработки базовых данных (на текущий момент известно, что на 10 мкм можно получить лишь 7 Кбайт памяти для хранения данных). В

таким случае пока речь идет лишь о базовом взаимодействии устройств на основе принципов молекулярного взаимодействия. Кроме разработки новых протоколов обмена информацией, ключевыми вопросами являются вопросы разработки новых методов обработки данных в режиме реального времени для минимизации требуемых объемов памяти совместно с новыми подходами организации хранения данных на наноуровне, что само по себе уже является революционным в условиях развития концепции Больших Данных.

Список литературы

- [1] Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. — СПб.: Питер, 2019. — 208 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-4461-1040-7.
- [2] Нестеренков, С.Н. Применение больших данных в электронном образовании / С.Н. Нестеренков, М.И. Макаров, Н.В. Ющенко, А.Д. Радкевич // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. материалов V Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13-14 марта 2019 года). В 2 ч. Ч. 2 / редкол. : В. А. Богущ [и др.]. - Минск : БГУИР, 2019. - С. 242-245.
- [3] Интернет-блог «Русские блоги». Протокол 6LoWPAN [Электронный ресурс] // RUSSIANBLOGS.COM : сервер радиолобителей России.[Б. м.], 2018. URL: <https://russianblogs.com/article/5844621883> (дата обращения: 28.03.2022).
- [4] Нестеренков, С.Н. Система определения вероятности успешного прохождения собеседования соискателем на основе экспертных оценок / С.Н. Нестеренков, В.Н. Видничук, Н.Н. Шинкевич // Информационные технологии и системы 2018 (ИТС 2018) : материалы междунар. науч. конф., Минск, 25 окт. 2018 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. - Минск, 2018. - С. 126-127.
- [5] Интернет-блог компании «TAdviser». Что такое интернет вещей: История возникновения и развития интернета вещей [Электронный ресурс] // TADVISER.RU : Про государство, бизнес, технологии.[Б. м.], 2017. URL: [https://www.tadviser.ru/index.php/Статья:Что_такое_интернет_вещей_\(Internet_of_Things,_IoT\)](https://www.tadviser.ru/index.php/Статья:Что_такое_интернет_вещей_(Internet_of_Things,_IoT)) (дата обращения: 28.03.2022).
- [6] Журнал «Московский экономический журнал». Перспективы внедрение концепции умных городов как основа устойчивого развития [Электронный ресурс] // CYBERLENINKA.RU : Научная электронная библиотека «КиберЛенинка».[Б. м.], 2021. URL: <https://cyberleninka.ru/article/n/perspektivy-vnedrenie-kontseptsii-umnyh-gorodov-kak-osnova-ustoychivogo-razvitiya> (дата обращения: 28.03.2022).
- [7] Нестеренков, С.Н. Использование экспертных оценок навыков для предсказания успешного прохождения соискателем собеседования на работу и формирования рекомендаций по изучению дополнительного материала/ С.Н. Нестеренков, В.Н. Видничук, Н.Н. Шинкевич // Высшее техническое образование: проблемы и пути развития : материалы IX Междунар. науч.-метод. конф., Минск, 1-2 ноября 2018 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: В.А. Богущ [и др.]. - Минск, 2018. - С. 326-328.

SOURCES OF BIG DATA AND MODERN DATA STORAGE

A. G. YASIUKEVICH
*Student of
BSUIR, software
engineer*

T. V. SLAVINSKY
*Student of
BSUIR,
software
engineer*

S. N. NESTERENKOV
*PhD, Associate Professor
Dean of the Faculty of
Computer Systems and
Networks*

A. N. MARKOV
*Senior lecturer of the
department, Deputy
head of the Center for
Informatization and
Innovative
Developments*

Center for Informatization and Development of the Belarusian University of State Informatics and Radioelectronics, Republic of Belarus.

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus.

E-mail: 85100186@study.bsuir.by, s.nesterenkov@bsuir.by, Ilyashilov@mail.by

Abstract. Big Data is a fairly well-known trend in the field of information technology today. Many talk about working with Big Data, but few give specific examples of working with them that would fully meet the definition of Big Data itself. Despite this, we cannot ignore the growing growth and development of data technologies.

Keywords: Big Data, mobile technologies, Internet of things, data storage methods.