Вопросы информационной безопасности «больших данных»

Гюляра Мурадова

Азербайджанский Технический Университет

gularamu.aztu.edu.az

***Аннотация*— В статье анализируются проблемы защиты личной информации, сохранения приватности, утечки конфиденциальных данных, а также связанная с этим недооценка рисков. Приведены методы обработки больших потоков данных с применением «классических» систем безопасности. Рассматривается применение существующих технологий «больших данных» в целях защиты информации.**

Ключевые слова— технологии «больших данных»; защита данных; конфиденциальность; риск несанкционированного доступа

I. ВВЕДЕНИЕ

Роль науки в развитии цивилизации, соотношение науки и техники, представление о связанных с ними современных социальных и этических проблемах, ценность научной рациональности и ее исторических типов — в части влияния феномена «больших данных» на развитие общества, а также важности применения научных методов для извлечения общественной пользы и больших массивов данных является необходимым в понимании, изучении технологий Big Data. Большие объемы данных требуют надежного хранения и кропотливого анализа, который позволяет получить ответы на вопросы: где лежит информация конкретного типа? кто имеет к ней доступ? как оперативно найти ту или иную информацию определенного типа? как оптимизировать информационные потоки? как оптимизировать бизнес-процессы, использующие информацию того или иного типа? [1]

Создание глобального киберпространства, появление новых информационных технологий порождают множество острейших проблем, с регулированием которых не справляются юридические законы. Глобальное информационное пространство создает пользователю практически неограниченные возможности перемещения по информационным сетям любой информации, даже не ставя в известность об этом ее собственника. Поэтому защита интеллектуальной собственности в таком пространстве является трудной задачей в техническом и юридическом планах, а также требует больших экономических затрат. Необходимость поиска специфических принципов, норм и методов оценки различных действий, связанных с использованием компьютерной техники, приобретает все большую актуальность. Способы добычи информации на протяжении веков делились на два класса: законные и незаконные. Происходит размывание границ недозволенного и вседозволенного [2]. Чтобы большие наборы данных стали революционным преимуществом, их нужно анализировать так же быстро, как они поступают в хранилище. За считанные микросекунды необходимо решить, стоит ли сохранять определенный бит данных и будет ли он иметь смысл в сочетании с другими данными. Для высокоскоростного анализа больших объемов информации нужно слаженное взаимодействие между самими потоками данных и вычислительной инфраструктурой. Например, лицо известного преступника, опознанное среди тысяч других изображений, может инициировать сигнал «стоп», выявленная схема кредитного мошенничества - рассылку предупреждений, а растущий отток клиентов - предложение купонов. Аналитическая система должна выявлять любую существенную аномалию в данных до того, как данные будут сохранены, чтобы можно было оперативно принять меры [3].

II. НЕЗАЩИЩЕННОСТЬ ДОСТУПА К «БОЛЬШИМ ДАННЫМ»

Ежедневно по сетям передаются петабайты важнейшей информации предприятий, государственных учреждений и рядовых пользователей. Это создает быстро растущий риск несанкционированного доступа и нарушений безопасности. При объединении социальных сетей с облаками неизбежно надо рассмотреть проблему безопасности. Для регулирования доступа к облачным ресурсам нужны будут пользовательские группы, роли и политики. Процесс регулирования может опираться на социальные взаимосвязи: используя их, можно создать эволюционирующую систему контроля доступа, которая будет автоматически адаптироваться при добавлении, удалении пользователей или изменении их взаимосвязей. Поскольку облачные ресурсы, как правило, динамичны, необходимы правила назначения прав доступа, автоматически адаптирующиеся по мере появления новых ресурсов и подключения новых пользователей к социальной сети. При этом с зарождением новых данных в облаке правила доступа к ним могут генерироваться с использованием схем мгновенной классификации. Автором [4] анализируется подобная сеть в виде социального графа с политиками безопасности и пользовательскими группами − новые пользователи по мере присоединения к сети автоматически

распределяются по группам в зависимости от своих социальных связей .

Развитие технологии «больших данных» в медицине неизбежно столкнется и с внешним препятствием, лежащим в юридическом поле: у любой медицинской информации есть владельцы, которые хотят получить право самостоятельно управлять своими медицинскими записями – это становится все более важно в глобальном мире, где пациенты часто переезжают с места на место, меняют работу и поставщиков медицинских услуг. Но и в этом случае врачам требуется доступ к данным о пациентах в любой точке и в любое время. Данные не всегда можно сделать анонимными. В геномике деидентифицировать данные практически невозможно. Генетическая информация однозначно указывает на конкретного пациента, его очень легко реидентифицировать, и вопрос с приватностью данных здесь должен быть решен на законодательном уровне [5]. Позволят ли люди обрабатывать компаниям их персональные данные и вообще всю поступающую от клиентов информацию? Информация предоставляется бесплатно, потом анализируется, обрабатывается компьютером, производится новый продукт, который затем снова продается людям. Вопрос в том, что корректно ли использовать таким образом «собственность» клиента. [6,7]. Больные не догадываются, что данные о них и их действиях могут быть повторно использованы для других целей. Коммуникация избегает социального контроля и возникает угроза негативного влияния на человека.Бесконтрольность означает-вседозволенность.

Не существует четкой правовой основы для сбора и последующего использования «больших данных». Например, пользователи социальной сети Facebook ежедневно регистрируют 2,7 млрд. комментариев [8]. Для социальных сетей нужна политик безопасности, опирающиеся на уже имеющиеся доверительные отношения Существует сложность управления всеми этими данными, которые многократно превышают возможности обработки. Нужны «оправданные меры» для защиты данных.

III. ОБРАБОТКА БОЛЬШИХ ПОТОКОВ ДАННЫХ С ПРИМЕНЕНИЕМ «КЛАССИЧЕСКИХ» СИСТЕМ БЕЗОПАСНОСТИ

В настоящее время происходит процесс скрещивания классических систем безопасности и полномасштабных аналитик на базе «больших данных». Рассмотрим применение технологий Big Data для решения задач обеспечения информационной безопасности.

Security Intelligence –технология, которая может получать своевременный анализ данных, от систем обеспечения информационной безопасности (ИБ) и смежных систем (кадровых, ИТ-мониторинга и т.д.), связывать их между собой, представлять данные о работе процессов ИБ для разных категорий пользователей: представителей бизнеса, руководителей подразделений ИБ, специалистов ИБ. Есть попытки связать термины Security Intelligence и Big Data. В этом контексте стоит упомянуть специальную систему IBM Security Intelligence with Big Data, которая может анализировать терабайты самых разных электронных сообщений, финансовых счетов и web-трафика. Ее цель − определить угрозы для систем безопасности и потенциальные возможности мошенничества [9,10,11].

DDoS-атаки позволяют довести до отказа практически любую систему, не оставляя юридически значимых улик [12]. Поток данных, генерируемый зомби-сетью для проведения DDoS-атаки можно отнести к разряду «большая данных», причем из него нужно выделить реальных пользователей, запросы которых надо передать по назначению, а паразитный трафик задержать. Решение может быть распределенным, построенным по принципам сети распространения контента (Content Distribution Network, CDN). Владельцы такой сети стремятся максимально приблизить данные к пользователю, передавая наиболее востребованные из них напрямую к провайдерам доступа, которые пользуются сервисом. Эту архитектуру можно «перевернуть», установив у провайдеров доступа - фильтры, каждый из которых обрабатывает получаемые от пользователей запросы и фиксирует те из них, которые являются элементами DDoS-атаки. В этом случае каждый кластер отвечает за обработку запросов, исходящих от пользователей только одного провайдера, чтобы не допустить сбора всего трафика.

Технологии «больших данных» можно применять для поиска аномалий в сетевом трафике, для защиты от целевых атак, использующих нестандартные приемы работы с сетью, такие как туннелирование в IPv6 или DNS-VPN, когда канал утечки организуется с помощью подконтрольного злоумышленнику DNS-сервиса, обменивающегося с троянцем запросами по заранее определенному протоколу. Стандартными средствами обеспечения безопасности, рассчитанными на массовые нападения, практически невозможно блокировать целевые атаки, поэтому сейчас появляются компании, которые предлагают сервисы анализа всего сетевого трафика организации для выделения необычных фрагментов [13]. В Virtual Private Network VPN - технология виртуальных защищенных сетей, созданная VPN-туннель может гибко настраиваться, чтобы соединять подсети, отдельные компьютеры и даже заключать в себя трафик отдельного заданного приложения. На концах туннелей могут находиться перечисленные объекты, однако, наряду с сетевыми объектами, VPN-туннель может опираться на индивидуальную аутентификационную информацию о конкретном пользователе. Можно построить сложную систему туннелей, с разными свойствами защиты. Одни туннели могут обеспечивать только аутентификацию отправителя и получателя информации и целостность потока данных. Другие – обеспечивать шифрование информации, причем для различных туннелей могут

применяться различные криптоалгоритмы. Туннели могут вкладываться друг в друга. Различные защищенные сети могут быть полностью изолированы, либо могут взаимодействовать по контролируемым правилам [14].

Система предотвращения вторжений (Intrusion Prevention System) − программная или аппаратная система сетевой и компьютерной безопасности, обнаруживающая вторжения или нарушения безопасности и автоматически защищающая от них. [15]. Такие системы, работающие на компьютерах компании, могут быть перенесены на облачные платформы, поддерживающие технологии «больших данных». Специальный агент собирает сведения о файлах, процессах, состоянии памяти и сети и пересылает их в облако для обработки с целью выявления аномалий в поведении приложений. По ряду признаков можно определить типы вредоносного кода − от вирусов, кардинально меняющих поведение процессов, до программ-невидимок, вмешивающихся в работу ОС [13].

В основу решения анализа корпоративных баз может быть положен аппарат управления рисками — система собирает данные из различных источников и высчитывает риски для каждого отдельного сотрудника, а потом вычисляет общий риск. Чтобы управление рисками было действительно эффективным, надо осуществлять его на непрерывной основе в течение всего жизненного цикла проекта [16]. Учитывая большие объемы сведений, целесообразно применение технологий «больших данных». Набирает обороты и событийный видеоанализ, который способен обеспечить контроль над определенными сценариями поведения объектов наблюдения. В перспективе эта функция видеоаналитики будет наиболее востебована. Сегодня технологии «больших данных» применяются для обработки видеопотоков, и есть решения по распределенному выявлению объектов и событий — например в распределенной обработке потоков данных, поступающих с камер видеонаблюдения для обеспечения автоматической реакции на нерегулярные явления. Системы обработки видеопотоков должны быть распределенными, и правила для них должны быть универсальными для всех охраняемых объектов, поэтому здесь использование технологии «больших данных» востребовано [13].

IV. ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ «БОЛЬШИХ ДАННЫХ» В ЦЕЛЯХ ЗАЩИТЫ ИНФОРМАЦИИ

Данные необходимо хранить, анализировать, обеспечивать безопасный обмен между ведомствами и обрабатывать для предоставления услуг гражданам. InfoWatch Traffic Monitor Enterprise осуществляет мониторинг и анализ данных, отправляемых за пределы корпоративной сети через почтовые системы, Web, системы обмена сообщениями, путем распечатки на локальные и сетевые принтеры и копирования на съемные устройства; выполняет автоматическую классификацию передаваемой информации; предотвращает утечки конфиденциальных данных, блокируя процесс передачи в случае обнаружения нарушения политики безопасности; обеспечивает безопасное хранение данных для анализа и проведения расследований. IWTM технология позволяет повысить точность конфиденциальных данных в общем хранилище неструктурированной информации, определять тип и тематику информации [1].

Усовершенствованная технология защиты от вредоносных программ FireAMP - интеллектуальное решение для защиты от вредоносных программ. Эта технология использует облачные технологии и средства анализа «больших данных» для обеспечения полного контроля, необходимого для обнаружения, анализа и устранения угроз, которые могут быть не обнаружены другими средствами. Также доступны средства защиты от вредоносных программ для мобильных устройств и виртуальных машин [17].

Одним из альтернативных вариантов, способных улучшить ситуацию с безопасностью, является модель нулевого доверия (Zero-Trust Model, ZTM). Использование модели ZTM приведет к генерации огромных объемов данных в реальном времени. Этот агрессивный подход к обеспечению сетевой безопасности предполагает установление контроля за всеми имеющимися данными исходя из предположения, что каждый файл таит в себе потенциальную угрозу. К выдвигаемым требованиям относятся обеспечение безопасного доступа ко всем ресурсам, предоставление доступа только тем сотрудникам, кому это действительно положено, обязательная проверка систем и исключение какого-либо доверия к ним, инспектирование всего трафика, ведение журналов, наблюдение за функционированием систем, проектирование систем изнутри наружу [18].

По оценкам Gartner, анализ «больших данных» сыграет ключевую роль в выявлении кибератак. Если сейчас лишь 8% глобальных организаций применяют анализ «больших данных» для выявления угроз безопасности и мошеннических действий, то к 2016 году их доля вырастет до 25%. «Большие данные» изменят многие продукты обеспечения безопасности компьютерных сетей, включая средства сетевого мониторинга, аутентификации и авторизации пользователей, выявления мошеннических действий. Они окажут влияние на системы управления, оценку рисков и обеспечение соблюдения нормативных требований. «Большие данные» изменят природу управления безопасностью, в том числе межсетевые экраны, антивирусы и системы предотвращения потерь данных [18].

ЗАКЛЮЧЕНИЕ

Цель аналитики «больших данных» для обеспечения безопасности состоит в массовой обработке большого количества сведений в режиме реального времени. Имеются значительные перспективы в разработке

информационных решений в комплексных системах безопасности. Однако актуальными вопросами на сегодняшний день остаются:

1. Происхождение данных: подлинность и целостность данных, используемых для аналитики по мере расширения «больших данных» источников.

2. Конфиденциальность информации в растущем потоке данных.

3. Визуализировать инструменты, которые помогут аналитики понимают данные своих систем.

Проблемы интеллектуальной собственности в глобальной сети должны рассматриваться в правовой плоскости. Необходимо регулирование и саморегулирование сетевых сообществ. Обращение к анализу процессов происходящих в сфере информационных технологий с помощью применений технологий Big Data, позволит избежать многих негативных последствий при коммуникации, и создать условия для обеспечения информационной безопасности.

ЛИТЕРАТУРА

[1] А. Данкевич, А. Насонов, “Конфиденциальность Больших Данных,” Открытые системы, №3, 2013.

[2] А.Шмид, Заметки о Big Data, 2012. http://www.interface.ru/home.asp?artId=34991

[3] B.Gerhardt, K.Griffin, R.Klemann, Получение преимуществ в условиях разобщенности систем анализа больших наборов данных (Big Data), Cisco, 2012.

[4] Вэй Тан, Б. Блейк, И. Салех, “Аналитика Больших Данных и социальные сети,” Открытые системы,2013, № 08.

[5] http://tokoorama.jofo.ru/464497.html

[6] Предиктивная медицина и Большие Данные http://apptractor.ru/info/articles/lektsiya-lorensa-dzheykobsa-prediktivnaya-meditsina-i- bolshie-dannyie.html

[7] L. Jacobs, “Big Data and Predictive Medicine.” University of Zurich Higher School of Economics, 2014.

[8] B. Gerhardt, K. Griffin, R. Klemann, “Unlocking Value in the Fragmented World of Big Data Analytics”, Cisco IBSG, 2011, p.3.

[9] Cloud Security Alliance: Big Data Analytics for Security Intelligence, 2013.

[10] И. Шабанов, Что такое Security Intelligence и зачем это нужно, 2014. http://www.antimalware.ru/analytics/Technology_Analysis/Security_Intelligence

[11] IBM Security Intelligence with Big Data. http://www-03.ibm.com/security/solution/intelligence-big-data/

[12] https://ru.wikipedia.org/wiki/DoS-атака

[13] В. Коржов, “Большие Данные в службе безопасности,” Открытые системы, 2013,№ 9,с.48-49

[14] Технологии и стандарты сетевой защиты информации http://www.s-terra.com/solutions/standards/

[15] https://ru.wikipedia.org/wiki Система_предотвращения_вторжений

[16] А.Закис, “Как внедрить управление рисками,” Intelligent Enterprise, 2003,№13

[17] Обзор решений Sourcefire® Безопасность для реального мира http://www.cisco.com/assets/global/RU/pdfs/brochures/Sourcefire-Solutions-Overview-Brochure.pdf

[18] А.Банафа, “Сетевой безопасности нужны Большие Данные,” Computerworld Россия, 2015,№ 1