УДК 004.056.5

КИБЕРБЕЗОПАСНОСТЬ БОЛЬШИХ ДАННЫХ В МЕДИЦИНЕ

А.А. Беляк
инженер ЦИИР БГУИР

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

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
E-mail: alexbeljak99@gmail.com, s.nesterenkov@bsuir.by

А.А. Беляк
Окончил Белорусский государственный университет информатики и радиоэлектроники по специальности «Программируемые мобильные системы» факультета компьютерного проектирования. Работает в отделе информационных технологий ЦИИР БГУИР в качестве инженера.

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.

Аннотация. В данной статье рассмотрены современные проблемы кибербезопасности и конфиденциальности больших данных применительно к сфере здравоохранения. Была произведена оценка, как проблемы безопасности и конфиденциальности возникают в случае больших данных в здравоохранении, и были описаны пути их решения. Основной акцент делался на недавно предложенных методах, основанных на анонимизации и шифровании, сравнения их преимуществ и недостатков, а также рассмотрение направлений будущих исследований.

Ключевые слова: кибербезопасность, жизненный цикл больших данных, большие данные в медицине.

Введение.

Большие данные в значительной степени изменили способы управления, анализа и использования данных в любой отрасли. Одной из наиболее перспективных областей, где большие данные могут быть применены – это здравоохранение. Большие данные в здравоохранении обладают значительным потенциалом для улучшения результатов лечения пациентов, прогнозирования вспышек эпидемий, получения ценных сведений, раннего выявления излечимых заболеваний, снижения стоимости оказания медицинской помощи и улучшения качества жизни в целом. Однако принятие решения о допустимом использовании данных при сохранении кибербезопасности и права пациента на конфиденциальность является сложной задачей. Большие данные, независимо от того, насколько они полезны для развития медицинской науки и жизненно важны для успеха всех организаций здравоохранения, могут быть использованы только при условии решения вопросов кибербезопасности и конфиденциальности. Чтобы обеспечить безопасную и надежную среду больших данных, важно определить ограничения существующих решений и наметить направления будущих исследований.

Безопасность больших данных в медицине.

Ввиду того, что организации здравоохранения хранят, обрабатывают и передают огромные объемы данных для обеспечения качественного и надлежащего медицинского обслуживания, недостатками являются отсутствие технической поддержки и минимальная безопасность. Ситуация осложняется тем, что отрасль здравоохранения остается одной из самых восприимчивых к утечкам данных [1]. В частности, злоумышленники могут использовать методы и процедуры интеллектуального анализа данных, чтобы узнать конфиденциальные данные и обнародовать их, в результате чего происходит утечка данных [1, 2]. В то время как внедрение мер безопасности остается сложным процессом, способы обхода средств контроля безопасности постоянно совершенствуются, что в значительной мере повышает требования к средствам обеспечения безопасности с каждым днём.

Жизненный цикл кибербезопасности больших данных.

CCW (The Chronic Conditions Data Warehouse) следует формальной модели жизненного цикла информационной безопасности, состоящей из четырех основных этапов, которые служат для выявления, оценки, защиты и мониторинга угроз безопасности данных пациентов. Эта модель жизненного цикла постоянно совершенствуется с акцентом на постоянное внимание и непрерывный мониторинг [2, 3].

В данной работе предлагается модель, которая основывается на этапах формальной модели, представленной выше с незначительными улучшениями, чтобы обеспечить политику и механизмы, которые обеспечивают устранение угроз и атак на каждом этапе жизненного цикла больших данных. На рисунке 1 представлены основные элементы жизненного цикла больших данных в здравоохранении.

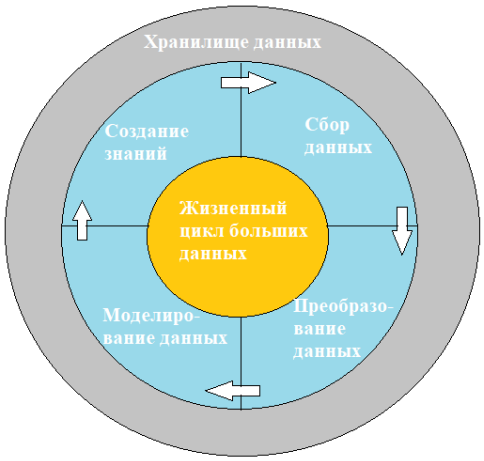

Рисунок 1. Элементы жизненного цикла больших данных в здравоохранении

Анализируя рисунок 1, можно выделить следующие элементы жизненного цикла:

1. Этап сбора данных. Он включает в себя сбор данных из различных источников в различных форматах. С точки зрения безопасности, технологии обеспечения безопасности больших данных медицины является очень важным требованием данного этапа. Поэтому важно собирать данные из доверенных источников, сохранять конфиденциальность пациентов (в базе данных не должно быть попыток идентифицировать отдельных пациентов) и обеспечить безопасность и защиту этого этапа [6]. Действительно, для обеспечения защиты всех данных и информационных систем от несанкционированного доступа, раскрытия, модификации, дублирования, перенаправления, уничтожения, потери, неправильного использования или кражи необходимо использовать некоторые проверенные меры безопасности.

2. Этап преобразования данных. Данный этап предполагает фильтрацию и

классификацию данных на основе их структуры и выполнение необходимых преобразований для проведения значимого анализа. В более широком смысле, фильтрация, классификация и преобразование данных необходимы для повышения качества данных перед этапом анализа или моделирования и удаления или соответствующей обработки неточностей, пропущенных значений, дубликатов данных и так далее. С другой стороны, собранные данные могут содержать конфиденциальную информацию, что делает чрезвычайно важным принятие достаточных мер предосторожности при преобразовании и хранении данных [4, 5]. Для того чтобы гарантировать безопасность собранных данных, они должны оставаться изолированными и защищенными путем обеспечения безопасности на уровне доступа и контроля доступа (использование обширного списка каталогов и баз данных в качестве центрального хранилища учетных данных пользователей, шаблонов входа в приложения, политик паролей и настроек клиента).

3. Этап моделирования данных. На этом этапе могут применяться контролируемые методы интеллектуального анализа данных, такие как кластеризация, классификация и ассоциация. Кроме того, существует несколько методов обучения, которые повышают точность и надежность конечной модели. С другой стороны, очень важно обеспечить безопасную среду обработки данных. Фактически, на этом этапе основное внимание специалистов по сбору данных сосредоточено на использовании мощных алгоритмов сбора данных, которые могут извлечь конфиденциальные данные. Поэтому процесс сбора данных и сетевые компоненты в целом должны быть настроены и защищены от атак на основе сбора данных и любых нарушений безопасности, которые могут произойти, а также убедиться, что только уполномоченный персонал работает на этом этапе.

4. Этап создания знаний. На этапе моделирования появляется новая информация и ценные знания, которые могут быть использованы лицами, принимающими решения. Эти созданные знания считаются конфиденциальными данными, особенно в конкурентной среде. Действительно, медицинские организации знают, что их конфиденциальные данные (например, личные данные пациентов) не должны быть обнародованы. Соответственно, соблюдение требований безопасности и проверка являются основной задачей на этом этапе.

Технологии, используемые для обеспечения кибербезопасности больших данных.

Наиболее широко используемыми технологиями являются:

1. Аутентификация. Она выполняет жизненно важные функции в любой организации: обеспечение доступа к корпоративным сетям, защита личности пользователей и гарантия того, что пользователь действительно тот, за кого он себя выдает.

2. Шифрование. Шифрование данных является эффективным средством предотвращения несанкционированного доступа к конфиденциальным данным. Оно полезно для предотвращения таких нарушений, как перехват пакетов и кража устройств хранения данных [7].

3. Маскирование данных. Оно заменяет чувствительные элементы данных на неидентифицируемое значение. Используется стратегия деидентификации наборов данных или маскировки персональных идентификаторов, таких как имя, номер социального страхования. Таким образом, маскирование данных является одним из наиболее популярных подходов к анонимизации данных в реальном времени. k-анонимичность, впервые предложенная Свани и Самрати, защищает от раскрытия личности, но не может защитить от раскрытия атрибутов. Другие методы анонимизации относятся к классам добавления шума в данные, замены ячеек в столбцах и замены групп из k записей k копиями одного представителя [8].

4. Контроль доступа. После аутентификации пользователи могут войти в информационную систему, но их доступ по-прежнему будет регулироваться политикой управления доступом, которая обычно основана на привилегиях и правах каждого работника, пациента или доверенной третьей стороны.

5. Мониторинг и аудит.

Заключение.

Таким образом, были описаны основные этапы и технологии, используемые для обеспечения кибербезопасности больших данных в медицине, а также рассмотрены их ограничения. Кроме того, существует больше различных методов, таких как шифрование на основе атрибутов, контроль доступа, гомоморфное шифрование, шифрование пути хранения и так далее. Однако проблема всегда остается актуальной.

Список использованных источников

[1] David Houlding. Health Information at Risk: Successful Strategies for Healthcare Security and Privacy. – white paper. 2011. – 250 p.

[2] James W. Transforming healthcare through big data, strategies for leveraging big data in the healthcare industry. – Institute for Health. 2013. – 322 p.

[3] Tom W. Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale 4th Edition. – O'Reilly Media, 2015. - 756 p.

[4] 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. Богуш [и др.]. - Минск : БГУИР, 2019. - С. 242-245.

[5] Калоша, А.Л. Система анализа качества текстовых коллекций / А.Л. Калоша, М.А. Медунецкий, М.П. Хоронеко, А.А. Александров, А.И. Гридасов, С.Н. Нестеренков // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. материалов VI Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 20-21 мая 2020 года): в 3 ч. Ч. 2 / редкол. : В. А. Богуш [и др.]. - Минск : Бестпринт, 2020. - С. 369-375.

[6] Bernard M. Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance. – Wiley, 2015. - 256 p.

[7] Кукареко, А.В. Способы машинного обучения для выявления ошибок выполнения упражнений на smart-тренажере / А.В. Кукареко, С.Н. Нестеренков // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. материалов VI Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 20-21 мая 2020 года): в 3 ч. Ч. 2 / редкол. : В. А. Богуш [и др.]. - Минск : Бестпринт, 2020. - С. 214-224.

[8] Samrati P. Protecting respondents identities in microdata release. – Institute for Health. 2014. – 201 p.

CYBERSECURITY OF BIG DATA IN HEALTHCARE

A.A. BELIAK
Engineer at BSUIR

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

Belarussian State University of Informatics and Radioelectronics, Republic of Belarus
E-mail: alexbeljak99@gmail.com, s.nesterenkov@bsuir.by

Abstract. This article examined current cybersecurity and privacy issues of big data as they relate to the healthcare industry. How security and privacy issues arise in the case of big data in healthcare has been assessed, and ways to address them have been described. The main focus was on recently proposed methods based on anonymization and encryption, comparing their advantages and disadvantages, and considering directions for future research.

Keywords: Cybersecurity, Big Data lifecycle, Big Data technologies, Big Data in healthcare.