УДК 004.056

ПРОБЛЕМЫ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В СИСТЕМАХ УПРАВЛЕНИЯ БОЛЬШИМИ ДАННЫМИ

М.А. Полтавцева
Санкт-Петербургский политехнический университет Петра Великого
Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29
E-mail: poltavtseva@ibks.spbstu.ru

Ключевые слова: информационная безопасность, кибербезопасность, системы управления данными, СУБД, Большие данные, защита данных.

Аннотация: В докладе рассматриваются проблемы обеспечения кибербезопасности систем управления Большими данными. Дается краткий обзор состояния этого направления, вводится понятие самой системы управления Большими данными, определяются ее архитектурные особенности. Выделяются свойства этого класса программного обеспечения с точки зрения информационной безопасности, основные угрозы и уязвимости. Автором приводится набор методов обеспечения конфиденциальности, целостности и доступности рассматриваемых систем, выделяются проблемные области исследований.

1. Введение

Цифровизация различных областей человеческой деятельности обеспечила развитие информационных технологий во всех отраслях современной экономики и промышленности. Крупномасштабные информационные системы становятся неотъемлемой частью систем управления и мониторинга, в том числе, при обеспечении информационной безопасности [1]. В то же время в основе таких систем находятся системы обработки Больших данных. Рост информации поступающей из различных источников, необходимость ее обработки и совместного анализа обуславливает востребованность и распространение соответствующих платформ управления информацией.

Системы управления Большими данными, как закономерный этап эволюции СУБД, занимают нишу не только в коммерческих продажах и маркетинговых системах, но и в финансовом и банковском секторе, промышленности, экологических системах, системах электронного государства и других. Поэтому существенно возрастает как ценность обрабатываемой информации, так и требования (в том числе, нормативные) по ее защите. Сегодня обеспечение комплексной безопасности решений по управлению Большими данными является одной из важных проблем обеспечения информационной безопасности крупномасштабных систем.

Современные исследования сосредоточены на отдельных вопросах защиты больших данных [2-3] или безопасности инфраструктуры [4], и не рассматривают комплексную кибербезопасность систем управления большими данными как отдельного, нового класса современного программного обеспечения управления и мониторинга.

2. Особенности систем управления большими данными

2.1. Понятие систем управления большими данными

На сегодняшний день нет четкого терминологического определения больших данных. Как правило, под ними понимается информация, для обработки которой не достаточно традиционных решений и используются горизонтально – масштабируемые системы [5-6]. Характерными особенностями таких данных является высокая скорость поступления информации, большой объем и разнообразие сведений [5]. «Технологии Больших данных», в свою очередь, включают многообразие как традиционных, так и новых решений в области горизонтально масштабируемых информационных систем. Системы управления Большими данными являются закономерным развитием современных СУБД. В процессе эволюции от файловых систем до современных промышленных серверов в построении систем управления данными последовательно сменялись стадии законченных программных пакетов и сочетаний готовых средств с программируемой средой. В современных высокопроизводительных системах управления данными используется большое разнообразие не только встроенных специализированных СУБД, но и программных сред.

Основным отличием систем управления данными (СУД) от остального программного обеспечения, при отсутствии выраженной границы между ними, можно указать архитектурную особенность. Системы управления данными имеют два входных потока: поток данных для обработки и хранения; поток запросов к данным. И один выходной поток, которым является результирующий набор информации, генерируемый в ответ на внешние запросы. При этом не имеет значения, являются ли запросы предустановленными или нет.

С точки зрения внутренних элементов, в системах управления Большими данными можно выделить следующие классы программного обеспечения (рис. 1):

- Транзакционные СУБД для хранения промежуточных данных в памяти («in memory» СУБД);
- Аналитические СУБД для долговременного хранения информации;
- Программные среды обработки и анализа данных.

Как правило, важным компонентом также является распределенная файловая система, необходимая для работы ряда аналитических СУБД.

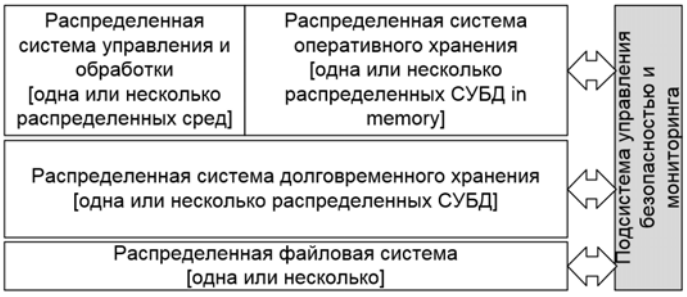

Рис. 1. Компоненты системы управления Большими данными.

В силу гетерогенности каждого класса компонент и, при этом, высокой связности между классами, возникает проблема построения подсистемы управления безопасностью таких решений, затрагивающей все уровни обработки информации.

2.2. Особенности современных СУД с точки зрения кибербезопасности

Особенности систем управления большими данными определяют их специфику с точки зрения обеспечения кибербезопасности. Исходя из понятия и характеристик больших данных[5-6], выделим ряд важных характеристик современных СУД.

Во-первых, большой объем данных, не только с точки зрения целевой обработки, но и с точки зрения систем защиты. В связи с большим числом элементов информации, возникают задачи ведения больших реестров гранулированных данных, политик безопасности и их распределенного хранения. Запрос прав доступа, направленный при совершении операции на другой узел системы может замедлить целевой процесс до недопустимых значений.

Во-вторых, большая скорость поступления данных. Помимо требований к скорости обработки запроса доступа, необходимо оперативно вносить и учитывать информацию о поступающих данных, а также осуществлять противодействие нарушителю. Для решения этой задачи подсистема управления информационной безопасностью должна иметь высокую скорость реакции на изменения в данных.

В-третьих, в современных СУД используется множество узлов, осуществляющих обработку информации, причем одни и те же данные могут поступать из разных источников; один и тот же источник данных может поставлять различные сведения; число и состав обработчиков информации при одном и том же наборе операций обработки может существенно меняться во времени с сохранением прозрачности, то есть, незаметно для источника данных и их получателя.

Первые два свойства выдвигают проектные требования к подсистеме управления безопасностью. Третья характеристика задает принципиально новые условия для обеспечения конфиденциальности, доступности и целостности информации, даже в условиях доверенной облачной среды. Также СУД обладают свойством, характерным для задач информационной безопасности в области традиционных СУБД: это двойственная природа защищаемого объекта, когда целью нарушителя может быть как данные, так и управляющее программное обеспечение.

Важным фактором является также смещение приоритетов при реализации кибербезопасности крупномасштабных систем, в частности по управлению данными [7]. Наиболее важной становится доступность системы, то есть, ее способность осуществлять прием и обработку данных, формирование результатов запросов. Целостность информации уступает доступности и наименее важной (за исключением систем, оперирующих с информацией, относящейся к государственной или иным видам тайн) становится конфиденциальность сведений. Это связано с тем, что издержки от нарушения доступности систем во много раз превышают издержки от других видов нарушений.

3. Обеспечение кибербезопасности систем управления большими данными

3.1. Угрозы и уязвимости систем управления большими данными

Высокая связность компонентов систем управления Большими данными и их разнообразие, а также тот факт – что подавляющее большинство из них является открытым программным обеспечением, обеспечивает множество возможных воздействий злоумышленника, даже в условиях доверенной инфраструктуры.

На уровне распределенной файловой системы возможен физический доступ к файлам данных или доступ через сетевую инфраструктуру самой файловой системы и ее модули управления.

На уровне распределенных СУБД для долговременного хранения возможен несанкционированный доступ к данным и атаки на превышение допустимой нагрузки средствами СУБД (повышение привилегий, логический вывод, инъекции в языки запросов и другие уязвимости), а реализация этой угрозы - через физический доступ к узлу обработки или через сетевую инфраструктуру (обмен данными внутри СУБД).

Тем же спектром уязвимостей, что и системы хранения, обладают СУБД транзакционной обработки in-memory, так как сегодня они допускают кластерную организацию с горизонтальным масштабированием. Дополнительной особенностью является чтение данных из этих СУБД напрямую при получении доступа к оперативной памяти, аналогично широко применяемому злоумышленниками чтению данных из кэша.

Распределенная среда управления и обработки, представляющая собой программируемый, в том числе - на популярных универсальных языках (python, java) уровень обработки, помимо уязвимостей любого распределенного программного обеспечения допускает подмену источников, получателей и обработчиков данных, несанкционированный доступ к информации в процессе обработки (в том числе, через оперативную память), атаки на превышение допустимой нагрузки.

3.2. Обеспечение кибербезопасности СУД

Обеспечение доступности – наиболее разработанная область обеспечения кибербезопасности систем управления Большими данными. Это связано с высокой критичностью приложений современных СУД и необходимостью обеспечения устойчивого функционирования в условиях программных сбоев и отказов оборудования. Обеспечение доступности осуществляется путем избыточности обоих компонентов системы:

- данных, путем их репликации в СУД и фрагментирования с шардированием нагрузки;
- программного обеспечения и оборудования, с использованием параллельной обработки на кластере с дублированием узлов.

Специфическими для информационной безопасности подходами к обеспечению доступности являются средства выявления и противодействия атакам на доступность. К этому классу относятся как DDOS –атаки, так и направленные атаки на ресурсы, данные и программное обеспечение с целью нарушения их работоспособности. Примером может служить использование вирусов – шифровальщиков.

Косвенно к этой области относятся и вопросы построения подсистемы управления безопасностью, в том числе, реализация контроля доступа с соблюдением требований по быстродействию всего комплекса.

Обеспечение целостности данных в системах управления ими традиционно осуществляется за счет обеспечения согласованности информации. В развитых универсальных СУБД для этого используются ограничения целостности и механизмы сериализации транзакций. В распределенной среде СУД задачи обеспечения целостности осложняются: распределенными транзакциями (и конфликтом между их сериализацией и производительностью); дублированием, в том числе – не идемпотентных транзакций, для обеспечения доступности; необходимостью согласования реплик данных в хранилищах. Несанкционированное изменение данных злоумышленником, нарушающее их целостность, также является существенной проблемой. Механизмы защиты локальных СУБД от несанкционированного доступа обладают рядом уязвимостей и не предназначены для обеспечения безопасности в условиях распределенной среды и выхода за гра-

ницы отдельных программных и физических компонентов при распределенной организации.

Обеспечение конфиденциальности сегодня одна из самых неразработанных областей информационной безопасности управления Большими данными. Так как функционал обеспечения конфиденциальности не связан напрямую с производительностью системы и решением ею ежедневных задач, и при этом имеет отрицательное влияние на производительность, как и в СУБД, он не входит в список приоритетных задач производителей и разработчиков соответствующего программного обеспечения. Открытое распространение большинства компонентов высокопроизводительных СУД дополнительно способствует высокой осведомленности нарушителя. Единственное, что сегодня предлагается в данной области – использование гомоморфного шифрования, обладающего хорошим потенциалом и для традиционных СУБД, однако его внедрение потребует значительных усилий, а криптостойкость многих гомоморфных алгоритмов не высока. Реализация комплексной политики безопасности в условиях множества обработчиков сетевой инфраструктуры систем управления Большими данными является неизученной областью как в теоретическом, так и практическом плане.

4. Заключение

Современные системы управления Большими данными обладают рядом особенностей, отличающих их как от традиционных СУБД, так и других классов программного обеспечения. Кибербезопасность СУД является важным компонентом обеспечения информационной безопасности современных систем управления и критической инфраструктуры. Специфические угрозы и уязвимости СУД обусловлены их распределенной структурой и архитектурными особенностями.

Во многих современных систем управления Большими данными такое свойство доступность является более значимым (экономически важным), чем целостность или конфиденциальность информации. Принципы и технологии построения распределенных систем, к которым относятся эти СУД, включают компоненты избыточности и управления производительностью, как обеспечения доступности, и управление транзакциями и согласованностью данных. И если эти аспекты развиваются вместе с теорией управления обработкой Больших данных, задача обеспечения конфиденциальности, включающая такие специфические аспекты как шифрование информации и управление доступом, остается не решенной. Наиболее острой проблемой являются теоретические и практические исследования в области новых подходов, моделей, методов и средств управления доступом и реализации политики безопасности в условиях множества узлов - обработчиков гетерогенного кластера.

Список литературы

1. Полтавцева М.А. Особенности применения технологий обработки больших данных в задачах обеспечения кибербезопасности // Методы и технические средства обеспечения безопасности информации. СПб.: Изд-во политехнического университета, 2018. № 27. С. 4-7.
2. Chechulina D., Shatilov K., Krendelev S. Fully Homomorphic Encryption for Secure Computations in Protected Database // Proceedings of the Federated Conference on Computer Science and Information Systems, 2015. P. 125-131.
3. Gupta A., Pandhi K., Bindu P. V. B., Thilagam P. S. Role and access based data segregator for security of big data // Procedia Technology, 2016. vol. 24. P. 1550-1557.

4. Jansen W., Grance T. Guidelines on Security and Privacy in Public Cloud Computing // NIST Special Publication. 2011. http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf.
5. Майер-Шенбергер В., Кукьер К.. Большие данные. Революция, которая изменит то, как мы живём, работаем и мыслим М.: Манн, Иванов, Фербер, 2014. 240 с.
6. Min Chen, Shiwen Mao, Yin Zhang, Victor C.M. Leung. Big Data. Related Technologies, Challenges, and Future Prospects. Spinger, 2014. 100 p. DOI:10.1007/978-3-319-06245-7.
7. Васильев Ю.С., Зегжда Д.П., Полтавцева М.А. Проблемы безопасности цифрового производства и его устойчивость к киберугрозам // Проблемы информационной безопасности. Компьютерные системы. 2017. № 4. С. 47-63.