УДК 004.056 doi: 10.15827/0236-235X.148.487-494 2024. Т. 37. № 4. С. 487–494

Аналитическая обработка больших массивов данных о событиях кибербезопасности с применением суперкомпьютерных вычислений

И.В. Котенко [1]✉, И.Б. Саенко [1], И.Б. Паращук [1], В.А. Десницкий [1], Л.А. Виткова [1]

[1] Санкт-Петербургский федеральный исследовательский центр РАН, г. Санкт-Петербург, 199178, Россия

Ссылка для цитирования
Котенко И.В., Саенко И.Б., Паращук И.Б., Десницкий В.А., Виткова Л.А. Аналитическая обработка больших массивов данных о событиях кибербезопасности с применением суперкомпьютерных вычислений // Программные продукты и системы. 2024. Т. 37. № 4. С. 487–494. doi: 10.15827/0236-235X.148.487-494
Информация о статье
Группа специальностей ВАК: 2.3.6
Поступила в редакцию: 20.06.2024 После доработки: 27.08.2024 Принята к публикации: 30.08.2024

Аннотация. Вопрос кибербезопасности критических инфраструктур осложняется неизбежностью обработки больших объемов данных о событиях безопасности. Это приводит к необходимости разработки информационной технологии, сочетающей аналитическую обработку с суперкомпьютерными вычислениями. Предложены общая схема такой технологии и архитектура реализующей ее системы. В системе выделены компоненты для обнаружения в реальном времени компьютерных атак, аномальной активности и нарушений политик безопасности. Кроме того, компоненты системы позволяют оперативно оценивать защищенность сетевых ресурсов, анализировать риски, принимать решения по защите сетевых ресурсов, расследовать компьютерные инциденты, визуализировать большие массивы данных о событиях кибербезопасности и взаимодействовать с суперкомпьютерным центром. При выборе решения использовались принципы датацентричности, открытой сервис-ориентированной архитектуры и платформенности. Представлено высоко- и низкоуровневое описание архитектуры системы. Продемонстрированы экспериментальные результаты, полученные в суперкомпьютерном центре «Политехнический». Оценка разработанной техники выполнялась с использованием набора данных HAI, собранного на испытательном стенде промышленной системы управления паровыми турбинами. Решена задача прогнозирования будущих состояний на основании предыдущих, полученных путем кластеризации системных событий. Реализованный метод прогнозирования показал, что точность зависит от количества учитываемых предыдущих состояний и дальности предсказания. Эти результаты подтвердили эффективность предложенной информационной технологии и продемонстрировали ее высокую производительность.

Ключевые слова: информационная технология, кибербезопасность, большие данные, событие кибербезопасности, суперкомпьютерные вычисления

Благодарности. Исследование поддержано РНФ, грант № 21-71-20078, в СПб ФИЦ РАН

Введение. Расширение масштабов задач по мониторингу и контролю кибербезопасности, рост объемов собираемых данных о событиях кибербезопасности, а также развитие средств и методов их надежного хранения привели к повышению актуальности разработки новых методов и алгоритмов анализа и обработки больших массивов данных в системах кибербезопасности [1–3]. В современных критических инфраструктурах эта задача требует привлечения технологий интеллектуальной аналитической обработки данных, инновационных методов оценки смыслового содержания информации об угрозах, а также методов и средств реализации высокопроизводительных вычислений, включая суперкомпьютерные [4–6].

Аналитическая обработка больших массивов данных нужна для оперативной и достоверной оценки состояния защищаемой системы, поддержки принятия решений и расследования компьютерных инцидентов. При этом исследование информации о событиях кибербезопасности, идентификация рисков, а также выработка мер по противодействию угрозам зачастую осуществляются в условиях неопределенности [7]. В современных критических инфраструктурах этот процесс представляет собой технологию целенаправленного поиска информации в массивах гетерогенных данных о подобных событиях. Эта технология подразумевает использование статистических, оптимизационных и других математических алгоритмов, позволяющих находить взаимозависимости (корреляция, классификация и т.п.) и синтезировать дедуктивную информацию [8]. Кроме того, аналитическая обработка должна осуществляться с использованием современных когнитивных методов и алгоритмов, таких как нечеткие, нейросетевые и нейро-нечеткие методы, биоинспирированные алгоритмы оптимизации, ме-

тоды распознавания образов, алгоритмы визуализации данных и проч. Учитывая, что обработке подлежат большие объемы данных, применение суперкомпьютерных вычислений становится необходимым условием достижения требуемой эффективности.

Целью статьи является изложение результатов разработки инновационной информационной технологии аналитической обработки больших массивов данных о событиях кибербезопасности, основанной на применении суперкомпьютерных вычислений.

Обзор релевантных работ

Многие авторы рассматривают различные подходы к применению алгоритмов и отдельных средств для обработки больших массивов гетерогенных данных, к анализу и оценке состояния политик безопасности на основе результатов такой обработки, а также к оценке их защищенности. Однако практическое применение этих подходов продолжает оставаться затруднительным.

В значительной степени это обусловлено необходимостью учета переходных процессов, протекающих в подобных инфраструктурах (например, в энергетике, железнодорожном транспорте, инфраструктуре управления большим городом). Они имеют многокритериальный характер требований, предъявляемых к кибербезопасности, и обусловливают постановку не только линейных, но и нелинейных нестатистических задач анализа и обработки больших массивов данных [9, 10]. Подходы к решению этих задач в рамках существующих методик не рассматривались.

В работах [11, 12] рассматриваются методики сбора текущей статистики и предобработки большого количества собранных гетерогенных данных о событиях кибербезопасности. Они влекут большие временные затраты, что негативно влияет на общее время обработки и оперативность оценки состояния, поддержки принятия решений и расследования компьютерных инцидентов.

Предложенные в работах [5–7] частные методики обработки и оценки защищенности данных работают с большими массивами информации и ориентированы на условия неопределенности, но не учитывают применение суперкомпьютерных вычислений.

Методы поиска и обработки информации, использующие анализ взаимозависимостей параметров кибербезопасности с точки зрения их корреляции, рассмотрены в работах [13, 14]. Подобные методы сложны, поскольку применение алгоритмов статистической обработки для корреляции событий безопасности связано с обеспечением соответствия исходных данных требованиям по их однородности.

Особого внимания заслуживают исследования, касающиеся суперкомпьютерных вычислений, в которых рассматриваются вопросы обеспечения кибербезопасности самого суперкомпьютера [15–17], а не использования его вычислительных мощностей для решения задач безопасности. В некоторых работах исследуется роль суперкомпьютера для обеспечения национальной безопасности или совершенствования вооружения [18, 19]. Выявлено, что вопрос применения суперкомпьютера для обеспечения кибербезопасности не получил широкого обсуждения в научной литературе.

Таким образом, анализ релевантных работ позволяет говорить не только об актуальности, но и об объективной необходимости формирования информационной технологии, позволяющей реализовать интеллектуальные подходы к аналитической обработке, применяя при этом суперкомпьютерные вычисления. Данная технология должна охватывать оценку состояния, поддержку принятия решений и расследование компьютерных инцидентов. Основными областями применения этой технологии являются критические информационные инфраструктуры, отличающиеся повышенными требованиями к кибербезопасности.

Содержание информационной технологии аналитической обработки больших массивов данных о событиях кибербезопасности

В широком смысле технология – это совокупность методов, процессов и материалов, используемых в какой-либо отрасли деятельности, а также научное описание способов производства. В узком смысле слова технология – это комплекс организационных мер, операций и приемов, направленных на изготовление, обслуживание, ремонт и/или эксплуатацию изделия с номинальным качеством и оптимальными затратами, обусловленных текущим уровнем развития науки, техники и общества в целом [20]. При этом процесс понимается как совокупность действий, направленных на достижение поставленной цели.

Существует несколько определений, поясняющих современную сущность информаци-

онной технологии. Наиболее близким к задачам аналитической обработки больших массивов данных о событиях кибербезопасности с помощью суперкомпьютера является следующее трактование: совокупность методов, производственных процессов и программно-технических средств, объединенных в технологическую цепочку, которые обеспечивают сбор, хранение, обработку, вывод и распространение информации для снижения трудоемкости процессов использования информационных ресурсов, для повышения их надежности и оперативности.

Различают три класса информационных технологий, ориентированных на различные предметные области: *глобальный*, включающий модели, методы и средства, формализующие и позволяющие использовать информационные ресурсы общества в целом; *базовый*, предназначенный для определенной области применения; *конкретный*, реализующий обработку определенных данных при решении конкретных функциональных задач пользователя (планирование, учет, анализ и проч.).

Информационная технология предусматривает технические, коммуникационные средства, организационно-методическое обеспечение и стандартизацию.

Требования, предъявляемые к информационной технологии: высокая степень разделения процесса обработки информации на этапы, включение всего набора элементов для достижения поставленной цели. Кроме того, необходимо наличие регулярного характера – этапы технологического процесса должны быть стандартизированы и унифицированы для более эффективного управления информационными процессами.

К свойствам информационной технологии относятся целесообразность, наличие компонентов и структуры, взаимодействие с внешней средой, целостность, развитие во времени.

Современные и перспективные критически важные инфраструктуры являются киберфизическими системами. Для них характерны: большой парк электронных устройств, огромные объемы данных о событиях безопасности, собираемые для последующего анализа, возможное наложение ограничений на коммуникационно-вычислительные ресурсы этих устройств, большое число пользователей, имеющих доступ к этим устройствам. В результате чего они оказываются подвержены атакам известных и новых видов, нередко целевого назначения.

Для выявления атак и принятия адекватных мер противодействия необходимо проводить сбор и анализ больших объемов разнородной информации по кибербезопасности в кратчайшие сроки, соответствующие реальному масштабу времени или близкому к нему. Эти функции реализуют системы управления информацией и событиями безопасности (*Security Information and Event Management*, SIEM) [21, 22].

Как правило, в SIEM-системах выделяют три уровня построения. На первом, нижнем уровне осуществляются сбор и предварительная обработка данных о событиях безопасности. На втором уровне реализуется поддержка хранилища данных. На третьем, верхнем уровне выполняются окончательный анализ всей собранной информации по кибербезопасности и выработка мер противодействия. Аналитическая обработка больших массивов данных предполагает реализацию функций второго и третьего уровней SIEM-системы.

В существующих и разрабатываемых перспективных SIEM-системах эти функции включают оценку состояния или текущей ситуации по безопасности – обеспечение осведомленности о безопасности, выработку и выбор вариантов мер противодействия атакам, расследование последствий и причин реализации атак.

В свою очередь, в осведомленность о безопасности входит следующее:

– восприятие ситуации, благодаря чему администратор владеет доступной оперативной информацией о текущей ситуации и накапливает ее;

– оценка воздействия, позволяющая понимать характер и последствия влияния атаки;

– отслеживание ситуации, заключающееся в понимании ее дальнейшего развития;

– анализ тренда атаки и намерений нарушителей;

– анализ причинно-следственных связей;

– оценка достоверности данных о ситуации и ее развитии, заключающаяся в прогнозировании будущих возможных действий нарушителей, в понимании их намерений, возможностей и ресурсов, а также в понимании собственных уязвимостей, возможных контрмер.

Кроме того, при оценке состояния кибербезопасности применяется визуальный анализ данных с помощью стандартных или специально разработанных для этой цели нестандартных моделей визуализации.

Таким образом, в содержательном плане информационная технология аналитической обработки данных о кибербезопасности включает обнаружение в реальном времени компьютерных атак на основе аналитического и имитацион-

ного моделирования и аномальной активности и нарушений критериев и политик кибербезопасности, оперативную оценку защищенности информационных, телекоммуникационных и других критически важных ресурсов, оперативный анализ и управление рисками кибербезопасности, выработку и выбор критериев оценки состояния, поддержку принятия решений, расследование компьютерных инцидентов на основе аналитической обработки больших массивов гетерогенных данных о событиях кибербезопасности, оперативную визуализацию больших массивов данных о событиях кибербезопасности.

Основной целью информационной технологии аналитической обработки больших массивов данных о событиях кибербезопасности, основанной на применении суперкомпьютерных вычислений, является обеспечение надежного и устойчивого сбора, предварительной и итоговой интеллектуальной обработки больших информационных объектов – данных об атаках и иных инцидентах, их достоверном и масштабируемом оценивании в интересах оперативного анализа состояния, поддержки принятия решений и расследования.

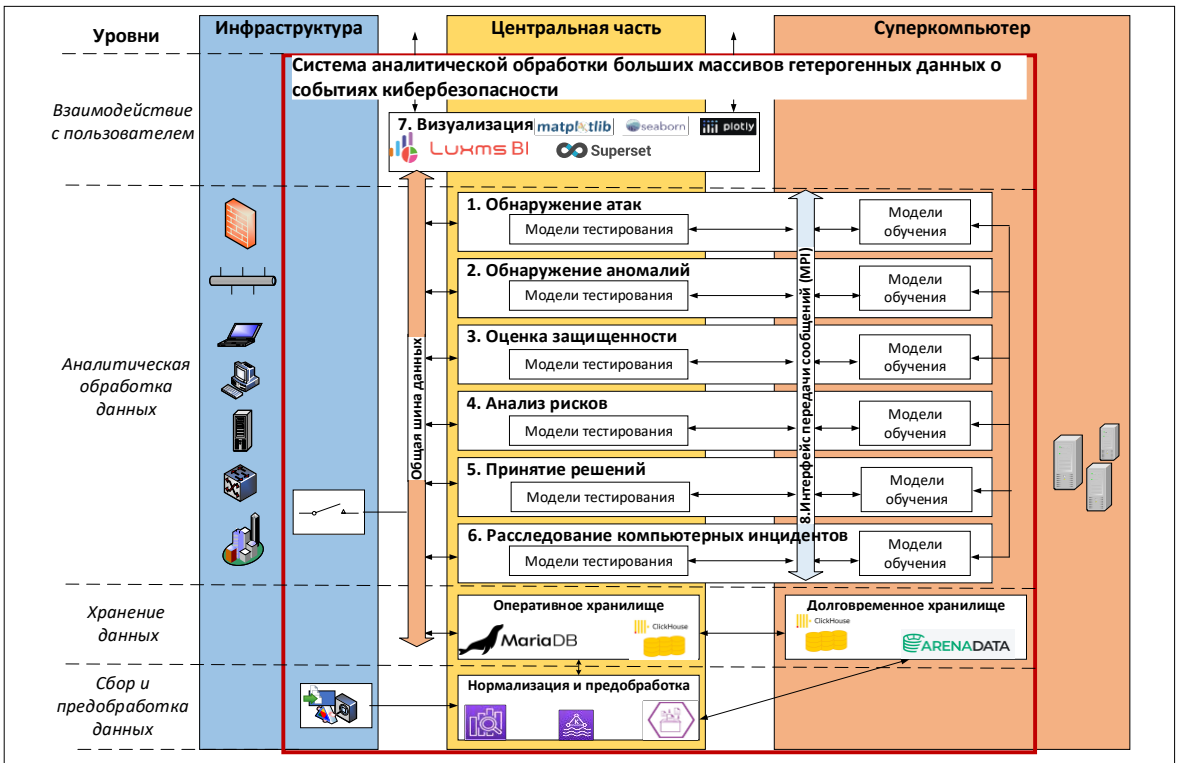
Архитектура системы аналитической обработки, использующей суперкомпьютерные вычисления

Рассмотрим особенности архитектуры разработанной системы (см. рисунок).

– Система является распределенной, включает часть защищаемой инфраструктуры, часть оборудования *суперкомпьютерного центра* (СКЦ) и центральную часть (ядро), в которой находятся аналитики кибербезопасности.

– Компоненты 1–6 предназначены для решения конкретных задач по анализу событий безопасности и принятию решений по противодействию атакам. Компонент 7 предназначен для визуального анализа событий безопасности, а также для представления промежуточных и конечных результатов работы других компонентов.

– Каждый из компонентов 1–6 состоит из двух частей: центральной, где обученные модели применяются для тестирования (решаются задачи по выявлению атак, аномалий, по оценке защищенности и проч.), и удаленной, расположенной на стороне СКЦ, где осуществляется обучение моделей. Обмен между ними происходит в компоненте 8 по технологии MPI.



Архитектура системы аналитической обработки больших массивов данных о событиях кибербезопасности, использующей суперкомпьютерные вычисления

Architecture of a system for analytical processing of large data sets on cybersecurity events using supercomputing

– На уровнях «Сбор и предобработка данных» и «Хранение данных» выделяются два типа хранилищ: оперативное, которое находится в ядре, и долговременное, расположенное в СКЦ.

– Обмен между компонентами осуществляется через общую шину данных, которая отвечает за доведение данных до администратора и регуляторов.

Технологический стек предлагаемой системы менялся в ходе исследований в связи с обновлением карты доступных решений. Изначально рассматривались программные продукты с открытым исходным кодом или доступные к использованию по лицензии GNU. В конечном итоге был сделан выбор в пользу решений, которые не ограничивают территориальное использование своих компонентов. Так, для уровня 7 были протестированы платформы анализа и визуализации данных Superset BI (https://superset.apache.org), Datalens yandex (https://cloud.yandex.com/services/datalens) и Luxms BI (https://luxmsbi.com).

При выборе решения основными критериями стали требования соответствия следующим принципам: датацентричность (логика управления процессами кибербезопасности рядом с данными), открытая сервис-ориентированная архитектура (открытое API для интеграции со сторонними программными продуктами), платформенность (поддержка процесса настройки правил корреляции, обнаружения, выбора уровня риска силами пользователя).

Для реализации уровня «Хранение данных» была предложена двухзвенная клиент-серверная архитектура, при которой на сервере находятся балансировщик нагрузки и серверная часть. Сервер анализаторов располагается внутри БД, основная логика управления процессами кибербезопасности реализована на языке PL/pgSQL. Кроме того, на сервере выполняются приложения, реализующие функции хранения данных, управления очередями сообщений (https://kafka.apache.org), управления конфигурацией и мониторингом сервисов (https://kubernetes.io), выполнения задач извлечения, преобразования и загрузки, а также обмена данными с внешними системами для контроля эффективности и качества работы анализаторов.

Предложенная двухзвенная архитектура превосходит трехзвенную по показателям скорости обработки данных и времени отклика. За счет сокращения количества звеньев экономится время на выборку данных из БД в сервер приложений, что позволяет снизить объем сетевого трафика. Таким образом, за счет устранения лишних этапов передачи информации платформа способна эффективно обрабатывать практически неограниченные объемы данных. Это позволяет в полной мере использовать возможности внешних СУБД для аналитической обработки данных. Для хранения и управления метаданными в системе используется Maria DB, основная БД Clickhouse (https://clickhouse.com), а для долговременного хранения данных – Arenadata Hadoop (https://www.arenadata.io/hadoop).

Экспериментальная оценка информационной технологии

Разработанная технология прошла экспериментальную оценку на кластере «РСК Торнадо» в СКЦ «Политехнический», который находится на 4-м месте в российском рейтинге и на 22-м – в мировом (https://rscgroup.ru/project/spbstu-politechnic/). Кластер содержит 612 узлов, каждый из которых имеет следующие характеристики: 2 процессора Intel Xeon CPU E5-2697 v3 @ 2.60 ГГц, 28 ядер и 56 потоков суммарно, 64 Гб оперативной памяти и 1 Пб общей для всех узлов памяти.

Оценка выполнялась с использованием набора данных HAI (https://www.kaggle.com/datasets/icsdataset/hai-security-dataset), который был собран на испытательном стенде промышленной системы управления паровыми турбинами, имитирующем выработку электрической и гидроаккумулирующей энергии. Длина временного ряда в наборе данных была равна 361 200, количество признаков – 86. Решалась задача прогнозирования будущих состояний на основании предыдущих, полученных путем кластеризации системных событий. Метод предсказания основан на рекуррентной нейронной сети, работающей в режиме классификации и состоящей из двух слоев – LSTM и Dense. Слой LSTM по умолчанию имел 512 входов, а слой Dense содержал количество выходов, равное количеству предсказуемых классов.

Реализованный метод прогнозирования показал, что его точность зависит от количества учитываемых предыдущих состояний (NPS) и дальности предсказания (PR). Так, для NPS = 1 получена точность 0,73 при PR = 1 и 0,61 при PR = 9. Для NPS = 4 получена точность 0,82 при PR = 1 и 0,68 при PR = 10.

Таблица представляет данные о времени, затраченном на построение матрицы состояний, при использовании СКЦ и обычного *пер-*

сонального компьютера (ПК) при различном количестве потоков.

Сравнительная оценка времени построения матрицы состояний в зависимости от количества потоков для СКЦ и ПК

Comparative assessment of state matrix construction time depending on the number of flows for SCC and PC

Количество потоков	Время, сек.	
	СКЦ	ПК
1	724,5	538,0
3	306,3	283,8
10	147,8	259,4
15	132,0	220,1

Время построения матрицы состояний при работе на 15 потоках СКЦ уменьшается на 40 %, по сравнению с работой ПК на 15 потоках, и на 75 % при работе ПК на одном потоке.

Таким образом, эксперименты показывают, что разработанная информационная технология для обработки больших массивов данных о событиях кибербезопасности, использующая суперкомпьютерные вычисления, демонстрирует существенный выигрыш во времени решения задач аналитической обработки.

Заключение

В статье рассмотрены основные положения и обосновано содержание разработанной информационной технологии аналитической обработки больших массивов данных о событиях кибербезопасности, использующей суперкомпьютерные вычисления. Описана архитектура и представлены технологические аспекты реализующей ее системы. Приведены экспериментальные результаты оценки разработанной технологии на СКЦ «Политехнический». Дальнейшие исследования связаны с апробацией разработанной технологии на различных типах защищаемых инфраструктур.

Список литературы

1. Alani M.M. Big data in cybersecurity: A survey of applications and future trends. J. of Reliable Intelligent Environments, 2021, vol. 7, pp. 85–114. doi: 10.1007/s40860-020-00120-3.

2. Verma R., Bhatt R. Security issues and challenges of big data analytics. Proc. Int. Conf. PDGC, 2022, pp. 61–66. doi: 10.1109/PDGC56933.2022.10053205.

3. Arya A., Malhotra H., Dayanand, Jeberson W. Big data analytics in cyber security. IJERT, 2017, vol. 5, no. 10, pp. 1–3.

4. Andrade R.O., Ontaneda N., Silva A., Tello-Oquendo L. et al. Application of big data analytic in cybersecurity. Proc. Int. Conf. ACC, 2020, pp. 26–32.

5. Котенко И.В., Саенко И.Б., Браницкий А.А., Паращук И.Б., Гайфулина Д.А. Интеллектуальная система аналитической обработки цифрового сетевого контента для защиты от нежелательной информации // Информатика и Автоматизация. 2021. Т. 20. № 4. С. 755–788. doi: 10.15622/ia.20.4.1.

6. Parashchuk I., Doynikova E., Saenko I., Kotenko I. Selection of countermeasures against harmful information based on the assessment of semantic content of information objects in the conditions of uncertainty. Proc. Int. Conf. INISTA, 2020, pp. 1–7. doi: 10.1109/INISTA49547.2020.9194680.

7. Kotenko I.V., Saenko I.B., Parashchuk I.B., Doynikova E.V. An approach for selecting countermeasures against harmful information based on uncertainty management. ComSIS, 2022, vol. 19, no. 1, pp. 415–433. doi: 10.2298/CSIS210211057K.

8. Tf M.R., Singh Y. An exploration on big data analysis and data mining methods. Proc. INCOFT, 2022, pp. 1–6. doi: 10.1109/INCOFT55651.2022.10094454.

9. Kamara M.K. Securing Critical Infrastructures. Xlibris US, Bloomington, 2020, 224 p.

10. Samanis E., Gardiner J., Rashid A. Adaptive cyber security for critical infrastructure. Proc. ICCPS, 2022, pp. 304–305. doi: 10.1109/ICCPS54341.2022.00043.

11. Ekpo U. Introduction to Cyber Security: Fundamentals. Independently published, NY, 2018, 37 p.

12. Srivastava N., Jaiswal U.C. Big data analytics technique in cyber security: A review. Proc. ICCMC, 2019, pp. 579–585. doi: 10.1109/ICCMC.2019.8819634.

13. Bothos M.A., Thanos K.G., Kyriazanos D.M. et al. Correlation and dependence analysis on cyberthreat alerts. ITU J.: ICT Discoveries, 2018, vol. 1, no. 2, pp. 1–6.

14. Zhang K., Zhao F., Luo S., Xin Y., Zhu H. An intrusion action-based IDS alert correlation analysis and prediction framework. IEEE Access, 2019, vol. 7, pp. 150540–150551. doi: 10.1109/ACCESS.2019.2946261.

15. Zhu G., Zeng Y., Guo M. A security analysis method for supercomputing users' behavior. Proc. Int. Conf. CSCloud, 2017, pp. 287–293. doi: 10.1109/CSCloud.2017.19.

16. Баранов А.В., Корепанов П.М., Кузнецов Е.Е. Обеспечение информационной безопасности научного суперкомпьютерного центра // Программные продукты и системы. 2023. Т. 36. № 4. С. 615–631. doi: 10.15827/0236-235X.144.615.

17. Yang B., Yu Y., Wang Z., Li Sh. et al. Research on network security protection of application-oriented supercomputing center based on multi-level defense and moderate principle. JPCS, 2021, vol. 1828, art. 012114. doi: 10.1088/1742-6596/1828/1/012114.

18. Агеева А.Ф. Роль суперкомпьютеров в вопросах национальной безопасности // Вестн. академии. 2023. № 1. С. 49–62. doi: 10.51409/v.a.2023.03.01.005.

19. Yalcin H., Daim T., Moughari M.M., Mermoud A. Supercomputers and quantum computing on the axis of cyber security. Tech. in Society, 2024, vol. 77, art. 102556. doi: 10.1016/j.techsoc.2024.102556.

20. Несмиянова И.О. Информационные технологии: этапы развития, понятие и классификация // Изв. ТулГУ. Экономические и юридические науки. 2020. № 1. С. 149–155.

21. Котенко И.В., Саенко И.Б., Захарченко Р.И., Величко Д.В. Подсистема предупреждения компьютерных атак на объекты критической информационной инфраструктуры: анализ функционирования и реализации // Вопросы кибербезопасности. 2023. № 1. С. 13–27. doi: 10.21681/2311-3456-2023-1-13-27.

22. Ададуров С.Е., Глухов А.П., Котенко И.В., Саенко И.Б. Интеллектуальные сервисы обеспечения информационной безопасности // Автоматика, связь, информатика. 2022. № 3. С. 27–30.

Software & Systems doi: 10.15827/0236-235X.148.487-494 2024, 37(4), pp. 487–494

Analytical processing of large data sets of cybersecurity events using supercomputing

Igor V. Kotenko [1]✉, Igor B. Saenko [1], Igor B. Parashchuk [1], Vasily A. Desnitsky [1], Lydia A. Vitkova [1]

[1] St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, 199178, Russian Federation

For citation

Kotenko, I.V., Saenko, I.B., Parashchuk, I.B., Desnitsky, V.A., Vitkova, L.A. (2024) 'Analytical processing of large data sets of cybersecurity events using supercomputing', *Software & Systems*, 37(4), pp. 487–494 (in Russ.). doi: 10.15827/0236-235X.148.487-494

Article info

Received: 20.06.2024 After revision: 27.08.2024 Accepted: 30.08.2024

Abstract. The issue of cybersecurity of critical infrastructures is complicated by the need to process large data sets of security events. This leads to the need to develop information technology that combines analytical processing with supercomputing. The authors proposed a general scheme of such technology and architecture of the system realizing it. The system contains components that realize real-time detection of computer attacks, abnormal activity and security policy violations. Furthermore, the system components allow promptly assessing the security of network resources, analyzing risks, making decisions on the protection of network resources, investigating computer incidents, visualizing large data sets of cybersecurity events and interacting with the supercomputer center. The authors used the principles of data-centricity, open service-oriented architecture and platform to select the solution. They presented a high-level and low-level description of the system architecture. The authors demonstrated experimental results obtained at the Polytechnic supercomputer center. They evaluated the developed technique using HAI dataset collected on a testbed of an industrial steam turbine control system. The authors solved the problem of predicting future states based on previous states obtained by clustering system events. The realized prediction method showed that the accuracy depends on the number of considered previous states and the prediction range. These results confirmed the effectiveness of the proposed information technology and demonstrated its high performance.

Keywords: information technology; cybersecurity; large data sets; cybersecurity event; supercomputing

Acknowledgements. The study was supported by the Russian Science Foundation grant no. 21-71-20078 in St. Petersburg Federal Research Center of the Russian Academy of Sciences

References

1. Alani, M.M. (2021) 'Big data in cybersecurity: A survey of applications and future trends', *J. of Reliable Intelligent Environments*, 7, pp. 85–114. doi: 10.1007/s40860-020-00120-3.

2. Verma, R., Bhatt, R. (2022) 'Security issues and challenges of big data analytics', *Proc. Int. Conf. PDGC*, pp. 61–66. doi: 10.1109/PDGC56933.2022.10053205.

3. Arya, A., Malhotra, H., Dayanand, Jeberson, W. (2017) 'Big data analytics in cyber security', *IJERT*, 5(10), pp. 1–3.

4. Andrade, R.O., Ontaneda, N., Silva, A., Tello-Oquendo, L. et al. (2020) 'Application of big data analytic in cybersecurity', *Proc. Int. Conf. ACC*, pp. 26–32.

5. Kotenko, I.V., Saenko, I.B., Branitskiy, A.A., Parashchuk, I.B., Gaifulina, D.A. (2021) 'Intelligent system of analytical processing of digital network content for protection against inappropriate information', *Informatics and Automation*, 20(4), pp. 755–788 (in Russ.). doi: 10.15622/ia.20.4.1.

6. Parashchuk, I., Doynikova, E., Saenko, I., Kotenko, I. (2020) 'Selection of countermeasures against harmful information based on the assessment of semantic content of information objects in the conditions of uncertainty', *Proc. Int. Conf. INISTA*, pp. 1–7. doi: 10.1109/INISTA49547.2020.9194680.

7. Kotenko, I.V., Saenko, I.B., Parashchuk, I.B., Doynikova, E.V. (2022) 'An approach for selecting countermeasures against harmful information based on uncertainty management', *ComSIS*, 19(1), pp. 415–433. doi: 10.2298/CSIS210211057K.

8. Tf, M.R., Singh, Y. (2022) 'An exploration on big data analysis and data mining methods', *Proc. INCOFT*, pp. 1–6. doi: 10.1109/INCOFT55651.2022.10094454

9. Kamara, M.K. (2020) *Securing Critical Infrastructures*. Bloomington: Xlibris US, 224 p.

10. Samanis, E., Gardiner, J., Rashid, A. (2022) 'Adaptive cyber security for critical infrastructure', *Proc. ICCPS*, pp. 304–305. doi: 10.1109/ICCPS54341.2022.00043.

11. Ekpo, U. (2018) *Introduction to Cyber Security: Fundamentals*. NY: Independently published, 37 p.

12. Srivastava, N., Jaiswal, U.C. (2019) 'Big data analytics technique in cyber security: A review', *Proc. ICCMC*, pp. 579–585. doi: 10.1109/ICCMC.2019.8819634.

13. Bothos, M.A., Thanos, K.G., Kyriazanos, D.M. et al. (2018) 'Correlation and dependence analysis on cyberthreat alerts', *ITU J.: ICT Discoveries*, 1(2), pp. 1–6.

14. Zhang, K., Zhao, F., Luo, S., Xin, Y., Zhu, H. (2019) 'An intrusion action-based IDS alert correlation analysis and prediction framework', *IEEE Access*, 7, pp. 150540–150551. doi: 10.1109/ACCESS.2019.2946261.

15. Zhu, G., Zeng, Y., Guo, M. (2017) 'A security analysis method for supercomputing users' behavior', *Proc. Int. Conf. CSCloud*, pp. 287–293. doi: 10.1109/CSCloud.2017.19.

16. Baranov, A.V., Korepanov, P.M., Kuznetsov, E.E. (2023) 'Information security of a supercomputer center', *Software & Systems*, 36(4), pp. 615–631 (in Russ.). doi: 10.15827/0236-235X.144.615.

17. Yang, B., Yu, Y., Wang, Z., Li, Sh. et al. (2021) 'Research on network security protection of application-oriented supercomputing center based on multi-level defense and moderate principle', *JPCS*, 1828, art. 012114. doi: 10.1088/1742-6596/1828/1/012114.

18. Ageeva, A.F. (2023) 'The role of supercomputers in matters of national security', *Bull. of the Academy*, (1), pp. 49–62 (in Russ.). doi: 10.51409/v.a.2023.03.01.005.

19. Yalcin, H., Daim, T., Moughari, M.M., Mermoud, A. (2024) 'Supercomputers and quantum computing on the axis of cyber security', *Tech. in Society*, (77), art. 102556. doi: 10.1016/j.techsoc.2024.102556.

20. Nesmiyanova, I.O. (2020) 'Information technologies: Stages of development, concept and classification,' *Izvestiya TulGU. Economic and Legal Sci*., (1), pp. 149–155 (in Russ.).

21. Kotenko, I.V., Saenko, I.B., Zakharchenko, R.I., Velichko, D.V. (2023) 'Subsystem for prevention of computer attacks against objects of critical information infrastructure: analysis of functioning and implementation', *Cybersecurity Issues*, (1), pp. 13–27 (in Russ.). doi: 10.21681/2311-3456-2023-1-13-27.

22. Adadurov, S.E., Glukhov, A.P., Kotenko, I.V., Saenko, I.B. (2022) 'Intelligent information security services', *Automation, communications, informatics*, (3), pp. 27–30 (in Russ.).

Авторы

Котенко Игорь Витальевич [1], д.т.н., профессор, главный научный сотрудник, руководитель лаборатории, ivkote@comsec.spb.ru
Саенко Игорь Борисович [1], д.т.н., профессор, главный научный сотрудник, ibsaen@comsec.spb.ru
Паращук Игорь Борисович [1], д.т.н., профессор, ведущий научный сотрудник, parashchuk@comsec.spb.ru
Десницкий Василий Алексеевич [1], к.т.н., доцент, старший научный сотрудник, desnitsky@comsec.spb.ru
Виткова Лидия Андреевна [1], к.т.н., старший научный сотрудник, vitkova@comsec.spb.ru

[1] Санкт-Петербургский федеральный исследовательский центр РАН, г. Санкт-Петербург, 199178, Россия

Authors

Igor V. Kotenko [1], Dr.Sci. (Engineering), Professor, Chief Researcher, Head of Laboratory ivkote@comsec.spb.ru
Igor B. Saenko [1], Dr.Sci. (Engineering), Professor,Chief Researcher, ibsaen@comsec.spb.ru
Igor B. Parashchuk [1], Dr.Sci. (Engineering), Professor, Leading Researcher, parashchuk@comsec.spb.ru
Vasily A. Desnitsky [1], Cand. of Sci. (Engineering), Senior Researcher, desnitsky@comsec.spb.ru
Lydia A. Vitkova [1], Cand. of Sci. (Engineering), Senior Researcher, vitkova@comsec.spb.ru

[1] St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, 199178, Russian Federation