УДК 351.86
DOI: 10.31673/2409-7292.2019.049297

І.В. Ткаченко, В.А. Козачок

СПОСІБ ОРГАНІЗАЦІЇ ОЦІНКИ СТАНУ КІБЕРЗАХИСТУ КРИТИЧНОЇ ІНФОРМАЦІЙНОЇ ІНФРАСТРУКТУРИ В РЕЖИМІ РЕАЛЬНОГО ЧАСУ З УРАХУВАННЯ ІНДИКАТОРІВ КІБЕРЗАГРОЗ

В статті розглянуто питання щодо організації комплексної оцінки стану кіберзахисту критичної інформаційної інфраструктури в режимі реального часу, з урахування індикаторів кіберзагроз. Оцінку стану кіберзахисту запропоновано здійснювати за рахунок використання різних типів даних, диференційованих джерел інформації, та програмних платформ, що здатні здійснювати обробку великих даних.

Ключові слова: інформаційна система, оцінки стану кіберзахисту критичної інформаційної інфраструктури, індикаторів кіберзагроз, великі дані.

Вступ

Комплексна оцінка стану кіберзахисту інформаційної інфраструктури будь якого масштабу є по своїй суті об’єднанням даних аудиту стану інфраструктури та даних щодо оцінки ризиків, що виникають в режимі реально часу. Для оцінки стану кіберзахисту ключовим моментом є вибір типів даних та джерел інформації, на основі яких вона буде здійснюватися, а також вибір ІТ рішення, функціонал якого дозволить проводити оцінювання стану кіберзахисту в режимі реального часу.

В ході аналізу типів даних та джерел інформації слід вказати, що всі можливі дані розподіляються на структуровані та не структуровані. На основі аналізу структурованих даних будується статистика, яка у певному вигляді відображає користувачу перелік наступних даних:

висновки аудиту інформаційної безпеки в установах, організаціях, де використовуються об'єкти критичної інформаційної інфраструктури, та виявлені порушення затверджених законодавством вимог захисту інформації;

висновки аудит стану спроможності кіберзахисту ІТ інфраструктури об'єкту критичної інформаційної інфраструктури, щодо виявлення, попередження інцидентів кібербезпеки, відповідніо стандартам (ISO 2700x, NIST, AICPA, HITRUST, COBIT, PCI DSS, GDPR, SOX, SWIFT, HIPPA, NYDFS);

інформація про стан пов’язаних інформаційними потоками з об'єктом критичної інформаційної інфраструктури організацій, установ, інших держав та стан організації кіберзахисту інформаційного обміну між ними, а також розрахункові показники відновлення його функціонування;

загальні тренди інформаційних загроз в кіберпросторі притаманні регіону розміщення об'єктів критичної інформаційної інфраструктури;

інформації державних, комерційних та не комерційних організацій про відомі способи, засоби та техніки виконання кібератак, аналізи виявлених кіберзлочинів, причин та наслідків їх виникнення;

оцінки ризиків виникнення кіберінцидентів для певного регіону та критичної інформаційної інфраструктури в цілому;

інформація щодо загальновідомих індикаторів кіберзагроз;

аналітичні звіти, інформаційні матеріали, описи алгоритмів кібератак, що були отримані оперативним шляхом;

описи вразливостей програмних засобів, зразки шкідливого програмного коду, пошкоджених системних журналів, лог файлів та повідомлень систем моніторингу, що були отримані в ході кібератак;

якісні показники та опис прийнятих рішень операційними центрами безпеки, щодо локалізації кібератак при тій чи іншій ступені їх розвитку;

загальнодержавні дані про населення, промисловий та економічний стан регіону(держави), на які впливає об'єкт критичної інформаційної інфраструктури.

Дану інформацію ми маємо можливість отримувати від державних, комерційних та не комерційних організацій.

Також, існує необхідність аналізу не структурованих даних та отримання корисної інформації, в тому числі такої, що надається в режимі реального часу, а саме:

оперативні дані щодо триваючих інцидентів інформаційної безпеки в регіоні, та щодо підвищення активності осіб (груп), що причетні до організації кіберінцидентів;

інформація, що надходить до держаного CERT про кіберінциденти в IT інфраструктурі державних органів, підрядних (аутсорсингових організацій), установ, підприємств, а також дані щодо переліку заходів, які вживаються (плануються до вживання) операційними центрами безпеки;

інформація про виявлені індикатори кіберзагроз в мережах об'єктів критичної інфраструктури, що підлягають аналізу (лог файли, системні журнали, повідомлення про аномальні активності та аномальну поведінку користувачів);

дані повідомлень систем моніторингу стану IT інфраструктури та повідомлень системи управління інформаційної безпеки та антивірусних систем;

інформація про аналіз стану мережевого трафіку та стану відмовостійкості IT інфраструктури, у піки активності та пасивності користувачів в мережах об'єктів критичної інфраструктури;

дані оцінки соціального настрою персоналу, що обслуговує об'єкти критичної інфраструктури;

інформація про зміни соціального настрою регіону, можливість виникнення локальних конфліктів, терористичних актів.

Джерелом таких даних можуть бути технічні складові: системи моніторингу IT - інфраструктури, системи антивірусного захисту, DLP системи, системи контролю доступу, лог файли активного мережевого обладнання, а також загальні джерела: соціальні мережі, новини, фото знімки, відео та аудіо файли, дані з будь яких відкритих джерел, дані розвідувальних, контррозвідувальних, військових організацій та правоохоронних органів, архів інтернету [1] та інші.

Обґрунтування вибору програмної платформи

Враховуючи різноманітну кількість платформ, які надають можливість обробки об'ємного масиву даних (dataset), в тому числі в режимі реального часу, пропонуємо використати стек Apache сумісних технологій та програмних продуктів (рис. 1), а саме:

HDFS (Hadoop Distributed File System) - розподілена файлова система, для зберігання файлів великих розмірів, поділених на блоки, що розподілені між вузлами обчислювального кластера. Всі блоки в HDFS мають однаковий розмір і кожен блок може бути розміщений на декількох вузлах, розмір блоку та коефіцієнт реплікації визначаються в налаштуваннях на рівні файлу. Завдяки реплікації забезпечується стійкість розподіленої системи до відмов окремих вузлів [2]

Для досягнення максимальної продуктивності обчислень в Hadoop використовуються технологія паралельної обробки потоків даних MapReduce, а також YARN модуль, що відповідає за управління ресурсами кластерів і планування завдань [3]

Наступним інструментом, що пропонується використати є Apache Spark - це фреймворк з відкритим вихідним кодом для паралельної обробки і аналізу слабкоструктурованих даних в оперативній пам'яті, що включає кілька бібліотек, які допомагають створювати додатки для машинного навчання (MLlib), обробки потоків (Spark Streaming) і графів (GraphX) [4] Також можливо використовувати Apache Solr - платформу, заснована на Apache Lucene. Основними можливостями якого є повнотекстовий пошук, індексація в режимі реального часу, динамічна кластеризація, інтеграція баз даних, обробка багатих документів (таких як Word,

PDF) та має можливість індексувати геолокацію за допомогою технологій обмеження вікон та геопросторових пошуків для даного регіону [5].

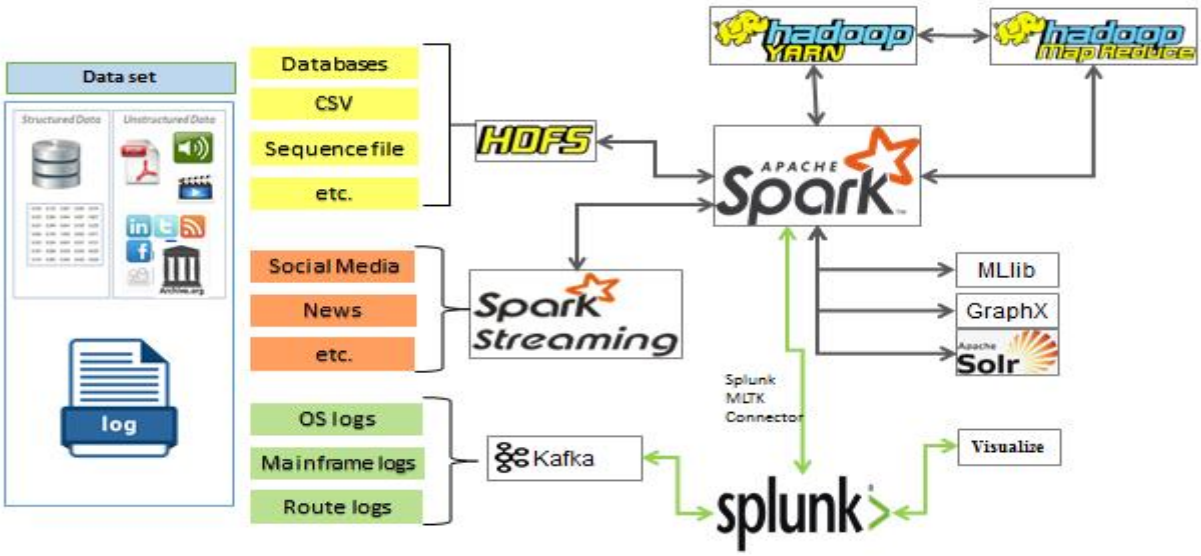

Рис. 1. Стек технологій Apache

Компоненти Hadoop поряд з Apache Spark можна використовувати в такий спосіб:

Apache Spark може працювати над HDFS, щоб використовувати розподілене тиражне сховище;

Apache Spark можна використовувати разом з MapReduce у тому ж кластері Hadoop або окремо як рамку обробки;

Пропонуємо MapReduce і Apache Spark використовувати разом, де MapReduce використовується для пакетної обробки, а Apache Spark - для обробки в режимі реального часу, що дозволить використовувати його, як у типових сценаріях оброблення даних, так в реалізації специфічних методів.

Один з найбільш успішних програмних продуктів для аналізу стану лог файлів в режимі реального часу на сьогоднішній день є Splunk, який ми пропонуємо застосувати у нашому рішенні [6].

Splunk – це система аналізу операційної діяльності в області інформаційних технологій, що збирає і аналізує великі обсяги машинних даних з усіх фізичних, віртуальних і хмарних середовищ IT-інфраструктури організації. Зібрані дані індексуються, в час доступу до даних, записаним раніше без моделювання, система перетворює машинні дані в формат «ключ - значення», після цього дані стають доступними для пошуку та аналізу через веб-інтерфейс. У продукті не використовується будь-яка зумовлена схема обробки даних, і орієнтована на роботу довільними форматами з системних журналів. Система дозволяє здійснювати пошук. як за даними в реальному часі, так за архівними даними, та на основі результатів пошуку дає можливість: аналізувати отримані результати за допомогою засобів візуалізації (використовується бібліотека D3.js), формувати звіти і попередження, створювати систему моніторингу та повідомлень в реальному часі.

Для формування черги лог файлів від IT інфраструктури до Splunk як правило використовується Apache Kafka - розподілений програмний брокер повідомлень, проект з відкритим вихідним кодом, що розробляється в рамках фонду Apache [7] Apache Kafka є розподілена, горизонтально масштабована система, що забезпечує нарощування пропускної здатності як при зростанні числа і навантаження з боку джерел, які можуть бути об'єднані в групи. Підтримується можливість тимчасового зберігання даних для подальшої пакетної

обробки. Однією з особливостей реалізації інструменту є застосування техніки, схожої з журналами транзакцій, використовуваними в системах управління базами даних.

Взаємодіє Splunk з Apache Spark досягається за допомогою використання коннектора Splunk Machine Learning Toolkit Connector for Apache Spark [8] Об'єднання даних з цих платформ дозволить створити систему підтримки прийняття рішень в режимі реального часу для забезпечення проактивного кіберзахисту.

Опис алгоритмів, що пропонується до застосування

Для обробки та аналізу даних за допомогою принципів машинного навчання, в залежності від типів даних, пропонуємо використати різні рішення (рис 2) наприклад:

NLP (Natural Language Processing) системи для обробки інформації з текстових файлів, новин та соціальних мереж– word2vec, Apache OpenNLP, GloVe. Крім цих моделей, знайшли застосування багато недавно розроблених технологій: FastText, Poincare Embeddings, sense2vec, Skip-Thought, Adaptive Skip-Gram;

Exif-аналіз для обробки фото та відео даних за допомогою програмних продуктів Jeffrey's Exif Viewer, FindEXIF.com, Panoramio, та інші;

Splunk indexer алгоритм пошуку порушень системи IT безпеки, запобігання кібератак, отримання інформації для бізнес-аналітики, оптимізації робочого процесу та збільшення продуктивності роботи з різноманітними великими масивами лог файлів [9].

Рис. 2. Рішення реалізації машинного навчання

Слід також зазначити, що обробка текстових файлів за допомогою word2vec базується на дистрибутивній семантиці та векторному представлені слів. Спочатку створюється словник, а потім обчислюється векторне подання слів, яке ґрунтується на контекстній близькості слова. Слова, що зустрічаються в тексті поруч та мають схожий зміст будуть мати близькі координати векторів-слів. Ґрунтуючись на контекстній близькості слів, технологія здійснює свої передбачення. Отримані вектори-слова можуть бути використані для обробки природної мови та машинного навчання. Дана технологія використовується для передбачення в Google [10]. У випадку оцінки стану кібербезпеки використання даної технології можливе для аналізу звітів, рапортів, новин та іншої текстової інформації, а також з метою пошук асоціацій, що надходить у запиті та видачі шарів (layer) з рівнем точності для обчислення критичних зон, що знаходиться у ньому. Виходячи з вищезазначеного dataset буде піддано детальному аналізу для визначення залежностей між подіями.

В ході застосування рішення (рис. 3) структуровані дані бути перевірені на достовірність, для того щоб відкинути пропаганду, фейкові повідомлення, (для аналізу новин

і соціальних мереж пропоную використати ресурс аналогічний Fakebox) [11]. Неструктуровані дані підлягають попередній обробці за допомогою описаних вище NLP систем для оцінки факторів пов'язаних загрозами, що несуть в собі кіберінциденти.

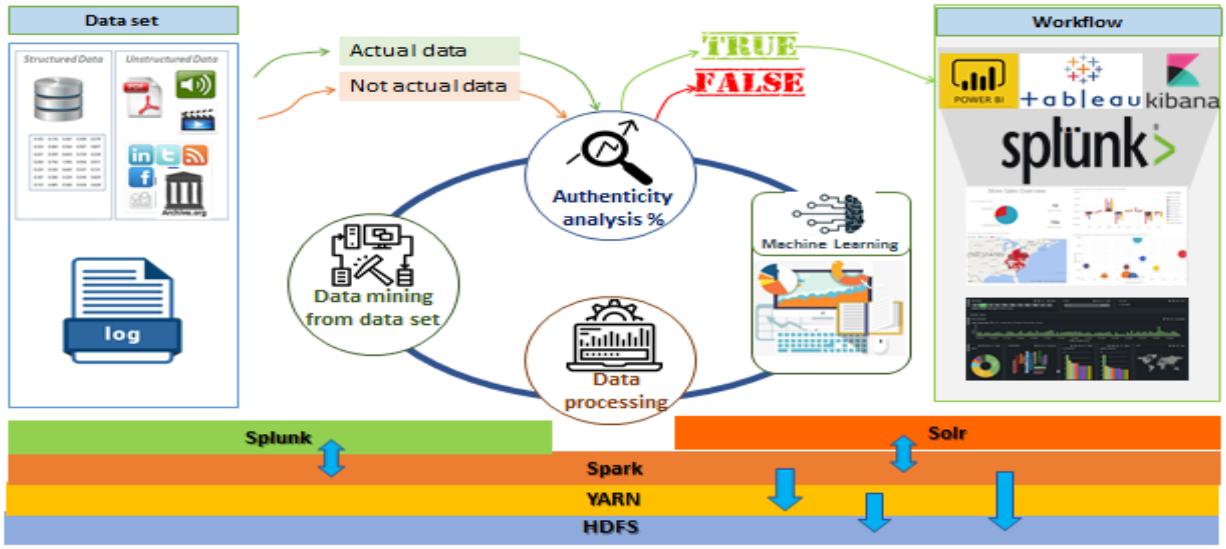

Рис. 3. Рішення для обробки текстових файлів

Після обробки вся інформація, яка може знадобитись для прийняття рішень візуалізується кінцевому користувачу використовуючи сервіси такі як: PowerBI, Tableau, Lucidworks Banana.

Висновки

Якщо на рівні прийняття рішень щодо кібероборони держави, кіберзахисту державних органів, комерційних та не комерційних установ, що експлуатують критичну інформаційну інфраструктуру буде застосовуватись інформаційна система, подібна за стеком технологій до запропонованої це дозволить:

моделювати та здійснювати оцінку економічних збитків, матеріальних та не матеріальних втрат за результатами чи в процесі кібератаки;

прогнозувати кількісні показники втрат даних при втраті контролю за складовими тієї чи іншої критичної інформаційної інфраструктури;

в єдиному інформаційному вікні, одержувати дані про кібератаку чи кіберінцидентів, з усіх можливих джерел на поточний момент, та здійснювати миттєвий порівняльного аналізу, класифікації, та віднесення її до вже відомого чи унікального виду;

контролювати рівень загрози кібератаки, та вживати проактивних заходів до початку незворотніх змін в складових критичної інформаційної інфраструктури;

надавати актуальні та достовірні рекомендації центрам операційної безпеки в режимі реального часу, з доступом до даних як в режимі онлайн так і офлайн.

Перелік посилань

1. https://uk.wikipedia.org/wiki/%D0%86%D0%BD%D1%82%D0%B5%D1%80%D0%BD%D0%B5%D1%82-%D0%B0%D1%80%D1%85%D1%96%D0%B2
2. https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
3. https://ru.bmstu.wiki/YARN_(Yet_Another_Resource_Negotiator)

4. Холден Карау, Рейчел Уоррен Эффективный Spark. Масштабирование и оптимизация. – Питер, 2018. – 352
5. https://ru.bmstu.wiki/Apache_Solr
6. https://www.splunk.com/en_us/blog/security/six-straight-years-splunk-named-a-leader-in-the-gartner-siem-magic-quadrant.html
7. Niya Narhid, Gwen Shapira, Todd Palino. Apache Kafka. Stream processing and data analysis. - . St Peterb. 2018. – 463
8. https://static.rainfocus.com/splunk/splunkconf18/sess/1523315089950001nFR9/finalPDF/Using-Spark-and-MLlib-1364_1538792259889001CncS.pdf
9. https://docs.splunk.com/Documentation/Splunk/8.0.1/Indexer/Basicclusterarchitecture
10. https://en.wikipedia.org/wiki/Word2vec
11. https://medium.com/@machineboxio/introducing-fakebox-detect-fake-news-with-machine-learning-f602c39aad04

Надійшла: 04.11.2019 Рецензент: д.т.н., доцент Гайдур Г.І.