УДК 004.056

Н.О. Маслова, канд. техн. наук, доц.,
М.А. Федорко, магістрант кафедри ПМІ,
Донецький національний технічний університет, м. Покровськ, Україна
nataliia.maslova@donntu.edu.ua

Особливості захисту даних великих обсягів

Досліджено питання забезпечення безпеки Big Data, зроблено огляд методів захисту; розкрито комплексність проблеми, необхідність багаторівневого захисту й застосування сучасних методичних, теоретичних та програмних розробок з перших кроків роботи з великими даними. З огляду на складність задачі обробки й захисту великих даних, зроблено пропозиції з побудови на перших етапах проектування сховищ для Big Data власних хмар; проаналізована можливість застосування вітчизняного криптографічного шифру «Калина».

Ключові слова: великі дані, інформаційна безпека, розміщення на хмарі, шифрування, захист.

DOI: 10.31474/1996-1588-2018-1-26-41-47

Вступ

Однією з особливостей сучасного етапу обробки інформації є безперервне збільшення обсягів оброблюваних та накопичуваних даних, ускладнення методів їх аналізу, візуалізації та захисту. Незважаючи на значну кількість наукових праць щодо рішення означених питань, проблема забезпечення безпеки Big Data є невирішеною та найбільш актуальною. Ситуація погіршується тим, що стандарти з захисту великих даних на даний час відсутні, а існуючі антивірусні системи не призначені для забезпечення необхідного рівня безпеки таких даних [1]. Крім того, на думку авторів, й самі дослідження в означеному напрямку недостатньо систематизовані, тому питання захисту даних великих обсягів потребує подальшого розвитку й дослідження та розробки у цьому напрямку є актуальними.

Опис предметної області

Термін «великі дані» (Big data) почали широко застосовувати з кінця 2000-х років. Популярність цієї теми й досі досить висока (рис.1).

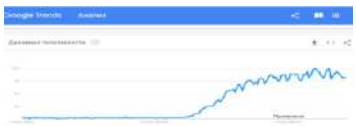

Рисунок 1 - Популярність тематики Big Data

Термін Big Date пов'язують з структурованими і неструктурованими даними великих обсягів та значного різноманіття форматів. При цьому їх формування, накопичення та обробка, як правило, виконуються в on-line режимі, завдяки роботі розподілених систем, застосуванню хмарних сервісів [2].

Приклади того, що може бути джерелом даних великих обсягів:

- GPS-сигнали для транспортної компанії;
- дані з датчиків промислового підприємства;
- цифровані книги в електронній бібліотеці;
- транзакції клієнтів банку;
- інформація про товари або покупки великої торговельної мережі і т.д.

Головна задача методів Big Data – обробка величезних обсягів даних й побудова на їх основі прогнозних моделей, виявлення прихованих зав'язків та взаємодій. Це актуальна задача, якій приділяють увагу, наприклад, такі компанії, як Informatica (ETL-технології розробки сховищ та керування даними); Hewlett Packard Enterprise; Imperva (розробка і виробництво продуктів для захисту систем управління базами даних та web-додатків); Apache Software Foundation (Apache - технології розробки програмного забезпечення з відкритим кодом).

Проблеми захисту Big data виникають на всіх етапах роботи з даними – при формуванні, передачі, накопиченні, зберіганні, аналізі та візуалізації.

Відомі на ринку IT компанії приділяють значну увагу розробці спеціалізованих рішень з забезпечення безпеки Big data, найбільш відомі з них наведено в таблиці 1.

Найбільший внесок у розробку технологій захисту Big Data зроблено міжнародним альянсом Cloud Security Alliance (CSA), Національним інститутом стандартів і технологій США (NIST) та Агентством Європейського Союзу з питань ме-

режевої та інформаційної безпеки (European Union Agency for Network and Information Security, ENISA).

Таблиця 1. Спеціалізовані рішення з захисту BD

Розробник	Рішення
IBM	Top tips for Big Data Security
Oracle	Enterprise Security for Big Data Environments
Forrester	Big Data Security Strategies For Hadoop Enterprise Data Lakes
Cloudera	Cloudera Security
Securosis	Securing Hadoop: Security Recommendations for Hadoop Environment
CSA	The Big Data Security and Privacy Handbook
NIST	Big Data Interoperability Framework
ENISA	Big Data Threat Landscape and Good Practice Guide

Одним з істотних обмежень проектів в області великих даних є ризики інформаційної безпеки, а особливо - порушення конфіденційності та цілісності великих даних. Дані, що використовуються для аналізу часто містять персональну або актуальну комерційну інформацію. А питання забезпечення цілісності можуть стосуватися як даних, які аналізуються, так і отриманих при їх обробці результатів.

Визначення загроз великих даних та рекомендації щодо їх запобігання є однією з пріоритетних задач Агенції Європейського Союзу з питань мережевої та інформаційної безпеки. Документ, розроблений ENISA - «Big Data Threat Landscape and Good Practice Guide» [3]. Автори наголошують, що робота з великими даними пов'язана з виникненням нових, характерних для Big Data ризиків інформаційної безпеки, які вимагають особливих підходів та процедур захисту.

В якості другої важливої розробки слід назвати огляд Міжнародного альянсу Cloud Security Alliance (CSA) «The Big Data Security and Privacy Handbook: 100 Best Practices in Big Data Security and Privacy». У огляді, що опубліковано у 2016 році, викладено 100 рекомендацій в області забезпечення безпеки великих даних [4].

Документ містить докладний опис кращих практик, що апробовані на ринку Big Data. В перелік включено як типові заходи кібер безпеки (наприклад, аутентифікацію і контроль доступу), так і сучасні криптографічні технології. В документі кожен пункт має два розділи, в яких вказується, чому ці заходи безпеки необхідні і як вони можуть бути реалізовані.

Основні розділи документу - етапи побудови системи захисту великих даних наведено в Таблиці 2.

Таблиця 2 Етапи захисту Big Data за версією CSA

	Етапи захисту Big Data
1	Захист обчислень в розподілених програмних системах
2	Захист нереляційних баз даних
3	Захист сховищ даних
4	Фільтрація і валідація даних
5	Моніторинг безпеки у режимі on-line
6	Забезпечення конфіденційності
7	Криптографічний захист
8	Гранульований контроль доступу
9	Аудит
10	Data provenance (походження даних)

Третій значний дослідник у сфері захисту великих даних - національний інститут стандартів і технологій США (NIST) запропонував специфікацію Interoperability Framework V1.0 [5], яка включає документи з описом всіх аспектів роботи з великими даними:

- Big Data Definitions;
- Big Data Taxonomies;
- Big Data Use Cases and Requirements;
- Big Data Security and Privacy;
- Big Data Architecture White Paper Survey;
- Big Data Reference Architecture;
- Big Data Standards Roadmap.

В документі вводиться концептуальна модель архітектури великих даних (NIST Big Data Reference Architecture, NBDRA), що являє собою систему великих даних. Система включає п'ять логічних функціональних компонентів. Об'єднуючою частиною є Security and Privacy Managament, у якому найбільша увага приділяється питанням ідентичності, авторизації, аудиту, безпеки мережевих пристроїв.

В найбільш узагальненому вигляді спеціалісти рекомендують зосередитися на чотирьох напрямках захисту великих даних:

- безпеці інфраструктури;
- забезпеченню конфіденційності даних;
- процедурах керування даними;
- безперервним моніторингом безпеки.

Основними інструментами та технологіями роботи з великими даними є бібліотеки масово-паралельної обробки невизначено структурованих даних, системи управління базами даних категорії SQL, у тому числі NoSQL, алгоритми MapReduce, проекти Hadoop та MPI.

Ці засоби придатні для обробки великих даних, але для забезпечення безпеки NIST рекомендує додавати декілька рівнів захисту. Один - для захисту програм, другий - для захисту даних. При цьому можливо застосування, наприклад, спеціального протоколу Kerberos, який контролює доступ до ресурсів Hadoop або інші програмні продукти, які реалізують функціональність

рольового доступу, наприклад, Apache Accumulo, Sentry і ін.

На ринку IT-технологій є досить повні рішення по розгортанню, супроводу або адмініструванню великих сховищ та апаратно-програмних рішень до них. Особливістю найкращих є наявність вбудованих систем та засобів безпеки, технологічні рішення та гармонічне поєднання застосованих програмних продуктів. На українському ринку послуги пропонує Корпорація EMC й її філіал Dell EMC. На сайті компанії - значний перелік продуктів [6], що забезпечують інтеграцію, зберігання, адміністрування та захист даних, дозволять розгортати додатки або розробляти власні програми для аналітики великих даних. Це, зокрема, сховище Isillon, платформа для розгортання хмари ECS з підтримкою Hadoop, високопродуктивні сервери, інтеграційні рішення. В якості засобів безпеки пропонуються Avamar 4.1, рішення Cisco, InfoMover, VNX. На сайті є й пропозиція «Спробувати», що досить доречно на перших етапах роботи з великими даними.

Також слід назвати рішення фірми IBM, що складається з трьох продуктів[7]:

• спеціалізованого програмно-апаратного комплексу для побудови аналітичних додатків і сховищ даних Netezza;

• nfoSphere BigInsights - рішення з аналізу та обробки неструктурованих даних на основі технологій Hadoop;

• засобу аналізу потокової інформації та комплексної обробки великих обсягів неструктурованих даних - InfoSphere Streams і Vivisimo.

Пропозиції компанії Hitachi Data Systems [8] – два спеціалізованих програмно-апаратних комплекси:

• платформа для зберігання і управління великими обсягами неструктурованих даних (Hitachi Content Platform, HCP);

• рішення для забезпечення файлового доступу до даних, з збереженням і керуванням великою кількістю файлів - Hitachi Network Attached Storage (HNAS).

Для захисту сховищ даних автори проектів пропонують застосування технік створення репозиторію SUNDR, розробку дайджестів завірених повідомлень; ротацію ключів; або створення власного хмарного сховища.

У цій роботі зупинимось на двох аспектах - дослідженні можливості створення власного сховища та виборі алгоритму криптографічного захисту.

Застосування власного сховища

Ідея створення власного сховища досить популярна в сьогоденні.

Сучасні технології дозволяють розгорнути власне хмарне сховище, та розмістити його на площі підприємства, на віртуальному хостингу, VPS чи видаленому сервісі. Інформація може поступати з офісних ПК, Інтернету, промислових датчиків чи з мобільних пристроїв.

Існує декілька підходів до розгортання власного хмарного сховища:

– використати готову комерційну платформу для розгортання свого хмарного рішення;

– знайти вихідний код реалізованого хмарного сховища та розгорнути його використовуючи свої апаратні потужності;

– розробити свою платформу для збереження файлів не використовуючи готові рішення.

Кожний підхід має свої плюси та мінуси.

Перший підхід заснований частіше на використанні рішень, які розроблюються комерційними компаніями. Ці рішення є по-більшості комплексними, надають багато-платформу підтримку хмари. Але в сфері безпеки спостерігається залежність від майстерності та рівня знань розробників компанії, а детальна інформація про застосовані захисні механізми надається не завжди.

Другий метод хоча і схожий на попередній з точки зору реалізації, вважається більш небезпечним для збереження даних, тому що в чужому коді складно провести ревізії на закладки, приховані функції й таке інше.

Третій метод є найбільш складним з точки зору проектування, розробки, та підтримки рішення. Він вимагає значних затрат на придбання апаратної частини, захищених каналів, аудит.

На початкових етапах роботи з великими даними компанії краще зосередитись на першому варіанті й досить швидко розгорнути свою хмарне рішення з відносно низькими зусиллями та достатнім функціоналом.

Існує велика кількість засобів для розгортання хмарних сховищ. До таких засобів відносяться Seafile, ownCloud, Pydio, BitTorrent Sync, Syncthing, HRCloud2, SparkleShare, Storage Made Easy, AeroFS, TeamDrive, arXshare, LimboMedia, EncBox, git-annex assistant, Tonido, Nextcloud, ownCloud, Cozy.

Серед цих засобів цікавими для розгляду є Bittorrent Sync, OwnCloud, Seafile, Cozy, SparkleShare, git-annex, Tonido.

З огляду на наявність досить великого переліку пропозицій від розробників, проведено спеціальне дослідження. Була складена таблиця основних вимог до засобу створення сховища з урахуванням необхідності розміщення великих даних на «хмарі» та їх безпеки, та виконано порівняння пропозицій.

Порівняння виконувалося з застосуванням чисельних метрик з ваговими коефіцієнтами. Загалом було проаналізовано близько 40 метрик, серед яких підтримувані платформи; наявність чи

відсутність обмежень на обсяг сховища; технологія побудови; схеми прийому-передачі даних; швидкісні можливості. У розділі безпеки основними метриками були: надійність та швидкодія алгоритмів захисту; можливість відпрацьовування в режимі on-line; наявність різних схем організація захисту - на стороні серверу, на стороні клієнту, при передачі; варіанти підключення алгоритмів безпеки за вибором замовника; наявність пропозицій захисту від розробника засобу розгортання хмари й т.д.

Результати порівняння можливостей інструментальних засобів створення власних сховищ на «хмарі» наведено у таблиці 3. У другому стовпчику - бальна оцінка в інтегрованому вигляді, виконана з застосуванням інформації з сайтів розробників [9-11].

Таблиця 3. Засоби створення «хмарних» сховищ

Програмний засіб	Бальна оцінка
Seafile	38
BitTorrent Synk	28
OwnCloud	17
Cozy	26
GitHub	36
Tonido	26
SparkleShare	20
Bitbucket	37

Тож за нашими оцінками перевагу має пропозиція Seafile. Також високі показники у Tonido, Bitbucket та GitHub. При цьому останні засоби не є строго інструментами для створення сховищ й вимагають спеціального застосування, але досить зручні у колективній роботі.

Означені експерименти можна застосовувати на початкових етапах проведення робіт з обробки великих даних з метою, наприклад, визначення обсягу й структури сховища, формулювання основних задач обробки та аналізу, оцінювання обсягів й складності вирішуваних задач, визначення ризиків та вимог до захисту даних, обрання технологічних рішень й оцінювання їх можливостей.

Обрання криптографічного алгоритму

Окремим питанням забезпечення безпеки й захисту великих даних є криптографія Big Data.

Окрім відомих та розповсюджених алгоритмів шифрування таких, як, наприклад, AES або RSA, на різних етапах захисту великих даних рекомендовано хешування паролів, наскрізне шифрування даних, пов'язане шифрування, шифрування на базі атрибутів (ABE), шифрування на базі ідентичності (IBE), конвергентне шифрування.

Приділяється увага безпеці на транспортному рівні (TLS), шифруванню на рівні захищених сокетів (SSL) й наявності надійних захисних механізмів на застосованих для зберігання, накопичування та обробки даних хмарах.

Особливостями криптографічних підходів до обробки Big Data є відокремлена обробка інформації та метаданих, організація пошуку за допомогою булевих запитів на зашифрованих даних, порівняння даних без їх розшифрування, виявлення дублікатних даних в великих масивах на основі ключів [4] й інші.

Можна зробити висновок, що дані повинні бути повністю зашифровані на кожному етапі обробки, зберігання або передачі. Але при цьому криптографічні процедури повинні бути швидкодіючими. Такими, що не завадять основній задачі – аналізу даних, а крім того, до них повинен бути організований ефективний та безперервний доступ.

Необхідність застосування швидких криптографічних алгоритмів при роботі з Big Data та вимога мінімізувати ризики, пов'язані з застосуванням неузгоджених один з одним алгоритмів або програмних продуктів з невідомими прихованими функціями, вимагає звернути увагу на вітчизняні розробки.

Так, одним з сучасних вітчизняних алгоритмів шифрування є шифр «Калина», в основі якого лежить національний криптографічний стандарт України ДСТУ 7624:2014 [12]. Стандарт визначає структуру шифру та режими його роботи. Шифр є прикладом блочного симетричного перетворення і підтримує розмір блоку і довжину ключа шифрування 128, 256 і 512 біт. Зараз це єдиний в світі стандарт блочного шифрування, що підтримує 512 - бітові симетричні ключі. При цьому довжина ключа повинна дорівнювати або в два рази перевищувати розмір блоку. Саме це забезпечує нормальний, високий і надвисокий рівень стійкості шифру.

Модель використовує Square-подібну SPN-структуру. Аналоги цих принципів застосовуються в алгоритмах AES / Rijndael, Whirlpool, Stribog, Grasshopper («Кузнечик») і ряд інших. Окрім блочного шифру, стандарт передбачає додаткові можливості, орієнтовані на сучасні системи криптографічного захисту, можливість ефективної реалізації на більшості сучасних програмних і програмно-апаратних платформ.

Орієнтація на сучасні архітектури пов'язана з узгодженням об'єму кешу L1 та розміру МДР-матриці, що за заявою авторів шифру суттєво підвищує швидкодію алгоритму. Відомо, [13], що для блокового шифру «Калина» розробниками була обрана МДВ-матриця розміром 64x64 біта (8x8 над полем $GF(2^8)$) як така, що забезпечує необхідні криптографічні властивості і

вимоги щодо швидкодії на сучасних програмних 64-бітових архітектурах.

Проведемо власне дослідження цього ствердження. Необхідність аналізу пов'язана з тим, що на рівні експерименту дослідження проводиться з застосуванням криптографічних алгоритмів, які включено в стандартні бібліотеки й результати залежать від особливостей реалізації, застосованої мови програмування, та технічних характеристик комп'ютерів.

Для експерименту обрано процесори:

- Intel Core i7-7700HQ CPU @ 2.8Gz 2.8 GHz. 2-4 ядра. Кеш 1-го рівня (L1) - 256 КБ, 2-го рівня (L2) - 1024 КБ, 3-го рівня (L3) - 6144Кб.
- Intel Core i5-4200U CPU @ 2.3GHz 2.4 GHz, 2-4 ядра. Кеш 1-го рівня (L1) - 128КБ, 2-го рівня (L2) - 512 КБ, 3-го рівня (L3) - 3072Кб
- Intel Pentium® CPU N3700. Тактова частота 1600-2400 МГц. Кеш 1-го рівня (L1) - 224КБ, 2-го рівня (L2) - 2048 КБ, 3-го рівня (L3) – немає.

Розмір оперативної пам'яті у всіх випадках дорівнював 6Гб, включено 2 ядра, без прискорення. Мова програмування C++, ОС Windows 10. Заміри виконувалися для режиму шифрування, для вхідних повідомлень (відкритих текстів) однакового обсягу

Результати, отримані експериментальним шляхом на різних пристроях наведено на рисунку 2. Послідовність графіків співпадає з переліком обраних процесорів (зверху вниз).

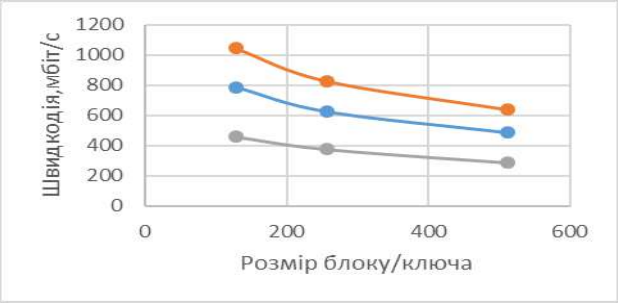

Рисунок 2 – Швидкодія алгоритму «Калина»

Таким чином, процесор з кешем першого рівня L1 у 256КБ показав найкращі результати на усіх трьох досліджених розмірах блоку (довжина ключа дорівнювала розміру блоку). Трохи нижча швидкодія виявилася для випадку, коли розмір кешу L1 дорівнював 128КБ. Третій випадок – кеш першого рівня 224КБ та відсутність кешу третього рівня L3. Кеш третього рівня зазвичай застосовується для взаємодії ядр процесора, та для тимчасового зберігання даних з низькою ймовірністю запиту. Але для цієї лінійки процесорів характеристики алгоритму «Калина» слід дослідити більш детально.

Крім того, дослідженнями підтверджено, що в цілому, час виконання алгоритму «Калина» на тих самих пристроях має той же рівень, що й результати стандартного AES – алгоритму з відповідною довжиною ключа. З оглядом на всі переваги вітчизняного алгоритму (нормальний, високий і надвисокий рівень стійкості, довжина блока і ключа 128, 256 і 512 бітів, односпрямована конструкція схеми розгортання ключів, циклове перетворення, стійкість до переборних атак та відомих методів аналізу), застосування шифру «Калина» для захисту великих даних на сучасних пристроях є перспективним.

Наприкінці наведемо ще деякі програмні й алгоритмічні підходи, що застосовуються для захисту файлів, що зберігаються в хмарних сховищах. Це:

• криптографічна файлова система EncFS;

• пропрієтарні програми, що позиціонуються як засіб для шифрування даних в хмарі, наприклад, Boxcryptor, Cloudfogger Truecrypt і ін.;

• локальне шифрування (наприклад, утиліта CryptSync);

• програми для резервного копіювання (наприклад, Duplicati);

• клієнт CarotDav.

Висновки

В роботі досліджено стан питання забезпечення безпеки великих даних.

Зроблено огляд відомих у цих сферах розробників методів захисту та рекомендацій щодо їх застосування.

Показана комплексність проблеми, необхідність багаторівневого захисту й застосування сучасних методичних, теоретичних та програмних розробок з перших кроків роботи з великими даними.

З огляду на складність задачі обробки й захисту великих даних, зроблено пропозиції з побудови на перших етапах проектування сховищ Big Data власних хмар, проаналізовані придатні до цього рішення.

Проаналізована можливість застосування вітчизняного криптографічного шифру «Калина» для шифрування даних для їх розміщення у сховищі.

Таким чином, зроблено ще один крок у напрямку дослідження можливостей забезпечення захисту технологій великих даних з застосуванням вітчизняних криптографічних стандартів.

Список літератури

1. Дмитрий Смирнов Защита Big Data: проблемы и решения https://www.it-weekly.ru/it-news/security/117831.html (дата звернення 03.08.2018).
2. J. L. Peñaloza-Figueroa and C. Vargas-Perez, "Big-data and the challenges for statistical inference and economics teaching and learning," *Multidiscip. J. Educ. Soc. Technol. Sci.*, vol. 4, no. 1, P. 64–87, 2017.
3. Big Data Threat Landscape and Good Practice Guide. URL: https://www.enisa.europa.eu/publications/bigdata-threat-landscape/at_download/fullReport (дата звернення 15.07.2018).
4. Big Data Security and Privacy Handbook: 100 Best Practices in Big Data Security and Privacy. Cloud Security Alliance https://downloads.cloudsecurityalliance.org/assets/research/big-data/BigData_Security_and_Privacy_Handbook.pdf (дата звернення 15.07.2018).
5. NIST Special Publication 1500-1. NIST Big Data Interoperability Framework. URL: https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf (дата звернення 15.07.2018).
6. Рішення для великих даних для трансформації вашого бізнесу. Інформація з сайту Dell EMC https://www.dellemc.com/uk-ua/big-data/solutions.htm#tab=tab-1 (дата звернення 05.08.2018).
7. https://www.ibm.com/analytics/netezza (дата звернення 03.08.2018).
8. https://www.hitachivantara.com/ru-ru/company/contact.html (дата звернення 03.08.2018)
9. https://admins.su/seafile-alternativa-dropbox-na-svoix-serverax-ustanovka-na-debian/ (дата звернення 03.08.2018).
10. https://git-scm.com/ (дата звернення 03.08.2018)
11. https://technet.microsoft.com/ru-ru/library/dn271884.aspx (дата звернення 03.08.2018)
12. Roman Oliynykov, Ivan Gorbenko, Oleksandr Kazymyrov, Victor Ruzhentsev, Oleksandr Kuznetsov, Yurii Gorbenko, Oleksandr Dyrda, Viktor Dolgov, Andrii Pushkaryov, Ruslan Mordvinov, Dmytro Kaidalov. DSTU 7624:2014. National Standard of Ukraine. Information technologies. Cryptographic Data Security. Symmetric block transformation algorithm. Ministry of Economical Development and Trade of Ukraine, 2015 (in Ukrainian)
13. Принципи побудови і основні властивості нового національного стандарту блокового шифрування України / Р. Олійников, І. Горбенко, О. Казимиров, В. Руженцев, Ю. Горбенко // Захист інформації, том 17, №2, квітень-червень 2015, С.142-157.

References

1. Dmitriy Smirnov Zashchita Big Data: problemy i resheniya https://www.it-weekly.ru/it-news/security/117831.html (data zvernennya 03.08.2018).
2. J. L. Peñaloza-Figueroa and C. Vargas-Perez, "Big-data and the challenges for statistical inference and economics teaching and learning," *Multidiscip. J. Educ. Soc. Technol. Sci.*, vol. 4, no. 1, P. 64–87, 2017.
3. Big Data Threat Landscape and Good Practice Guide. URL: https://www.enisa.europa.eu/publications/bigdata-threat-landscape/at_download/fullReport (data zvernennya 15.07.2018).
4. Big Data Security and Privacy Handbook: 100 Best Practices in Big Data Security and Privacy. Cloud Security Alliance https://downloads.cloudsecurityalliance.org/assets/research/big-data/BigData_Security_and_Privacy_Handbook.pdf (data zvernennya 15.07.2018).
5. NIST Special Publication 1500-1. NIST Big Data Interoperability Framework. URL: https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf (data zvernennya 15.07.2018).
6. Rishennya dlya velykykh danykh dlya transformatsiyi vashoho biznesu. Informatsiya z saytu Dell EMC https://www.dellemc.com/uk-ua/big-data/solutions.htm#tab=tab-1 (data zvernennya 05.08.2018).
7. https://www.ibm.com/analytics/netezza (data zvernennya 03.08.2018).
8. https://www.hitachivantara.com/ru-ru/company/contact.html (data zvernennya 03.08.2018)
9. https://admins.su/seafile-alternativa-dropbox-na-svoix-serverax-ustanovka-na-debian/ (data zvernennya 03.08.2018).
10. https://git-scm.com/ (data zvernennya 03.08.2018)
11. https://technet.microsoft.com/ru-ru/library/dn271884.aspx (data zvernennya 03.08.2018)
12. Roman Oliynykov, Ivan Gorbenko, Aleksandr Kazymyrov, Viktor Ruzhentsev, Aleksandr Kuznetsov, Yuriy Gorbenko, Aleksandr Dirda, Viktor Dolgov, Andrey Pushkarev, Ruslan Mordvinov, Dmitriy Kaydalov. DSTU 7624: 2014. Natsional'nyy standart Ukrainy. Informatsionnyye tekhnologii. Bezopasnost' kriptograficheskikh dannykh. Algoritm simmetrichnoy blochnoy transformatsii. Ministerstvo ekonomicheskogo razvitiya i torgovli Ukrainy, 2015 g. (na ukrainskom yazyke)

13. Pryntsypy pobudovy i osnovni vlastyvosti novoho natsional′noho standartu blokovoho shyfruvannya Ukrayiny / R. Oliynykov, I. Horbenko, O. Kazymyrov, V. Ruzhentsev, YU. Horbenko // Zakhyst informatsiyi,, tom 17, №2, kvíten'-cherven' 2015, S.142-157

Надійшла до редколегії 28.08.2018

N. MASLOVA, M. FEDORKO
Donetsk National Technical University, Pokrovsk, Ukraine

FEATURES OF BIG DATA PROTECTION

This article presents the Big Data Protection. The review of existing methods of data protection is conducted. The companies working in this field are named, the main results of their activity are listed. The complexity of the problem, the need for multilevel protection and the application of modern methodological, theoretical and software developments in the all stages of working with large data are shown.

Two important points were studied. The first is to creating a private cloud for a Big Data repository and ensure its protection. A few suggestions are discussed. It is shown that the SeaSize solution has the highest score and can be recommended at the first stages of Big Data storage development. The second issue considered in the article is the possibility and effectiveness encryption of Big Data with cryptographic code of the *Kalyna* cipher. The prospects of this direction for modern computing platforms are shown.

Keywords: Big Data, information security, cloud placement, encryption, protection.

Н.А. МАСЛОВА, М.А. ФЕДОРКО
Донецкий национальный технический университет, г. Покровск, Украина

ОСОБЕННОСТИ ЗАЩИТЫ ДАННЫХ БОЛЬШИХ ОБЪЕМОВ

Исследованы проблемы обеспечения безопасности Big Data, сделан обзор методов защиты; показана комплексность проблемы, необходимость многоуровневой защиты и применения современных методических, теоретических и программных разработок с первых шагов работы с большими данными. Учитывая сложность задачи обработки и защиты больших данных, сделаны предложения по построению на первых этапах проектирования хранилищ Big Data собственных облаков; проанализирована возможность применения отечественного криптографического шифра «Калина».

Ключевые слова: большие данные, информационная безопасность, размещение на облаке, шифрование, защита.