УДК 004.75

Лабжинський В. А.
Національний технічний університет України "Київський політехнічний інститут ім. Ігоря Сікорського"

СУЧАСНІ МЕХАНІЗМИ ПОБУДОВИ ХМАРНИХ ТЕХНОЛОГІЙ ЯК ПРІОРИТЕТНИЙ НАПРЯМОК ВДОСКОНАЛЕННЯ РОЗПОДІЛЕНИХ СИСТЕМ ОБРОБКИ ДАНИХ

Визначено основні види хмарних технологій, наводяться переваги та недоліки кожного виду. У якості основи дослідження прийнято архітектуру гібридної хмари. Зазначено, що проблема поділу даних в параметрі гібридної хмари полягає у мінімізації вартості виконання запиту навантаження і обмежена двома окремими обмеженнями. Здійснюється опис обмежень, та пропонується розподіл метрик продуктивності – вартість обробки запиту (тобто продуктивність), витрати та ризики.

***Ключові слова*:** гібридна хмара, репліка, продуктивність, витрата, ризик, розподілена система обробки даних, навантаження

Labzhynskyi V. A. Modern mechanisms of cloud technologies construction as a priority to improve distributed data processing system. The modern mechanisms of development of cloud technologies as a priority direction of improvement of distributed data processing systems are considered. There are three main types of cloud computing. The advantages and disadvantages of each type are determined. Architecture of hybrid cloud are adopted as the basis of the studying. The basics of load distribution describes in article. It is noted that the problem of the separation of the data in the parameter hybrid cloud is to minimize the cost of query execution load and are limited with two separate restrictions. Description of the restrictions are considered, and a distribution of performance metrics – the cost of query processing (i.e., performance), costs and risks are proposed.

Keywords: hybrid cloud, replica, performance, consumption, risk, distributed data processing system workload

Вступ та постановка проблеми. Питання впровадження комп'ютерних обчислень датується 1960 роком, автором даного напряму є Джон Маккарті [1]. На сьогоднішній день, масштабність використання хмарних технологій значно розширюються. Сучасні системи і техніки розгортання дозволяють економити як на обслуговуванні та персоналі, так і на інфраструктурі. Апаратне забезпечення може бути сильно спрощено при обробці даних та зберіганні інформації у віддалених центрах даних. Всі ці проблеми майже повністю перекладаються на провайдера послуг [2].

Основним недоліком хмарних технологій, по праву можна вважати повну залежність від постачальника послуг. Фактично підприємство (користувач) являється заручником провайдера сервісів, провайдера доступу в мережу Інтернет. З метою забезпечення надійності та безпеки даних необхідно застосування механізмів захисту до яких слід віднести:

- наявність та побудову дублюючих каналів зв'язку;
- наявність та побудову дублюючих потужностей, для можливості переходу на них;
- доступність інформації;
- безпека інформації.

Мета роботи. Розкрити сучасні механізми побудови хмарних технологій як пріоритетний напрямок вдосконалення розподілених систем обробки даних. Дослідити проблему поділу даних в параметрі гібридної хмари, яка полягає у мінімізації вартості виконання запиту навантаження і обмежена двома окремими обмеженнями. Здійснити опис обмежень, та визначити механізм розподілу метрик продуктивності – вартість обробки запиту (тобто продуктивність), витрати та ризики.

Аналіз останніх досліджень і публікацій. На сьогоднішній день, питання сучасних ефективних механізмів побудови хмарних технологій розглядало чимало, як зарубіжних так і вітчизняних вчених. Е. А. Ратушна, В. А. Ковальчук [3] розглядають хмарні технології в освітньому процесі.

Автори пропонують модель архітектури хмарних обчислень, з якої видно, що основу хмари становить інфраструктура як сервіс (IaaS – Infrastructure as a Service), платформа як сервіс (PaaS – Platform as a Service), програмне забезпечення як сервіс (SaaS – Software as a Service). До цієї ж ідеї приєднуються Лисенко В. І. і Колеснікова Т. А. [4], вони підходять до теоретичної складової формування хмарних технологій, як основи для реалізації навчальної діяльності.

У роботі [5] наведено огляд хмарних сервісів, які використовуються для освітньої галузі. Виділено переваги та недоліки поширених хмарних сервісів. Автори, в основі статті підходять до питання про те, що розміщення розподілених систем на гібридній хмарі, а також безпосередньо шифрування даних на стороні клієнта покращує показники безпеки системи загалом.

Т. А. Онуфрієва, А. С. Матросов [6] підійшли до розгляду інформаційної безпеки в інноваційному онлайновому сервісі – хмарні технології.

Архітектура інтеграції хмар розподіленої системи зберігання, оперативного оновлення і надання даних про потенційно небезпечні об'єкти на основі технології cloud computing розглядається у межах роботи [7]. А. В. Меленець на прикладі гібридної хмари показав архітектуру інтеграції хмар різних моделей розгортання на основі інтеграції як сервісу. Наведені в статті архітектури публічної, приватної і гібритної хмар розподіленої системи спільно з інтеграцією хмар як сервісу, дозволяють підвищити оперативність оновлення і надання інформації, а також організувати зберігання даних про потенційно небезпечні об'єкти на основі методів та інформаційної технології розробки та інтеграції розподілених систем.

Із зарубіжних джерел особливу цінність у підході до вивчення механізмів побудови хмарних технологій представляють роботи: R. Buyya [8], V. Egorova, D. Chechulina, S. F. Krendelev [9], J. Barr [10], B. Sosinsky [11], та ін..

Однак, незважаючи на масштабність наукових досліджень в рамках поставленої задачі, питання ефективної побудови хмарних технологій залишається відкритим і потребує подальшого детального вивчення.

Виклад основного матеріалу дослідження. В рамках інформаційних технологій і сучасного інформаційного простору, моделі розгортання хмар, поділяють на три основні види:

- приватні;
- загальнодоступні (публічні);
- гібридні [2, 4].

Перші відносяться до хмар заснованих на інфраструктурі та службах всередині підприємства. Приватні хмари знаходяться всередині корпоративної мережі. Особливістю розгортання приватної хмари, є те, що підприємство саме займається встановленням і підтримкою хмари. При цьому витратність як матеріальна, так і ресурсна, на створення внутрішньої хмари може бути дуже високою, а витрати на її експлуатацію можуть перевищувати вартість використання загальнодоступних хмар. До найбільш важливих переваг, слід віднести, більш детальний контроль над різними ресурсами хмари і, як наслідок, різні варіанти конфігурації, а також підвищений рівень безпеки.

Другий вид – це публічні хмари, сутність їх полягає у загальнодоступності та розгортанні за межами корпоративної мережі. Функціональні дії в рамках хмари покладено на власника.

До третього виду відносяться гібридні хмари, які є сполучною ланкою публічних і приватних хмар. Фундаментальну основу гібридної хмари складає розподіл потужностей, тобто делегування навантаження між хмарами (приватними і публічними), у разі, коли внутрішня IT-інфраструктура не справляється з поточними завданнями, частина

потужностей перекидається на публічну хмара, а також для надання доступу користувачам до ресурсів підприємства (до приватної хмари) через публічну хмару. До недоліків гібридної хмари слід віднести складність ефективного створення подібних рішень і управління ними, необхідність організації послуг одержаних з різних джерел, взаємодію між приватними і загальнодоступними компонентами.

Аналіз сучасної наукової літератури [12, 13] показав, що найближчим часом основною моделлю для більшості великих і середніх підприємств стане гібридна хмара.

Гібридна хмарна архітектура (рис. 1) складається з двох компонентів: компонент проектування даних, який відповідає за оптимальний розподіл в гібридному хмарі, і компонент обробки запитів, який з урахуванням поділу, приймає рішення про стратегію виконання запитів.

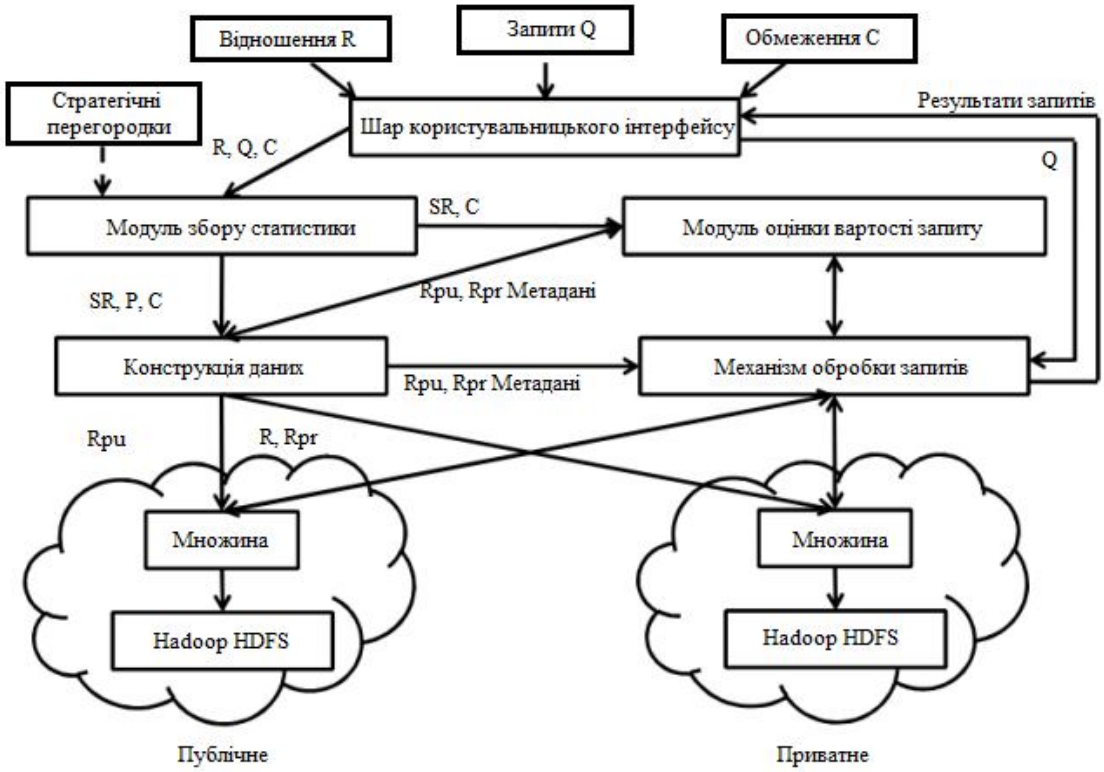

Рис. 1. Архітектура гібридної хмари*

*Власна розробка автора на основі літературних джерел [3, 5]

Оцінка витрат, необхідних для прийняття рішення про визначення оптимального розбиття залежить від стратегії обробки запитів на якій реалізований даний механізм.

Для компонента проектних даних, користувач представляє:

– сукупність відносин, $R = \{R_1, R_2, ..., R_M\}$

– сукупність робочого навантаження запиту, $Q = \{Q_1, Q_2, ..., Q_k\}$

– набір розподілу ресурсів і чутливі обмеження розкриття даних, C.

Система спочатку виконує завдання збору статистики по R і Q за допомогою модуля збору статистики. Цей модуль також створює набір предикатів P, виходячи з R, Q і заданої користувачем стратегії секціонування (вертикальної або горизонтальної).

Статистика SR створюється у вигляді гістограми з однаковою шириною. Рівень даних отримує набір P і статистику SR, а також обмеження C, а потім планомірно вирішує проблему поділу даних, DPP. У вирішенні DPP запиту, розрахунок вартості, $QCEst$ складової системи використовується для оцінки виконання витрат запитів $Q_i \in Q$. $QCEst$, у

свою чергу, оцінює час виконання запитів і визначає оптимальний план їх виконання за допомогою обробки запитів.

Зрештою, цей шар виводить дві часткові репліки

$$R, R_{pr} = \left\{ R_{1\,pr}, R_{2\,pr}, ..., R_{mpr} \right\} ma R_{pu} = \left\{ R_{1\,pu}, R_{2\,pu}, ..., R_{mpu}, \right\},$$

що відповідає приватним і публічним реплікам R.

Приватна хмара зберігає всі сторони репліки R_{pr}, в той час як у публічній хмарі зберігаються тільки ті репліки, R_{pr} і R_{pu}, які створюються шляхом розбиття множини p на приватні (P^{pr}) і публічні (P^{pu}) секції відповідно. Тому в рамках даної статті, посилання на "поділ" означає поділ множини P, при реплікації, це означає, що реплікація досягається за рахунок збереження R_{pr} на приватній стороні і руху R_{pu} у бік публічної множини. Нарешті, при розробці рівня даних також передбачається, зберігання конфіденційних даних у R_{pu}, які будуть зберігатися у загальнодоступній хмарі, після розробки шару даних і визначення реплік (R_{pr} і R_{pu}), а також місця подання (у відкритому або закритому вигляді), конфіденційних даних. З іншого боку, номери-конфіденційних даних у R_{pu}і R_{pr}, які зберігаються у форматі відкритого тексту на публічних і приватних хмарах відповідно.

Метадані по репліках R_{pu}і R_{pr} теж відправляються в підсистему обробки запитів.

При надходженні запиту Q, QPE перетворює Q план виконання, використовуючи правила перезапису та оцінки вартості різних стратегій для виконання Q^4.

Проблема поділу даних в параметрі гібридної хмари полягає у мінімізації вартості виконання запиту навантаження і обмежена двома окремими обмеженнями, перше з яких обмежує ресурси, які можуть бути надані у публічну хмару, а друга фіксує розкриття ризику того, що користувач готовий прийняти, конфіденційні дані публічної сторони. Рішення проблеми призводить до розбиття даних між публічною та приватною сторонами. В рамках даної наукової роботи пропонується виконати моделювання таких перегородок, використовуючи предикати.

У моделі поділу предиката в даній статті, використовуються прості предикати як фундамент, на якому можна здійснювати горизонтальний і вертикальний поділ стратегій, які лежать в основі рішень задачі розбиття даних. Використання предикатів дозволяє представити різні варіанти розбиття даних (між публічною і приватною хмарами) в рамках однієї загальної структури. Слід запропонувати загальний спосіб представлення різних стратегій секціонування, тобто горизонтальний або вертикальний, використовуючи загальний механізм замість розробки окремого позначення для них. Крім того, на відміну від стаціонарних стратегій, таких як цикл або хеш-секціонування, предикати пропонують адаптивну стратегію для контролю над тим, щоб конфіденційні дані були відтворені на публічній хмарі. Попередні методи завдання політики для контролю доступу і конфіденційності також розглядаються аналогічний підхід, наприклад, [9] використаний підхід, заснований на предикаті для визначення політик управління доступом в реляційних базах даних. У [10] використання запитів залежить, від зазначеної політики конфіденційності для структурованих даних. В рамках запропонованої моделі, завдання розбиття даних, полягає у наборі простих предикатів P які походять з відносин R і навантаження Q. Простий предикат, P_i, визначається однією з таких трьох форм:

1) $P_i \leftarrow$ Атрибут;
2) $P_i \leftarrow$ Атрибут op Вартість;
3) $P_i \leftarrow \left(P \wedge P_i\right) \left| \left(P \vee P\right) \right| \left(P_i\right)$,

де op включає $\{=, <, >, \leq, \geq\}$.

Предикати 1-го типу використовуються для визначення вертикального секціонування, в той час як предикати типів 2 і 3 можуть бути використані для горизонтального секціонування.

За допомогою предиката розбиття, є можливість моделювати завдання розбиття даних (DPP), у вигляді оптимізаційних задач, метою яких є поділ набору простих предикатів, $P = \{P_1, P_2, ..., P_l\}$, над гібридною хмарою, щоб загальна вартість виконання робочого навантаження Q була мінімальною.

Завдання розбиття (DPP) будується як оптимізаційна задача, яка визначається простим предикатом множини P', де $P' \subseteq P_t$:

$$\min_p \sum_{i=1}^{n} freg(Q_i) \times QPC_{Q_i}(P)$$

$$subject\_to\_store(P') + \sum_{i=1}^{n} freg(Q_i) \times \left(comm_{Q_i}(P') + proc_{Q_i}(P')\right) \leq PRA\_COST$$

$$sens\left(\bigcup_{p \in p'}\right) \times dis\_\cos t \leq DIS\_COST,$$

де $freg(Q_i)$ – доступ до частоти запиту Q_i;

$QPC_{Q_i}(P')$ – обробка запитів вартості Q_i враховуючи, що множина P' зберігається на публічній хмарі;

PRA_COST – максимально допустиме виділення коштів на витрати;

DIS_COST – максимально допустиме розкриття витрат.

Використання запропонованої моделі у поєднанні з обмеженнями (публічна сторона витрат і конфіденційні дані про ризики) дозволяє захопити кілька реалістичних сценаріїв у тих же рамках.

Прикладами таких сценаріїв можна назвати:

– користувачів, які не допускають зберігання конфіденційних даних у публічних хмарах, із-за законів/правил.

– користувачів, які хочуть досягти швидкості у продуктивності, і готові платити за ризик зберігання конфіденційних даних на публічній стороні. Крім того, такі загальні рамки також дозволяють вивчати різні співвідношення, які існують у предметній області на систематичній основі.

Формалізація задачі DPP описаної вище відноситься до трьох різних метрик продуктивності – вартість обробки запиту (тобто продуктивність), витрати та ризики.

Продуктивність (QPC): продуктивність запиту залежить від стратегії, використовуваної для виконання запиту. Враховуючи конкретний план запиту, є можливість оцінити витрати продуктивності з використанням стандартних методів оцінки вартості для розподіленої обробки запитів. Вартість включає в себе розрахунок на публічних і приватних хмарах, а також мережу обміну даними між приватними і публічними хмарами.

Витрати: всі хмарні провайдери, як правило, підтримують конкурентні цінові моделі і надають різні угоди про рівень обслуговування (SLA) для зберігання даних і зручності обробки.

Ризики: ризик розкриття є важливим питанням для організацій, які мають справу з конфіденційними даними, так як в тому випадку, якщо вони втратять секретну інформацію, вони будуть зобов'язані платити штрафи відповідності, а також можливі судові витрати [11].

В рамках описаної моделі, оцінюється сумарне розкриття витрат на набір предикатів $\left(P^{'}\right)$ як твір кількості чутливих кортежів над предикатом безлічі P' (обраховується як $sens\left(\bigcup_{p\in P^{'}}\right)$, так як там може бути перекриття предикатів). Крім того, ризик розкриття залежить від представлення даних, які використовуються для зберігання конфіденційних даних, що фіксується в задачі розбиття даних. Крім того, обчислена вартість розкриття обмежена користувальницьким значенням, DIS_COST[7].

Висновки. У межах даної статті у моделі поділу предиката використовуються прості предикати як фундамент, на якому можна здійснювати горизонтальний і вертикальний поділ стратегій, які лежать в основі рішень задачі розбиття даних. Використання предикатів дозволяє представити різні варіанти розбиття даних (між публічною і приватною хмарами) в рамках однієї загальної структури. Представлена модель може використовувати інші методи оцінки ризику зміни вартості розкриття інформації для предиката p_j.

Перспективи подальших розробок у даному напрямку спираються на можливість об'єднати ризики і витрати в принципову моду на одне обмеження. Це, вимагає нормалізувати дві витрати в єдину метрику.

Список використаної літератури

1. Публичные облака vs Частные облака // Tadviser. Государство. Бизнес. IT. – http://www.tadviser.ru/index.php

2. Гибридное облако для малого бизнеса // VMware. – http:// vmware.com./ru/smb/cloud-services-hosting

3. Ратушная Е.А., Ковальчук В.А. Облачные вычисления: новые технологии в образовании // Международный студенческий научный вестник. – 2014. – № 1. – С. 40.

4. Лысенко В. И. Облачные технологии в образовательном процессе / В. И. Лысенко, Т. А. Колесникова // 1-я Международная научно-техническая конференция «Полиграфические, мультимедийные и WEB-технологии (PMW-2016)», 16–20 мая 2016 г. – Харьков : ХНУРЭ, 2016. – Т. 2 : – С. 101-105.

5. Использование облачных технологий как способ повышения защищенности тестовых обучающих систем / Н.А. Маслова, Р.А. Сорокин // Искусственный интеллект. – 2013. – № 4. – С. 463–475.

6. Онуфриева Т.А., Матросов А.С. Информационная безопасность в инновационном интерактивном сервисе-облачные технологии // Международная научно-практическая конференция «Проблемы и перспективы технических наук». –2015. – С. 161-164.

7. Меленец А.В. Архитектура интеграции облаков распределенной системы хранения, оперативного обновления и предоставления данных о потенциально опасных объектах на основе технологии Cloud Computing / А.В. Меленец // Радіоелектронні і комп'ютерні системи. – 2012. – № 7 (59). – С. 54-59

8. Buyya, R. Goscinski Cloud computing. Principles and paradigms / R. Buyya, J. Broberg // John Wiley & Sons, Inc. – Hoboken, New Jersey, USA, 2011. – 637 p.

9. Egorova V., Chechulina D., &Krendelev S. F. (2013) New View on Block Encryption (Unpublished) Available: https://db.tt/vnE9wfg

10. Barr, J. Host Your Web Site in the Cloud: Amazon Web Services Made Easy / J. Barr // Amazon Web Services, LLC, a Delaware limited liability company, Seattle, USA, 2010. – 364 p.

11. Sosinsky, B. Cloud computing bible / B. Sosinsky // John Wiley & Sons, Inc. – Hoboken, New Jersey, USA, 2011. – 708 p.

12. Cloud Security Alliance. The Notorious Nine. Cloud Computing Top Threats in 2013. – https://downloads.cloudsecurityalliance.org/initiatives/top_threats/The_Notorious_Nine_Cloud_Computing_Top_Threats_in_2013.pdf

13. Паскова А. А., Бутко Р. П. Гибридные облака в IT-инфраструктуре предприятия // Экономика и экономические науки: Символ науки. – 2016. – № 10-2. – С. 73-75.

References

1. Public Clouds vs Private Clouds // Tadviser. Gosudarstvo. Biznes. IT. – http://www.tadviser.ru/index.php

2. Hybrid Cloud for Small Business // VMware/ – http:// vmware.com./ru/smb/cloud-services-hosting

3. Ratushnaja E.A., Koval'chuk V.A. Cloud Computing: New Technologies in Education] // Mezhdunarodnyj studencheskij nauchnyj vestnik. – 2014. – № 1. – PP. 40.

4. Lysenko V. I. Cloud Technologies in Educational Process / V. I. Lysenko, T. A. Kolesnikova // 1-st International scientific technical conference «Polygraph, multymedia and WEB-technology – 2016. – T. 2 : – PP. 101-105.

5. Use of Cloud Technologies as a Way to Increase Security of Test Training Systems / N.A. Maslova, R.A. Sorokin // Iskusstvennyj intellekt. – 2013. – № 4. – PP. 463–475.

6. Onufrieva T.A., Matrosov A.S. Information Security in an Innovative Interactive Service-Cloud Technologies // International scientific practical conference «Problem and prospect of engineerings sciences». – 2015. – PP. 161-164.

7. Melenec A.V. Cloud Integration Architecture of Distributed Storage System, Online Updation and Provision of Data on Potentially Dangerous Objects Based on Cloud Computing Technology / A.V. Melenec // Radioelektronni i komp'juterni sistemi. – 2012. – № 7 (59). – PP. 54-59.

8. Buyya, R. Goscinski Cloud computing. Principles and paradigms / R. Buyya, J. Broberg // John Wiley & Sons, Inc. – Hoboken, New Jersey, USA, 2011. – 637 p.

9. Egorova V., Chechulina D., &Krendelev S. F. (2013) New View on Block Encryption (Unpublished) Available: https://db.tt/vnE9wfg

10. Barr, J. Host Your Web Site in the Cloud: Amazon Web Services Made Easy / J. Barr // Amazon Web Services, LLC, a Delaware limited liability company, Seattle, USA, 2010. – 364 p.

11. Sosinsky, B. Cloud computing bible / B. Sosinsky // John Wiley & Sons, Inc. – Hoboken, New Jersey, USA, 2011. – 708 p.

12. Cloud Security Alliance. The Notorious Nine. Cloud Computing Top Threats in 2013. Available: https://downloads.cloudsecurityalliance.org/initiatives/top_threats/The_Notorious_Nine_Cloud_Computing_Top_Threats_in_2013.pdf

13. Paskova A. A., Butko R. P. Hybrid Clouds in IT-Infrastructure of Enterprise // Jekonomika i jekonomicheskie nauki: Simvol nauki, 2016. – № 10-2. – PP. 73-75.

Автор статті

Лабжинський Володимир Анатолійович – доцент, кандидат технічних наук, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського»

Author of the article

Labzhynskyi Volodymyr Anatoliiovych – candidate of sciences (technical), National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Дата надходження в редакцію: 14.04.2017 р.

Рецензент: доктор технічних наук, професор М. М. Климаш *Національний університет «Львівська політехніка»*