

DOI: 10.36910/6775-2524-0560-2020-40-07

УДК: 004.934

Коляда Костянтин Вячеславович, асистент

<https://orcid.org/0000-0002-3962-5791>

Романкевич Віталій Олексійович, д.т.н., зав. каф.

<https://orcid.org/0000-0003-4696-5935>

Орлова Марія Миколаївна, к.т.н., доцент

<https://orcid.org/0000-0002-6617-4631>

Марковський Олександр Петрович, к.т.н., доцент

<https://orcid.org/0000-0003-3483-4233>

КПІ ім. Ігоря Сікорського, м. Київ, Україна

МЕТОД ВІДНОВЛЕННЯ ДАНИХ ПРИ ЇХ РОЗПОДІЛЕНОМУ ЗБЕРІГАННІ НА ВІДДАЛЕНИХ СХОВИЩАХ

Коляда К.В., Романкевич В.О., Орлова М.М., Марковський О.П. Метод відновлення даних при їх розподіленому зберіганні на віддалених сховищах. Запропоновано метод, який дозволяє зменшити надлишковість резервування, за рахунок врахування специфіки реальних втрат інформації, що зберігається на віддалених носіях і гарантує її відновлення в разі втрати доступу до одного сховища в результаті катаклізмів або до трьох блоків на різних сховищах.

Ключові слова: рознесене зберігання даних, хмарні технології, відновлюючі коди, лінійні коди.

Коляда К.В., Романкевич В.А., Орлова М.Н., Марковський А.П. Метод восстановления данных при их удаленном хранении на удаленных хранилищах. Предложен метод, который позволяет уменьшить избыточность резервирования за счет учета специфики реальных потерь информации, которые сохраняются на удаленных носителях. Метод гарантирует восстановление данных в случае потери доступа к одному хранилищу в результате катаклизмов или трех блоков в разных хранилищах.

Ключевые слова: распределенное хранение данных, облачные технологии, восстанавливающие коды, линейные коды.

Koliada K., Romankevich V., Orlova M., Markovskiy O. Method for data recovery in their distributed storage on remote memory nodes. The method is proposed that makes it possible to reduce excess of redundancy by taking into account the specifics of real information losses that are stored on remote media. The method guarantees data recovery in case of loss of access to one repository as a result of cataclysms or three blocks in different repositories as a result of hardware breakdown or program errors.

Keywords: distributed data storage, cloud technologies, erased codes, linear codes.

Постановка наукової проблеми.

Досягнутий в останнє десятиліття прогрес глобальних мереж кардинально змінив організацію комп'ютерної обробки інформації на користь розподілених систем зберігання даних та віддалених обчислень. Широкого розповсюдження набула практика розподіленого зберігання інформації користувачів на віддалених сховищах. Крім таких очевидних переваг, як практичну відсутність обмежень на об'єм пам'яті, доступність інформації користувача для широкого кола інших користувачів, зручність обміну даними, сучасна технологія зберігання інформації в хмарах забезпечує якісно більш високий рівень надійності.

Втрата даних при їх зберіганні на магнітних, оптичних чи твердотільних (флеш-пам'ять) накопичувачах може відбуватися в результаті виходу з ладу обладнання, помилок в роботі програмного забезпечення, хакерських атак, дій вірусних програм, комп'ютерних хробаків, катаклізмів політичного, терористичного, природного, військового чи техногенного характеру. При цьому можуть втрачатися як окремі блоки даних, так і всі інформація на певному сховищі. При зберіганні даних користувачів хмарі, основну роль в забезпеченні надійності відіграє рознесення даних користувача по різним, географічно віддаленим, сховищам. Це обмежує втрати даних під дією наведених вище причин, відносно невеликим об'ємом. Втрачені дані можуть бути відновлені за рахунок використання резервних блоків, які формуються з інформаційних блоків.

В цьому ракурсі резервування та відновлення втрачених даних відіграють важливу роль в технології забезпечення високої надійності розподіленого зберігання інформації користувачів. Відповідно, подальший розвиток технологій віддаленого зберігання даних диктує необхідність адекватного вдосконалення методів їх резервування та відновлення втраченої інформації.

Таким чином, наукова задача підвищення ефективності резервування та відновлення даних користувачів при їх розподіленому зберіганні на віддалених сховищах є актуальною для сучасного етапу розвитку інформаційних та комп'ютерних технологій.

Аналіз технологій резервування та відновлення даних при їх віддаленому зберіганні.

Швидкий розвиток технологій віддаленого зберігання даних користувачів стимулював інтенсивні дослідження в області забезпечення надійності, резервування та відновлення втрачених даних.

Характерною особливістю сучасного етапу розвитку технології забезпечення надійності віддаленого зберігання даних є багаторівневність [1].

Перший рівень в забезпеченні надійності зберігання даних утворюють засоби, що функціонують на рівні секторів магнітних чи оптичних носіїв. В процесі їх форматування здійснюється контрольне зчитування інформації з них: при виявленні двох чи більше спотворених байтів, сектор позначається як дефектний і запис на нього інформації блокується. Крім того, кожен сектор включає, крім 512-ти інформаційних, 4 резервних байти, які являють собою контрольні символи коду Ріда-Соломона [2] і, відповідно, забезпечують можливість виправлення двох байтів або виявлення факту спотворення більшої кількості байтів. Тобто, на першому рівні використовуються класичні корегуючі коди [2], які передбачають реалізацію двох фаз виправлення даних: локалізацію спотворених символів та їх відновлення.

В сховищах віддаленого зберігання даних користувачів, другий рівень в забезпеченні надійності утворюють резервні накопичувачі. Найбільш розвинутою технологією цього рівня є широко відома RAID (Redundant Array of Independent Disks) [3]. Система RAID-3 використовує один резервний накопичувач, на якому зберігаються суми за модулем два всіх однойменних інформаційних блоків на основних накопичувачах сховища. Це дозволяє відновити дані з будь-якого одного накопичувача.

Система RAID-5 також використовує в якості резервних даних суму за модулем два однойменних блоків всіх інформаційних накопичувачів. Проте, на відміну від RAID-3 резервні блоки розміщуються не на одному накопичувачу, а на всіх $n+1$ накопичувачах. RAID-5 дозволяє відновлювати тільки один втрачений блок з однаковим номером і всі втрачені блоки на одному чи різних накопичувачах, за умови, що вони мають різні номери.

Система RAID-6 використовує два резервні блоки для відновлення двох із n інформаційних блоків. Відповідно, система потребує для своєї реалізації 2 резервних накопичувача. Реально використовуються $n+2$ накопичувача з рознесенням резервних блоків по різних накопичувачах. Фактично RAID-6 гарантує відновлення 2 однойменних блоків за умови збереження інших n інформаційних та резервних блоків. Іншими словами, RAID-6 дозволяє відновлювати дані з будь-яких двох втрачених накопичувачів.

Сучасні моделі RAID виходять з того, що локалізація втраченої інформації точно відома [4]. Це визначає доцільність використання для відновлення втрачених даних максимально рознесених кодів, або MDS (Maximum Distance Separable)-кодів. Для відновлення двох чи більшої кількості втрачених блоків можуть бути використані різні різновиди MDS-кодів: коди Ріда Соломона [2], матричні коди типу EVENODD [2], RDP [4], X-коди [5], В-коди [3], С-коди [6].

Основні недоліки систем забезпечення відновлення даних в сховищах типу RAID полягають в тому, вони не захищені від катаклізмів, здатних зруйнувати всю інформацію, що зберігається в сховищі, не дозволяють відновлювати більше двох блоків, не враховують специфічних вимог щодо надійності зберігання даних кожного з користувачів.

Аналіз невирішених аспектів проблеми ефективного відновлення втрачених даних.

Вказані недоліки можуть бути подолані на третьому рівні забезпеченні надійності віддаленого зберігання даних [7], тобто на рівні користувача. Досягнутий на сьогодні прогрес технології глобальних мереж надає можливість здійснювати резервування і відновлення даних на рівні користувача або групи користувачів. При цьому повною мірою можна задіяти можливості рознесеності зберігання інформації для локалізації її можливих втрат. Тільки цей чинник здатен реально забезпечити захист даних від втрат внаслідок катаклізмів. Проте рознесення зберігання даних користувача негативно впливає на швидкість доступу до даних – їх потрібно зібрати з різних сховищ [8].

Таким чином, для підвищення надійності зберігання даних користувачів з урахуванням їх специфічних вимог, резервування має здійснюватися на рівні, вищому ніж рівень окремого сховища.

Вище зазначалося, що прийнята в системі RAID-6 орієнтація на відновлення лише двох блоків даних може вважатися прийнятною лише для середнього рівня вимог до надійності зберігання. Разом з тим, існує коло користувачів, специфіка яких висуває суттєво більш жорсткі вимоги до надійності зберігання даних. Крім того, постійно розширюється використання твердотільних накопичувачів, типу flash-накопичувачів. Ці пристрої мають значно меншу технічну надійність, яка сильно залежить від часу зберігання та кількості циклів перезапису інформації [9]. З наведеного випливає наявність об'єктивної необхідності в розробці методів резервування, здатних відновлювати більшу ніж два втрачені блоки даних [10].

Аналіз можливостей використання згаданих вище MDS-кодів для відновлення більш ніж двох блоків даних показує, що при зі зростанням числа блоків, що можуть бути реконструювані при втратах, швидко зростає потрібна для цього кількість резервних блоків. Тому виникає потреба в розробці методу резервування, який би забезпечував вирішення задачі відновлення даних з урахуванням практики їх втрат в реальних системах віддаленого зберігання з використанням помірної кількості резервних блоків.

Мета досліджень. Таким чином, мета досліджень полягає в підвищенні ефективності відновлення даних, шляхом зменшення надлишковості резервування за рахунок урахування реальних особливостей втрати доступу до інформації при її віддаленому зберіганні.

Метод відновлення всіх блоків даних на одному сховищі та трьох блоків на різних сховищах.

Для досягнення поставленої мети досліджено характер втрат доступу до блоків даних користувачів. Характер втрат даних значною мірою визначається їх причиною. Втрати даних, викликані поломками обладнання, старінням матеріалу носія, помилками в роботі програмного забезпечення найбільш точно описуються в математичному сенсі потоками Пуассона [4]. При цьому інтенсивність втрати одного блоку даних за даними Google Cloud Storage становить близько $7 \cdot 10^{-6}$ год⁻¹ [6]. Використовуючи профілактичний періодичний контроль збереженості блоків даних можна обмежити кількість втрачених блоків.

Для втрат даних користувачів, зумовлених дією зовнішніх катаклізмів та хакерських чи вірусних атак характерним є неможливість доступу до всіх даних сховища.

Таким чином, метод має гарантувати можливість реконструкції всіх блоків даних користувача, що зберігаються на одному сховищі, втрата яких зумовлена зовнішніми катаклізмами чи хакерськими атаками та відносно невелику кількість інформаційних блоків, доступ до яких втрачено з причин відмов обладнання та помилок в роботі програм.

Для досягнення поставленої розроблено метод відновлення інформаційних блоків при їх розподіленому зберіганні не більш ніж по r на кожному з віддалених сховищ, який гарантує реконструкцію всіх r блоків при втраті доступу до одного сховища, на якому вони зберігаються, а також трьох блоків, які зберігаються на різних сховищах.

В рамках дослідження розглядається така модель збереження інформації на віддалених носіях. Користувач зберігає на віддалених серверах n інформаційних блоків B_1, B_2, \dots, B_n . Для їх відновлення при формуються m резервних блоків R_1, R_2, \dots, R_m у вигляді лінійних перетворень над інформаційними блоками:

$$\forall i \in \{1, 2, \dots, m\} : R_i = \bigoplus_{j=1}^n a_{i,j} \cdot B_j,$$

де $a_{i,j} \in \{0, 1\}$ – бінарні коефіцієнти, які утворюють матрицю A і визначають спосіб формування резервних блоків даних:

$$A = \|a_{i,j}\|, i = 1, 2, \dots, m, j = 1, 2, \dots, n.$$

Для гарантування можливості відновлення не менше трьох втрачених від загального числа $q = m + n$ блоків користувача необхідним є виконання наступних умов:

1. Кожен стовпець матриці A має містити не менше 3-х одиниць, що гарантує відновлення одного інформаційного блоку в ситуації, коли крім нього втрачено два резервних.

2. Хемінгова відстань між будь-якими двома стовпцями матриці A має бути не менше двох, що гарантує відновлення пари інформаційних блоків за умови, що втрачено один із резервних блоків. Іншими словами, після втрати одного резервного блоку всі стовпці матриці A мають бути різними.

3. Матриця Δ , утворена будь-якими трьома стовпцями матриці A , має містити в собі ортогональну квадратну підматрицю Θ , що визначає можливість представлення будь-яких трьох інформаційних блоків у вигляді лінійної комбінації інших інформаційних та резервних блоків, тим самим гарантуючи відновлення будь-яких трьох втрачених інформаційних блоків.

В роботі запропоновано наступний метод побудови матриці A формування резервних блоків для відновлення трьох втрачених блоків. Метод полягає в виконанні наступної послідовності дій:

1. Обчислюється значення k кількості двійкових розрядів для кодування номеру стовпчика за умови, що їх нумерація починається з одиниці: $k = \lceil \log_2(n+1) \rceil$

2. В перші k рядків кожного стовпчика матриці A записується його порядковий номер, починаючи з одиниці.

3. В $(k+1)$ -му рядку тих стовпчиків, номери яких мають парну кількість одиниць, встановлюється одиниця, а тих, що мають непарну кількість одиниць – нуль.

4. В $(k+2)$ -му рядку тих стовпчиків, номери яких мають одну одиницю, встановлюється одиниця, а тих, що мають дві і більше одиниць – нуль.

Покажемо, що матриця A побудована згідно з запропонованим методом задовольняє сформульованим вище трьома умовам.

Оскільки нумерація стовпців згідно викладеного вище починається з одиниці, що гарантує хоча б однієї одиниці в перших k рядках кожного стовпчика матриці A . У відповідності з п.4 в $(k+2)$ -му рядку тих стовпчиків, номери яких мають лише одну одиницю, встановлюється одиниця, так, що кількість одиниць у кожному стовпчику не може бути меншою за два.

Якщо Хемінгова відстань між парою номерів дорівнює одиниці, то це можливо лише за умови, що кількість одиниць в цих номерах відрізняється на одиницю, тобто один із номерів парний, а інший – непарний. Описаний вище метод передбачає додавання одиниці в $(k+1)$ -му рядку лише тих стовпців, сума одиниць в k перших позиціях парна. Це означає, що з урахуванням $(k+1)$ -х перших компонент Хемінгова відстань між будь-якою парою стовпців матриці A гарантовано не менша двох.

Будь-які три різні номери відрізняються мінімум в двох розрядах. Це означає, що серед перших k рядків матриці A , побудованою за викладеним вище методом, для будь-яких трьох стовпців є мінімум два рядки, компоненти яких не рівні між собою і ці два рядки різні.

Якщо таких рядків рівно два, то один не є інверсією іншого. В протилежному випадку існували б два однакових номери, що протирічить викладеному методу. Якщо цих рядків рівно два і в цих двох рядках існує розряд, який дорівнює нулю в обох рядках, то серед інших перших k рядків матриці A має бути рядок, що складається з одиниць, інакше номер одного із стовпців дорівнює нулю, що входить в протиріччя з викладеним методом.

Таким чином, якщо серед перших k рядків вибраної трійки стовпців матриці A є рівно два таких, компоненти яких не рівні між собою і ці два рядки різні, але не інверсні, то можна розглядами два випадки. У першому випадку існує компонента, яка дорівнює нулю в обох рядках. Але в цьому випадку обов'язково існує і рядок, що складається з одиниць і який, разом з двома зазначеними вище рядками утворюють ортогональну систему Θ . В другому випадку в парі різних і не інверсних рядків в обох з них є дві одиниці і точно існує компонента, яка в обох рядках дорівнює одиниці. Якщо в обраних трьох стовпців матриці A в перших k рядках існує рядок, всі компоненти якого дорівнюють одиниці, то цей рядок, разом з двома згаданими вище утворює ортогональну систему Θ . Якщо в вибраних трьох номерах нема рядка, що складається з усіх одиниць, то це означає, що точно один з цих номерів має парну кількість одиниць. Згідно з п.4 викладеного методу в $(k+2)$ -му рядку цей стовпець містить одиницю, а два інші – нулі. Відповідно $(k+2)$ -й рядок разом з двома, що розглядаються, утворюють ортогональну систему Θ . Таким чином доведено, що за умови, що номери трьох вибраних стовпчиків матриці A відрізняються точно в двох розрядах (рядках), в матриці A точно існує ортогональна квадратна матриця Θ , стовпці якої є компонентами вибраних стовпців.

Якщо номери трьох вибраних стовпчиків матриці A відрізняються більш ніж в двох розрядах, то існує не менше трьох різних рядків з не рівними між собою компонентами.

За аналогією з викладеним вище можна розглядати два випадки. У першому в усіх трьох різних рядках існує компонента яка дорівнює нулю, а в другому випадку такої компоненти нема.

Якщо в цих трьох рядках існує компонента, яка дорівнює в них нулю, то дві інших компоненти утворюють пари $\{0,1\}$, $\{1,0\}$, $\{1,1\}$. При цьому обов'язково існує і рядок, що складається з одиниць. В протилежному випадку треба допустити існування нульового номеру, що входить в протиріччя з

викладеним методом. В цій ситуації рядок вибраної трійки номерів, що складається з одиниць разом з рядками, що містять пари $\{0,1\}$, $\{1,0\}$ утворюють ортогональну систему Θ . В другому випадку серед k перших рядків обраної трійки стовпців матриці A існує мінімум три різні рядки, причому жодна компонента не дорівнює нулю в усіх трьох рядках, ці три рядки утворюють ортогональну систему Θ якщо серед них є хоча б один з непарним числом одиниць. Якщо серед трійки рядків нема рядків з непарною кількістю одиниць, то вони не утворюють ортогональної системи. Якщо серед інших рядків вибраної трійки номерів є рядок, що складається з одиниць, то він утворює ортогональну систему Θ з будь-якою парою наведених вище стовпців.

Якщо три різних рядки трійки номерів дорівнюють $\{0,1,1\}$, $\{1,0,1\}$ та $\{1,1,0\}$ і всі інші рядки вибраної трійки номерів однакові з наведеними трьома або складаються з нулів, то серед вибраних номерів обов'язково будуть такі, число одиниць в котрих парне. Причому кількість таких стовпців обов'язково непарна тому, що кожен рядок трійки номерів містить парну кількість одиниць. Згідно з п.4 викладеного методу в $(k+2)$ -му рядку в таких стовпцях розміщуються одиниці. Відповідно, $(k+2)$ -й рядку міститься завжди непарна кількість одиниць. Цей рядок разом з будь-якою парою рядків $\{0,1,1\}$, $\{1,0,1\}$ та $\{1,1,0\}$ утворюють ортогональну систему. Наприклад, номери 7, 9 і 14 містять чотири таких рядка: $\{0,1,1\}$, $\{1,0,1\}$, $\{1,0,1\}$ та $\{1,1,0\}$. Сума одиниць в номері 7 непарна, в номері 9 парна і в номері 14 – непарна. Тому $(k+2)$ -й рядок формується у вигляді $\{0,1,0\}$, тобто містить непарну кількість одиниць. Очевидно, що цей рядок разом з рядками $\{0,1,1\}$, $\{1,0,1\}$ утворюють ортогональну систему Θ .

Таким чином, доведено, що матриця A сформована за викладеним методом повністю відповідає умовам, які зумовлюють гарантовану можливість всіх втрачених інформаційних блоків за умови втрати не більше трьох від загального числа інформаційних та резервних блоків.

З запропонованого методу очевидно, що кількість резервних блоків для гарантованого відновлення інформаційних пакетів за умови втрати не більше трьох від загальної кількості блоків визначається формулою:

$$m \geq \lceil \log_2(n+1) \rceil + 2.$$

На основі проведених теоретичних досліджень запропоновано метод, який гарантує реконструкцію всіх r блоків при втраті доступу до одного сховища, на якому вони зберігаються, а також трьох блоків, які зберігаються на різних сховищах.

Розроблений метод передбачає, що стовпці матриці A розділяються на $g = \lceil n/r \rceil$ фрагментів так, що в перших $g-1$ фрагментах міститься по r стовпців, а g -й фрагмент містить $n-r \cdot (g-1)$ стовпців.

Побудову матриці A пропонується здійснювати наступним порядком:

1. Встановити індекс j в нуль: $j=0$.
2. $(i+j) \bmod r$ -тий рядок кожного i -того фрагменту матриці A , $i \in \{1, 2, \dots, g\}$, заповнюється $(r-j)$ -розрядним кодом, що складається одиниць на $(r+j-1-i) \bmod r$ – позиції та циклічно наступних за нею в рамках r -бітового фрагменту $r-1-j$ нулів.
3. Здійснюється інкремент індексу j : $j=j+1$. Якщо $j < r$, то виконується перехід на повторне виконання п.2.
4. Невизначені елементи перших $\lceil \log_2(n+1) \rceil$ рядків матриці A заповнюються таким чином, щоб для кожного стовпця вони утворювали унікальний номер, починаючи з одиниці.
5. Наступні рядки матриці A заповнюється таким чином, щоб кількість одиниць в кожному зі стовпців матриці A була не меншою трьох.

Наприклад, для $r=4$ і $n=15$ матриця A формування резервних блоків, побудована за розробленим методом, має наступний вигляд.

$$A = \begin{vmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{vmatrix}$$

При втраті всіх блоків $V_9, V_{10}, \dots, V_{12}$, які зберігаються на 3-му сховищі, відповідні стовпці матриці A містять в собі ортогональну підматрицю Θ :

$$\Theta = \begin{vmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix}$$

Відповідно, втрачені блоки відновлюються наступним порядком:

$$\begin{aligned} V_9 &= R_6 \oplus V_2 \oplus V_3 \oplus V_5 \oplus V_6 \oplus V_{14} \\ V_{10} &= R_3 \oplus V_2 \oplus V_4 \oplus V_7 \oplus V_8 \oplus V_{13} \oplus V_{14} \oplus V_{15} \\ V_{11} &= R_4 \oplus V_1 \oplus V_3 \oplus V_5 \oplus V_7 \oplus V_8 \oplus V_{10} \oplus V_{13} \\ V_{12} &= R_1 \oplus V_{11} \oplus V_4 \oplus V_5 \oplus V_6 \oplus V_7 \oplus V_8 \oplus V_{14} \end{aligned}$$

При втраті трьох блоків V_1, V_7, V_{12} , що зберігаються на різних сховищах, відповідні стовпці матриці A містять ортогональну підматрицю Θ :

$$\Theta = \begin{vmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{vmatrix}$$

Відповідно, втрачені блоки відновлюються наступним порядком:

$$\begin{aligned} V_7 &= R_3 \oplus V_2 \oplus V_4 \oplus V_6 \oplus V_{10} \oplus V_{13} \oplus V_{14} \oplus V_{15} \\ V_{12} &= R_1 \oplus V_7 \oplus V_4 \oplus V_5 \oplus V_6 \oplus V_8 \oplus V_{11} \oplus V_{12} \oplus V_{14} \\ V_1 &= R_5 \oplus V_2 \oplus V_3 \oplus V_6 \oplus V_9 \oplus V_{10} \oplus V_{12} \oplus V_{15} \end{aligned}$$

Таким чином, запропонований метод формування резервних блоків і відновлення інформаційних блоків при їх розподіленому зберіганні не більш ніж по r на кожному з віддалених сховищ, гарантує реконструкцію всіх r блоків при втраті доступу до одного сховища, на якому вони зберігаються, а також будь-яких трьох інформаційних блоків, які зберігаються на різних сховищах.

Оцінка ефективності. Основний ефект запропонованого методу полягає в тому, що суттєвим чином зменшується надлишковість резервування в порівнянні з відомими методами за рахунок урахування особливостей втрати доступу до даних при їх віддаленому зберіганні. Надлишковість резервування зменшується в μ раз, чисельне значення μ визначається формулою:

$$\mu = \frac{(r-2) \cdot \log_2(n+1) + 2}{r+2}.$$

Зокрема, для наведеного прикладу кількість резервних блоків для відновлення 4-х, довільно локалізованих по сховищам блоків, дорівнює 10, а в запропонованому методі, який дозволяє відновлювати 4 блоки, що зберігаються на одному сховищі і три, довільно локалізованих блоків, потрібно 6 резервних блоків. Відповідно, надлишковість резервування зменшується в $\mu = 1.67$ раз.

Висновки та перспективи подальшого дослідження. В результаті проведених досліджень теоретично обґрунтовано, розроблено метод відновлення інформаційних блоків при їх розподіленому зберіганні не більш ніж по r на кожному з віддалених сховищ, який базується на використанні для формування резервних блоків матриці, що є ортогональною в межах суміжних r стовпців, а також будь-які три стовпці якої містять ортогональну підматрицю, що гарантує реконструкцію всіх r блоків при втраті доступу до одного сховища, на якому вони зберігаються, а також трьох блоків, які зберігаються на різних сховищах. Метод використовує меншу кількість резервних блоків, тобто має нижчий рівень надлишковості резервування в порівнянні з відомими методами, орієнтованими на відновлення будь-яких r втрачених блоків.

Подальший розвиток запропонованого підходу вбачається в дослідженні можливостей прискорення процесу відновлення втрачених при віддаленому зберіганні даних за рахунок оптимізації їх розміщення на віддалених сховищах за критерієм мінімізації часу доставки блоків, потрібних для реконструкції втраченої інформації користувачів.

Список бібліографічного опису

1. Bardis N., Markovskiy O. (2013) Method for Cloud Storage Data Recovery with Limited Loss of Acces. Proc. of 4-th International Conference on Mathematics and Computers in Sciences and Industry, 24-26 August 2017, Corfu, Greece, MCSI-17, P.55-61.
2. Blaum M., James Lee Hafner, Steven Hetzler (2013) Partial MDS Codes and Their Application to RAID Type of Architectures, IEEE Trans. Inf. Theory. vol.59, no.7, P. 4510-4519.
3. Im S., Shin D. (2011) Flash-aware RAID techniques for dependable and high-performance flash memory SSD, IEEE Trans. Comput. vol.C-60, no.1, P. 80-92.
4. Li M., Shu J. (2011) C-codes: Cyclic lower-density MDS array codes constructed using starters for RAID-6, IBM New York USA. Res. Report RC25218.
5. Li J., Tang X., Tian C. (2019) A Generic Transformation to Enable Optimal Repair in MDS Codes for Distribution Storage System, IEEE Trans. Inf. Theory, vol.65, no.9, P.6257-6267.
6. Rhea S. (2011) Maintenance-free global data storage, IEEE Internet Comput. vol. 5, no.5, P. 40-49.
7. Koliada K.,V., Bardis N.G., Markovskiy O.P. (2019) Usage of linear erasure codes for increasing reliability and efficiency of information delivery on the internet, International Journal of Circuits, Systems and Signal Processing, vol.13, P. 585-592.
8. Rashmi K.V., B. Nihar , K. Ramchandran K. (2018) Information-Theoretically Secure Erasure Codes for Distrubute Storage, IEEE Trans. Inf. Theory, vol.64, no.3, P. 1621-1645.
9. Kai X., Zhu S., Li P. (2015) A Construction of New MDS Symbol-Pair Codes, IEEE Trans. Inf. Theory, vol.61, no.11, P. 5828-5834.
10. Yehezkeally Y. and Schwartz M. (2020) Reconstruction Codes for DNA Sequences With Uniform Tandem-Duplication Errors, IEEE Trans. Inf. Theory, vol.66, no.5, P. 2658-2669.

References

1. Bardis N., Markovskiy O. (2013) Method for Cloud Storage Data Recovery with Limited Loss of Acces. Proc. of 4-th International Conference on Mathematics and Computers in Sciences and Industry, 24-26 August 2017, Corfu, Greece, MCSI-17, P.55-61.
2. Blaum M., James Lee Hafner, Steven Hetzler (2013) Partial MDS Codes and Their Application to RAID Type of Architectures, IEEE Trans. Inf. Theory. vol.59, no.7, P. 4510-4519.
3. Im S., Shin D. (2011) Flash-aware RAID techniques for dependable and high-performance flash memory SSD, IEEE Trans. Comput. vol.C-60, no.1, P. 80-92.
4. Li M., Shu J. (2011) C-codes: Cyclic lower-density MDS array codes constructed using starters for RAID-6, IBM New York USA. Res. Report RC25218.
5. Li J., Tang X., Tian C. (2019) A Generic Transformation to Enable Optimal Repair in MDS Codes for Distribution Storage System, IEEE Trans. Inf. Theory, vol.65, no.9, P.6257-6267.
6. Rhea S. (2011) Maintenance-free global data storage, IEEE Internet Comput. vol. 5, no.5, P. 40-49.
7. Koliada K.,V., Bardis N.G., Markovskiy O.P. (2019) Usage of linear erasure codes for increasing reliability and efficiency of information delivery on the internet, International Journal of Circuits, Systems and Signal Processing, vol.13, P. 585-592.
8. Rashmi K.V., B. Nihar , K. Ramchandran K. (2018) Information-Theoretically Secure Erasure Codes for Distrubute Storage, IEEE Trans. Inf. Theory, vol.64, no.3, P. 1621-1645.
9. Kai X., Zhu S., Li P. (2015) A Construction of New MDS Symbol-Pair Codes, IEEE Trans. Inf. Theory, vol.61, no.11, P. 5828-5834.
10. Yehezkeally Y. and Schwartz M. (2020) Reconstruction Codes for DNA Sequences With Uniform Tandem-Duplication Errors, IEEE Trans. Inf. Theory, vol.66, no.5, P. 2658-2669.