УДК 004.052.042

МАРКОВСЬКИЙ О.П.,
ІВАНОВ Д.Г.,
ВАНЧУГОВ Б.Ю.

ОРГАНІЗАЦІЯ РЕЗЕРВУВАННЯ ТА ВІДНОВЛЕННЯ ДАНИХ ПРИ ЇХ ВІДДАЛЕНОМУ ЗБЕРІГАННІ

В статье предложен метод резервирования и восстановления данных пользователя, хранящихся на нескольких удаленных узлах. Разработанный метод позволяет восстанавливать данные в наиболее часто встречающихся на практике ситуациях: при потере доступа к одному из узлов хранения или любым двум носителям. Детально описаны математическая идея метода и процедуры восстановления данных из узла хранения, доступ к которому утрачен. Разработанная процедура восстановления иллюстрируется примерами. Приведены теоретические и экспериментальные оценки эффективности предложенного метода.

In paper the method of reservation and recovering of user data stored on some remote memory units is proposed. Developed method makes it possible to recover data for frequent occurrence in practice situation: in case of access loosing for one of remote storage unit or for any two storage devices. The mathematical idea of proposed method and procedure for recovering of data from access lost storage unit are described in details. A numerical example for developed recovering procedure are given.

Вступ

Початок другого десятиліття XXI-го століття знаменує новий виток спіралі розвитку комп'ютерної обробки даних, а саме: використання технологій віддаленого надання обчислювальних і програмних ресурсів. На початку 80-х років минулого століття за зміну обчислювальним центрам з концентрованими комп'ютерними ресурсами прийшли доступні і максимально наближені до користувачів персональні комп'ютери. Нині, випередження темпів збільшення складності задач користувачів в порівнянні зі зростанням потужності персональних комп'ютерів, а також динамічний прогрес засобів телекомунікацій та мережових технологій, стимулюють відродження концентрації обчислювальних ресурсів з віддаленим доступом до них користувачів.

Технології віддаленого надання користувачеві ресурсів на комерційній основі з використанням Інтернету отримали назву "хмарних обчислень". Поняття ресурсу в цьому контексті включає власне обчислювальні ресурси, програмне забезпечення, а також ресурси пам'яті. Саме віддалене надання ресурсів зовнішньої пам'яті на сьогоднішній день набуло найбільшого поширення [1].

Вузловою проблемою ефективності віддаленого надання ресурсів пам'яті є забезпечення надійного доступу до даних користувача. Причинами втрати доступу до даних можуть бути: фізичне пошкодження носіїв, втрата інформації на них в результаті дії вірусів або некоректних дій програмного забезпечення, тимчасовий вихід з ладу обладнання або перевантаження вузла зберігання даних, втрата доступу вузла до мережі, комерційні чи природні катаклізми. Швидке збільшення кількості користувачів систем віддаленого зберігання інформації загострює проблему ефективного доступу до неї.

Для забезпечення надійного та оперативного доступу користувача до його інформації, що зберігається на віддалених носіях потрібні спеціальні механізми резервування та відновлення даних, доступ до яких постійно чи тимчасово втрачено.

Таким чином, наукова задача створення ефективних методів та засобів резервування і відновлення даних, що зберігаються на віддалених носіях є актуальною та важливою з огляду на особливості сучасного етапу розвитку інформаційних технологій.

Аналіз технологій резервування даних в системах їх віддаленого зберігання

При віддаленому зберіганні даних користувача вони розподіляються по окремим вузлам зберігання інформації. В рамках окремого вузла організовано розподілення даних по носіях, доступ користувачів до даних, їх захист, резервування в разі втрати доступу до одного або декількох носіїв.

Для забезпечення високого рівня неперервності доступу кожному користувачеві до своєї інформації, що зберігається на віддалених носіях найчастіше використовується їх резервування.

В якості критеріїв ефективності систем резервування найчастіше виступають:
- обчислювальна складність процедури відновлення даних з використанням резервних носіїв;
- кількість носіїв ($К_В$), дані з яких можуть бути відновлені в разі втрати доступу до них;
- відношення кількості резервних носіїв ($К_Р$) до числа носіїв, дані з яких можуть бути відновлені ($\alpha=К_Р/К_В$).

Проведений аналіз літературних джерел [1-3] показав, що найчастіше причинами втрати доступу до даних віддалених користувачів стає вихід з ладу окремих носіїв (або стирання інформації на них), є а також тимчасова нездатність вузлу обслужити запит користувача.

При втраті даних на окремих носіях в результаті чи то виходу їх із ладу, або стирання (помилкового чи цілеспрямованого) даних на них, кількість q таких носіїв залежить від часу t, що пройшов від моменту останнього звертання до них. Якщо вважати, що користувач зберігає свої дані на М носіях і залежність ймовірності $P(t)$ втрати даних до них від часу підпорядкована експоненційному закону, то ймовірність того, що до q з них буде втрачено доступ, визначається наступною формулою:

$$P(t) = C_M^q \cdot (1 - e^{-\lambda \cdot t})^q \cdot e^{-(M-q)\cdot\lambda\cdot t}, \qquad (1)$$

де λ- інтенсивність втрати даних на одному носієві в результаті його виходу з ладу чи стирання.

З формули (1) та аналізу статистики виходу з ладу носіїв [4] випливає, що ймовірність втрати доступу до одного носія доволі невелика, на 2-3 порядки меншою є ймовірність виходу з ладу двох носіїв. Тому, можна вважати, що реально, за умови, коли час між зверненнями не перевищує року, кількість носіїв, доступ до яких втрачено в результаті виходу їх з ладу або стирання даних не перевищує 2-х. Більшість реальних систем резервування даних розраховано саме на таку кількість носіїв, доступ до яких може бути втрачено [4].

Тимчасова нездатність вузлу обслужити запит користувача здебільшого пов'язана з перевантаженістю вузла зберігання даних, виходом з ладу його апаратних чи програмних компонентів, економічними чи природними катаклізмами.

Проблема забезпечення надійного доступу до даних, що містяться на віддалених від користувачів вузлах зберігання інформації спонукала до створення ряду технологій резервування.

Найбільш простою схемою резервування є використання простого дублювання даних на двох носіях. До такого типу відносяться системи Intermemory [2] та RAID-1 [3]. Використання простого дублювання пов'язане зі значними затратами об'єму пам'яті. При цьому, воно не гарантує відновлення даних при втраті доступу до обох носіїв, на яких зберігаються копії даних.

Значно меншого об'єму пам'яті потребує схема резервування, що передбачає для групи носіїв використання одного контрольного, на якому зберігається сума за модулем 2 відповідних даних всіх носіїв групи. Ця схема дозволяє доволі просто відновити дані при втраті доступу до одного з носіїв групи. Найбільш відомим застосуванням описаної схеми резервування є система RAID-1 [3]. Проте ця схема не дозволяє відновлювати дані при втраті доступу до більш як одного носія.

Найбільшого поширення на практиці набули технології відновлення даних на основі корегуючих та erasure кодів [4]. При відновленні даних з носіїв, до яких втрачено доступ, як правило, не має потреби в їх локалізації. Класичні корегуючі коди, такі, як коди Хемінга, БЧХ, Ріда-Соломона орієнтовані на послідовне виконання двох процедур: локалізації спотвореної частини даних та їх виправлення. З цієї причини при використанні загаданих вище класичних корегуючих кодів для відновлення даних з носіїв, до яких втрачено доступ, потрібна їх модифікація. Модифіковані коди Ріда-Соломона, зокрема, використовуються в системі відновлення даних з носіїв RAID-6 [3].

Більш ефективно використання для цієї цілі спеціальних erasure кодів. Більшість таких кодів [4] мають за основу лінійні перетворення, і це зумовлює швидке зростання кількості резервних носіїв при збільшенні числа носіїв до яких втрачено доступ.

Загальною рисою відомих технологій відновлення даних з носіїв, до яких втрачено доступ є те, що вони реалізовані в рамках окремого вузла зберігання інформації. Це означає, що в разі втрати доступу до вузла в результаті тимчасового виходу його з ладу, перевантаження, вірусної атаки, відключення від мережі, техногенних або природних катаклізмів, відомі механізми відновлення даних або доступу до них для конкретного користувача не спрацьовують.

Таким чином, існуючі методи відновлення доступу до даних в системах їх віддаленого зберігання не гарантують вирішення цієї задачі в разі втрати доступу до вузла зберігання інформації.

Ціллю досліджень є розробка та дослідження методу відновлення даних з віддалених від користувача носіїв для найбільш поширених на практиці ситуаціях, включаючи втрату доступу до вузла зберігання інформації.

Метод відновлення даних з носіїв вузла, до якого втрачено доступ

Проведений аналіз показав, що існуючі системи резервування інформації на носіях, які реалізуються на рівні вузлів зберігання не забезпечують повною мірою надійності та оперативності доступу до даних конкретного користувача - фактичного їх володаря. Тільки останній може у відповідності до цінності для нього даних, особливостей їх використання та вимог щодо оперативності доступу до них визначати політику резервування інформації на віддалених носіях. Це диктує доцільність застосування паралельно з існуючими системами резервування засобів забезпечення надійності та оперативності доступу до даних конкретного користувача.

З позицій користувача найбільш частими причинами відмов у доступі до віддаленої інформації є тимчасова втрата взаємодії з вузлом зберігання та втрата доступу до одного та рідше до двох носіїв.

Нехай інформація користувача, що зберігається на s носіях, розміщена на n вузлах зберігання інформації (надалі ВЗІ або просто вузол). При цьому на кожному з вузлів задіяно для даних конкретного користувача однакова кількість носіїв - $m=s/n$. Дані, що розміщено на j-тому носієві, $j\in\{1,..,m\}$, i-го вузла, $i\in\{1,..,n\}$, позначаються як a_{ij}.

Система віддаленого зберігання даних дає можливість швидкого доступу до будь-якої користувацької інформації. Як зазначалося вище, на практиці для користувача причинами відмові в доступі найчастіше є дві ситуації:

1. Один з вузлів тимчасово не доступний.
2. Не доступними є один або, що щонайбільше, два носія різних вузлів.

Отже, постає необхідність одночасного вирішення 2-ох задач:

1. Відновлення даних будь-якого з M носіїв одного, наприклад, k-ого вузла: $a_{k1},a_{k2},...,a_{km}$, де $k\in\{1..n\}$.

2. Відновлення даних будь-яких 2-х носіїв a_{qe} і a_{gr}, де $q,g\in\{1,..,n\}$ і $e,r\in\{1,..,m\}$.

Для вирішення вказаних задач пропонується використати m+2 додаткових носіїв, які розділяються на дві групи. Перша група складається з m додаткових носіїв, дані на яких $b_1,b_2,..,b_m$ формуються як суми за модулем два даних користувача з однойменних носіїв всіх вузлів на яких зберігається його інформація:

$$\forall j=1,...,m: b_j=\bigoplus_{i=1}^{n} a_{ij} \qquad (2)$$

Друга група з двох додаткових носіїв, на яких формуються і зберігається сума c за модулем 2 поліноміальних добутків даних всіх можливих пар однойменних носіїв на яких записані дані користувача:

$$c=\bigoplus_{j=1}^{m}\ \bigoplus_{\substack{i=1,..,n-1\\ k=i+1,..,n}} (a_{ij}\otimes a_{kj}), \qquad (3)$$

де символом '⊗' позначено операцію поліноміального множення. В силу того, що розрядність поліноміального добутку вдвічі більша за розрядність множників, то для зберігання суми c добутків використовується два додаткових носія. Практично, код c являє собою набір сум поліноміальних добутків фрагментів, на які розбиваються дані.

Додаткові носії для зберігання даних $b_1,b_2,..,b_m$ і коду c можуть бути розміщені як на окремому вузлі, так і безпосередньо у користувача.

При втраті користувачем доступу до k-го вузла, $k\in\{1,..,n\}$ дані $a_{k1},a_{k2},...,a_{km}$ з його носіїв відновлюються у відповідності з наступним виразом:

$$\forall j\in\{1,..,m\}: a_{kj}=b_j\oplus\bigoplus_{l=1,l\neq k}^{n} a_{lj}. \qquad (4)$$

При втраті доступу до двох довільних носіїв, на яких розміщено інформацію користувача, постає задача відновлення даних a_{qe} і a_{gr} і тут можна розглядати три випадки:

- втрачено доступ до двох носіїв, що знаходяться на одному вузлі, тобто $q=g$. Для цього випадку відновлення $a_{q,e}$ і a_{gr} реалізується за допомогою формули (3):

$$\begin{aligned} a_{qe}&=b_e\oplus\bigoplus_{l=1,l\neq q}^{n} a_{l,e},\\ a_{gr}&=b_r\oplus\bigoplus_{l=1,l\neq g}^{n} a_{lr} \end{aligned} \qquad (5)$$

- втрачено доступ до двох носіїв на різних вузлах, тобто $q\neq g$ і при цьому номери e і r носіїв зайнятих даними користувача на різних вузлах також різні: $e\neq r$. Для цього випадку відновлення $a_{q,e}$ і a_{gr} також реалізується за формулою (4).

- втрачено доступ до двох носіїв на різних вузлах, тобто $q \neq g$ але при цьому номери e і r носіїв зайнятих даними користувача на різних вузлах однакові: $e=r$. У цьому випадку відновлення a_{qr} і a_{gr} відбувається за дещо складнішою процедурою. Спочатку визначається значення суми за модулем 2 вказаних кодів a_{qr} і a_{gr} :

$$\delta = a_{qr} \oplus a_{gr} = b_r \oplus \bigoplus_{l=1, l\neq q, l\neq g}^{n} a_{lr} \qquad (6)$$

Значення поліноміального добутку η кодів a_{qr} і a_{gr} : $\eta = a_{qr} \otimes a_{gr}$ визначається наступним чином. Сума c за модулем 2 поліноміальних добутків даних всіх можливих пар одноіменних носіїв на яких записані дані користувача може бути розділена на дві компоненти, з яких перша γ не залежить від a_{qr} і a_{gr}, а друга - залежить від них:

$$\begin{aligned} c &= \bigoplus_{j=1}^{m} \bigoplus_{\substack{i=1,..,n-1 \\ k=i+1,..,n}} (a_{ij} \otimes a_{kj}) = \\ &= \bigoplus_{j=1, j\neq r}^{m} \bigoplus_{\substack{i=1,..,n-1 \\ k=i+1,..,n}} (a_{ij} \otimes a_{kj}) \oplus \bigoplus_{\substack{i=1,..,n-1 \\ k=i+1,..,n}} (a_{ir} \otimes a_{kr}) \end{aligned} \qquad (7)$$

Очевидно, що перша компонента γ суми (7) має вигляд:

$$\gamma = \bigoplus_{j=1, j\neq r}^{m} \bigoplus_{\substack{i=1,..,n-1 \\ k=i+1,..,n}} (a_{ij} \otimes a_{kj}) \qquad (8)$$

Тоді сума за модулем 2 (7) може бути представлена у наступному вигляді:

$$\begin{aligned} c &= \gamma \oplus \bigoplus_{\substack{i=1,..,n-1 \\ k=i+1,..,n}} (a_{ir} \otimes a_{kr}) = \\ &= \gamma \oplus (a_{qr} \oplus a_{gr}) \otimes \bigoplus_{i=1, i\neq q, i\neq g}^{n} a_{ir} \oplus (a_{qr} \otimes a_{gr}) = \\ &\gamma \oplus \delta \otimes \bigoplus_{i=1, i\neq q, i\neq g}^{n} a_{ir} \oplus (a_{qr} \otimes a_{gr}) \end{aligned} \qquad (9)$$

Відповідно, чисельне значення γ визначається за наступною формулою:

$$\eta = a_{qr} \otimes a_{gr} = c \oplus \gamma \oplus \delta \otimes \bigoplus_{i=1, i\neq q, i\neq g}^{n} a_{ir} \qquad (10)$$

Таким чином, значення кодів a_{qr} і a_{gr} з пари носіїв, до яких втрачено доступ може бути віднайдено як розв'язання системи рівнянь:

$$\begin{cases} a_{qr} \oplus a_{gr} = \delta \\ a_{qr} \otimes a_{gr} = \eta \end{cases} \qquad (11)$$

Кожен з кодів a_{qr} і a_{gr} має певну розрядність w, яка вимірюється мільйонами бітів. При обчисленні поліноміального добутку коди a_{qr} і a_{gr} розділяються на d-розрядні фрагменти, так, що код η добутку складається з $(2 \cdot d\text{-}1)$-розрядних фрагментів.

В силу симетричності операцій додавання за модулем 2 та поліноміального множення система рівнянь (11) має два розв'язки.

Можна показати, що для кожного з d-розрядних фрагментів система (11) може бути зведена до системи з $2 \cdot d$-1 лінійних рівнянь. Наприклад, якщо d=4, і біти фрагменту a_{qr} позначаються як x_1, x_2, x_3, x_4, біти фрагменту a_{gr} позначаються як y_1, y_2, y_3, y_4, біти фрагменту δ позначаються як $\beta_1, \beta_2, \beta_3, \beta_3$, біти η позначаються як $\rho_1, \rho_2, ..., \rho_7$, то друге рівняння системи (11) може бути представлено у вигляді наступних бітових рівнянь:

$$\begin{cases} x_1 \oplus y_1 = \beta_1 \\ x_2 \oplus y_2 = \beta_2 \\ x_3 \oplus y_3 = \beta_3 \\ x_4 \oplus y_4 = \beta_4 \\ x_1 \cdot y_1 = \rho_1 \\ x_2 \cdot y_1 \oplus x_1 \cdot y_2 = \rho_2 \\ x_3 \cdot y_1 \oplus x_2 \cdot y_2 \oplus x_1 \cdot y_3 = \rho_3 \\ x_4 \cdot y_1 \oplus x_3 \cdot y_2 \oplus x_2 \cdot y_3 \oplus x_1 \cdot y_4 = \rho_4 \\ x_4 \cdot y_2 \oplus x_3 \cdot y_3 \oplus x_2 \cdot y_4 = \rho_5 \\ x_4 \cdot y_3 \oplus x_3 \cdot y_4 = \rho_6 \\ x_4 \cdot y_4 = \rho_7 \end{cases} \qquad (12)$$

Підстановкою виразів для y_1, y_2, y_3, y_4 з перших 4-х рівнянь: $y_1 = \beta_1 \oplus x_1, ..., y_4 = \beta_4 \oplus x_4$ системи (11) в наступні 7 рівнянь цієї системи отримується наступна система лінійних бітових рівнянь:

$$\begin{cases} x_1 \oplus y_1 = \beta_1 \\ x_2 \oplus y_2 = \beta_2 \\ x_3 \oplus y_3 = \beta_3 \\ x_4 \oplus y_4 = \beta_4 \\ x_1 \cdot \beta_1 \oplus x_1 = \rho_1 \\ x_2 \cdot \beta_1 \oplus x_1 \cdot \beta_2 = \rho_2 \\ x_3 \cdot \beta_1 \oplus x_2 \cdot \beta_2 \oplus x_2 \oplus x_1 \cdot \beta_3 = \rho_3 \\ x_4 \cdot \beta_1 \oplus x_3 \cdot \beta_2 \oplus x_2 \cdot \beta_3 \oplus x_1 \cdot \beta_4 = \rho_4 \\ x_4 \cdot \beta_2 \oplus x_3 \cdot \beta_3 \oplus x_3 \oplus x_2 \cdot \beta_4 = \rho_5 \\ x_4 \cdot \beta_3 \oplus x_3 \cdot \beta_4 = \rho_6 \\ x_4 \cdot \beta_4 \oplus x_4 = \rho_7 \end{cases} . \qquad (13)$$

Очевидно, що система (12) може бути доволі просто розв'язана і результаті можуть бути отримані бітові значення x_1, x_2, x_3, x_4, поточного фрагменту a_{qr} і біти як y_1, y_2, y_3, y_4 одноіменного фрагменту a_{gr}.

Таким чином доведено, що запропонований метод забезпечує відновлення даних при втраті доступу до будь-яких двох носіїв, на яких розміщено дані користувача.

Оцінка ефективності

Нескладно показати, що обчислювальна складність розв'язання системи лінійних рівнянь (13) визначається як $O(2\cdot d^2+d)$ так, що обчислювальна складність розв'язання системи (11) становить $O(2\cdot w\cdot d+w)$. В перших двох випадках обчислювальна складність відновлення даних з двох носіїв визначається складністю операції додавання за модулем 2: $O(2\cdot w/d)$. Враховуючи, що ймовірність третього випадку при відновленні даних з двох носіїв приблизно оцінюється як n^{-1}, то, в середньому, обчислювальна складність операцій відновлення даних з двох носіїв при використанні запропонованого методу визначається як $O(2\cdot w\cdot(n+d^2)/(n\cdot d)) \approx O(2\cdot w/d)$. Обчислювальна складність відновлення даних з носіїв вузла, доступ до якого втрачено визначається як $O(m\cdot w/d)$.

На практиці, час відновлення залежить не скільки від обчислювальної складності операцій реконструкції даних, стільки від кількості носіїв, до яких треба звернутися, щоб отримати дані, потрібні для відновлення втраченої інформації.

Кількість додаткових носіїв для зберігання резервних кодів $b_1,b_2,..\ ,b_m$ дорівнює m, а для збереження коду c - два носії, вважаючи на те, що розрядність поліноміального добутку вдвічі перевищує довжину множників. Таким чином, кількість h резервних носіїв, що використовуються у запропонованому методі визначається формулою:

$$h = m + 2 \qquad (14)$$

Як зазначалося вище, практично всі відомі засоби відновлення доступу до віддалених даних користувача реалізуються в рамках окремого вузла зберігання даних. Це означає, що резервування відбувається на рівні самого вузла і, відповідно, відомі засоби практично не забезпечують вирішення вказаної задачі в разі втрати доступу користувача до вузла або тимчасовій непрацездатності самого вузла. Разом з тим, статистика відмов в доступі до даних, що віддалено зберігаються, свідчить про те, що такий тип втрати доступу зустрічається доволі часто. Реалізація запропонованого методу резервування та відновлення доступу до даних орієнтована на рівні користувача і передбачає, що зберігання основних і резервних його даних відбувається на різних вузлах. Відповідно, викладений метод забезпечує ефективне відновлення для користувача доступу даних, які зберігаються на тимчасово недоступному віддаленому вузлі.

Крім того, запропонований метод дозволяє ефективно для користувача вирішувати задачу забезпечення неперервності доступу до своїх даних в разі втрати доступу до двох довільних носіїв, локалізованих як на одному вузлі зберігання даних, так і на різних вузлах.

Таким чином, запропонований метод дозволяє ефективно вирішувати задачу резервування і відновлення даних користувача для найбільш поширених на практиці випадків втрати доступу до них.

Задача відновлення даних при втраті доступу до одного вузла та двох довільних носіїв може бути вирішена з використанням відомих корегуючих кодів, таких, зокрема як коди Ріда-Соломона, циклічні та БЧХ коди, а також erasures коди. При цьому, для вирішення, за допомогою зазначених кодів, задачі відновлення даних з носіїв одного вузла потребує $2\cdot m$ додаткових носіїв. При цьому зазначені вище корегуючі коди забезпечують відновлення даних і при втраті інформації з двох довільних носіїв. Таким чином, загальна кількість носіїв для резервування даних для їх відновлення при втраті доступу до носіїв одного вузла зберігання або двох довільних носіїв становить $2\cdot m$, тобто менше в порівнянні з оцінкою (13) для запропонованого методу.

Процедура відновлення даних з використанням корегуючих кодів передбачає розв'язання систем лінійних рівнянь. Для відновлення інформації з носіїв одного вузла потрібно розв'язувати систему з m лінійних рівнянь. Обчислювальна складність розв'язання такої системи становить $O(2\cdot m^2)$ для одного d-розрядного фрагменту. Обчислювальна складність відновлення даних з носіїв вузла, доступ до якого втрачено, визначається як $O(m^2\cdot w/d)$. Тобто, використання запропонованого методу дозволяє зменшити обчислювальну складність приблизно в m раз.

При вирішенні задачі відновлення даних з двох довільних носіїв за допомогою корегуючих кодів, розв'язується система з 4-х рівнянь, обчислювальна складність цієї процедури становить $O(8\cdot w/d)$. Порівняння з наведеним вище аналогічним показником для запропонованого методу доводить, що його застосування дозво-

ляє зменшити обчислювальну складність приблизно в 4 рази.

В відомих корегуючих кодах, одні і ті ж самі механізми використовуються для відновлення даних при всіх варіантах втрати доступу до носіїв. Відповідно, для вирішення різних задач витрачаються однакові обчислювальні ресурси. В запропонованому методі передбачена більш гнучка процедура відновлення даних, яка дозволяє витрачати різні за об'ємом обчислювальні ресурси в різних випадках втрати доступу до віддалених носіїв.

Таким чином, запропонований метод відновлення даних при втраті доступу до носіїв одного вузла зберігання даних або двох довільних носіїв забезпечує зменшення числа резервних носіїв та обчислювальної складності процедури відновлення в порівнянні з корегуючими кодами. Досягнутий ефект забезпечується за рахунок вузької спеціалізації запропонованого методу на найбільш поширених на практиці ситуаціях втрати доступу користувача до даних, що віддалено зберігаються.

Важливим і принципово новим моментом є те, що в рамках запропонованого методу з'являється можливість відновлення інформації цілого вузла за рахунок використання резервних даних користувачів, що користуються цим вузлом для віддаленого зберігання своїх даних. Це значно підвищує живучість систем віддаленого зберігання інформації.

Висновки

Таким чином, в роботі запропоновано метод організації резервування та відновлення даних при їх зберіганні на віддалених від користувача носіях з урахуванням найбільш поширених на практиці ситуацій: втраті доступу до окремого вузла зберігання або не більше ніж двох віддалених носіїв.

Дослідження запропонованого методу забезпечення неперервності доступу користувача до його віддалених даних довели, що за рахунок спеціалізації та використання більш гнучких процедур відновлення даних, він забезпечує більшу ефективність резервування в порівняння з відомими методами, корегуючими та ensure кодами.

Розроблений метод може бути ефективно використано в перспективних "хмарних" технологіях віддаленого зберігання даних.

Список літератури

1. Иванов Д. Г. Организация резервирования в системах распределенного хранения данных // Вісник Національного технічного університету України "КПІ" Інформатика, управління та обчислювальна техніка, – Київ: ВЕК+ – 2012 – № 56. с.160-164.
2. Луцкий Г.М., Иванов Д.Г. Метод восстановления данных на основе скошенных матриц в системах распределенного хранения // Известия высших учебных заведений. Проблемы полиграфии и издательского дела. Информационные технологии.- М.:УПИПК МГУП им. И.Федорова.- 2013.- № 2. - С.47-52
3. Heuert U., Ivanov D. Paladin: Secure and redundant cloud storage // Herald of the Merseburg University of Applied Sciences. - Merseburg: Elbe Drucketei Wittenberg GmbH.- 2011.-№ 8.-S.220-228.
4. Plank J. S., Thomasson M.G., On Practical Use of LDPC Erasure Codes for Distribute Storage Application: Technical Report UT-CS-03-510.- Department of Computer Science, University of Tennessee.-2004.