

*АНТОНЮК А. І.,
ЦИМБАЛ І. В.*

ВИЗНАЧЕННЯ ПЕРІОДУ ЗАПАМ'ЯТОВУВАННЯ СТАНУ ЦЕНТРАЛІЗОВАНОГО ІНФОРМАЦІЙНОГО ФОНДУ

В роботі пропонується вирішення актуальної задачі визначення періоду запам'ятовування стану серверів, призначених для розташування централізованого інформаційного фонду невеличкої установи, з метою підвищення рівня надійності та готовності фонду до роботи користувачів. Встановлена функціональна залежність періоду запам'ятовування від кількості створених копій, параметрів стану централізованого інформаційного фонду та рівня надійності функціонування засобів, які використовуються для створення копій.

Solution of actual task of determination of period of memorizing of the state of the servers intended for the location of the centralized informative fund of small establishment is in-process offered, with the purpose of increase of level of reliability and readiness of fund to work of users. Set functional dependence of period of memorizing from the quantity of the created copies, parameters of the state of the centralized informative fund and level of reliability functioning of facilities, that is used for creation of copies.

Вступ

Одним з основних способів забезпечення збереження інформації в теперішній час є резервне копіювання даних, сутність якого полягає в створенні копій даних, призначених для відновлення вказаних даних в оригінальному або новому місці їх розташування у разі їх ушкодження або руйнування [1]. Алгоритми та стратегії створення резервних копій різноманітні і постійно удосконалюються [2-6]. Рішення, які здебільшого приймаються, відносяться до великих організацій, перед якими на першому місці стоїть завдання економії місця зберігання і автоматизації процесу створення резервних копій.

Метою цієї роботи є пошук простого рішення для невеликої організації, де працюють, як правило, один-два програмісти в умовах обмеженого фінансування та постійного дефіциту часу. Остання обставина примушує їх постійно вибирати, яка робота більш пріоритетна, і виконувати в першу чергу більш пріоритетну роботу. У цих умовах, роботи із створення резервних копій, як менш пріоритетні, постійно відкладаються на потім, що може призвести у результаті до втрати актуальності існуючих резервних копій, і, у разі руйнування інформації, відновлення актуального стану інформації стане неможливим.

Головними чинниками, які призводять до руйнування інформації, є: логічні помилки розробників комп'ютерних програм; людський чинник – помилки користувачів або адміністраторів

комп'ютерної мережі; можливі "вірусні" або "хакерські" атаки; ненадійна робота апаратури.

Додатковий чинник, який можливо є головним в наш час, – це нестабільне електропостачання. Припустимо, що будівля, в якій розміщується установа, була побудована більше 30 років тому. Це означає, що електропроводка не розрахована на сучасне навантаження і часто виходить з ладу, що у свою чергу, може призвести до виходу з ладу апаратури і до руйнування інформації. Деякі автори стверджують, що 95% установ стикалися з втратою даних через відключення електрики [7].

Ще одна особливість – електропостачання може бути перервано і в день, під час роботи установи, і в вечері або в ночі, коли працюють лише сервери. Ця обставина обмежує застосування автоматичного способу копіювання, бо залишений на ніч процес резервного копіювання може бути несподівано перервано на такий великий час, що навіть джерело безперервного живлення не спасе від руйнування інформації. Таким чином, адміністратор має бути присутнім щоб своєчасно прийняти оперативне рішення. Наприклад, іноді достатньо своєчасно увімкнути напругу за допомогою автоматичного вимикача. А для цього адміністратор має знати період копіювання заздалегідь.

Два головних чинника, які дозволяють керувати процесом запам'ятовування, це – період копіювання та кількість створених копій.

Кінцевою метою роботи в цих умовах є проста, бажано арифметична, формула, яка дозволить швидко вирахувати період копіювання в залежності від інтенсивностей руйнування та відновлення інформації, а також від кількості копій та заданої ймовірності руйнування всіх копій.

Постановка задачі

Припустимо, є деяка установа з кількістю робочих місць не більше 100. В установі є централізований інформаційний фонд (ЦІФ), який розміщується на 4-х серверах:

1. "спільне поле пам'яті" – сервер, який призначений для оперативного обміну інформацією між користувачами установи;

2. "електронний архів" – сервер, який призначений для постійного зберігання тих документів, які надходять до установи, які відправляються з установи та інших офіційних документів установи;

3. "сайт установи" – сервер, який призначений для функціонування офіційного сайту установи;

4. "міні АТС" – сервер, який призначений для організації телефонного зв'язку в установі.

Цей набір серверів умовний, головне полягає в тому, що кожен сервер є часткою ЦІФ і їх стан необхідно періодично копіювати, при цьому інтенсивність оновлення ЦІФ для кожного серверу своя і, відповідно, період копіювання для кожного сервера свій.

При цьому під терміном стан розуміється не лише стан інформації користувачів установи, а і стан операційної системи та інших програмних засобів кожного сервера, бо оновлення програм є також процес безперервний. Таким чином окрема резервна копія являє собою образ стану окремого сервера.

Копіювання здійснюється на зовнішні накопичувачі на магнітних дисках (НМД). Припустимо, що для кожного сервера для резервного копіювання виділено по одному зовнішньому НМД, де зберігаються декілька поколінь копій конкретного сервера.

Такий технологічний прийом дозволяє при руйнуванні інформації відновити процес функціонування ЦІФ з моменту останнього запам'ятовування, або, у випадку зруйнованої останньої копії, з останньої справної копії. Додатково для відновлення ЦІФ в актуальний стан необхідно повторити усі роботи, пов'язані зі змі

ною стану фонду, з моменту запам'ятовування. При цьому адміністратор ЦІФ повинен мати можливість заздалегідь оцінити період копіювання, який забезпечить йому мінімальні витрати часу на створення резервних копій в умовах дефіциту часу і необхідності збереження актуального стану інформації.

Реалізація процесів запам'ятовування та відновлення вимагає додаткових витрат ресурсів обчислювальної системи. При цьому виникає дві протилежні вимоги. З одного боку необхідно мінімізувати втрати ресурсів на відновлення ЦІФ в актуальний стан і, з цієї точки зору, запам'ятовування доцільно робити якомога частіше (крайня межа – після кожної зміни стану фонду). З іншого боку, на час запам'ятовування закривається доступ користувачів до ЦІФ, що знижує коефіцієнт готовності фонду до роботи користувачів; а також додаткові витрати ресурсів на запам'ятовування знижують, в цілому, продуктивність обчислювальної системи. З цієї точки зору запам'ятовування доцільно робити якомога рідше.

Тому необхідно визначити такий період запам'ятовування поточного стану ЦІФ, при якому забезпечувався б необхідний час відновлення і мінімальне пониження коефіцієнта готовності фонду до роботи і продуктивності обчислювальної системи.

Рішення задачі синтезу стратегії управління процесом запам'ятовування актуального стану інформаційних фондів допускає два підходи:

У першому випадку період запам'ятовування (копіювання) актуального стану інформаційних масивів припускається відомим та визначається кількістю копій або поколінь копій, необхідна для забезпечення заданого рівня надійності зберігання інформації [8-10];

У другому випадку період запам'ятовування актуального стану інформаційних масивів є шуканою величиною, а кількість копій або поколінь копій величинами початковими [11].

Централізовані інформаційні бази даних включають десятки інформаційних масивів. У цих умовах створення великої кількості копій обмежується невисоким рівнем фінансування та наявністю лише по одному НМД для кожного сервера. З метою скорочення кількості копій в даному випадку використовується другий підхід до рішення задачі синтезу стратегії управління процесом запам'ятовування актуального стану серверів.

Моделювання процесу копіювання

Задача про визначення оптимального періоду запам'ятовування в системах з випадковими відмовами двох типів (що руйнують та неруйнують інформацію) розглядалася в [8], де автором встановлена залежність, що визначає оптимальний період запам'ятовування t як функцію від інтенсивностей a – руйнівних відмов і b – неруйнівних відмов, середньої тривалості відновлення t_b – руйнівних відмов і t_d – неруйнівних відмов, тривалості відновлення $t_e(a)$ – неруйнівних відмов, що настали на інтервалі відновлення після руйнівних відмов і від середньої тривалості запам'ятовування t_g :

$$t = \frac{2}{m} \ln\left(1 + \sqrt{\frac{t_g}{h}}\right), \tag{1}$$

де

$$m = a + b - bt_e(a);$$

$$h = \frac{1+at_b+bt_d+abt_b t_d}{m}.$$

Постановка задачі прийнята в [12], відповідає процесу виконання на ПК окремого завдання та створенню контрольних точок під час виконання цього завдання. У цій роботі ставиться задача встановити залежність величини періоду запам'ятовування поточного стану ЦіФ від параметрів процесу його функціонування.

Настання неруйнівних відмов не впливає в цілому на функціонування ЦіФ, оскільки не вимагає відновлення фонду і впливає зрештою лише на тривалості виконання окремих етапів робіт, що не істотно для поставленої задачі. Тому інтенсивність неруйнівних відмов дорівнюється до нуля, і в цьому випадку залежність (1) набирає такого вигляду:

$$t = \frac{2}{a} \ln\left(1 + \sqrt{\frac{at_g}{1+at_b}}\right). \tag{2}$$

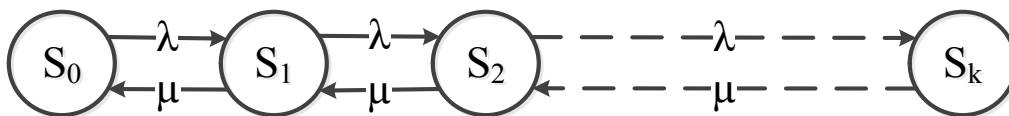


Рис. 1. Граф станів системи, яка запам'ятовує

- S_0 – усі покоління копій справні;
- S_1 – 1-е покоління справне, 0-е покоління зруйноване;
- S_2 – 2-е покоління справне, усі попередні покоління зруйновані;
- S_k – k-е покоління справне, усі попередні покоління зруйновані

Отже, для визначення оптимального періоду запам'ятовування поточного стану ЦіФ необхідно визначити середні значення часу запам'ятовування t_g і часу відновлення t_b .

У зв'язку з реально існуючою небезпекою руйнування інформації на зовнішньому НМД, обумовленої кінцевими значеннями параметрів надійності НМД, при запам'ятовуванні створюється n копій поточного стану(що називаються надалі поколінням копій). Витрати ресурсів комп'ютера на запам'ятовування стану ЦіФ пропорційні кількості створюваних копій, тобто:

$$t_g = ng, \tag{3}$$

де g – середній час створення однієї копії.

Цей факт є обмеженням для кількості копій в поколінні. Тому з метою підвищення надійності функціонування зберігаються k поколінь, що забезпечують відновлення утримуваного фонду при руйнуванні $(k-1)$ поколінь.

Покоління, що має хоч би одну копію що забезпечує відновлення ЦіФ, називається справним поколінням. Усі k поколінь, кожне з яких складається з n копій, утворюють систему запам'ятовування стану ЦіФ (ЗС).

Якщо припустити, що процеси руйнування і відновлення ЦіФ апроксимуються показовими законами розподілу вірогідності з відомими значеннями інтенсивностей, то систему запам'ятовування стану ЦіФ можна звести до системи, процес функціонування якої наведений на мал. 1.

Система, що тут запам'ятовує, розглядається як обслуговуючий канал, процес руйнування інформації окремого покоління ЦіФ – як процес надходження заявок на обслуговування, процес обслуговування заявок – як процес відновлення системи; тобто система, що запам'ятовує, розглядається як система масового обслуговування.

На вхід обслуговуючого каналу поступає простий потік заявок на відновлення з інтенсивністю

$$\lambda = \frac{\lambda_1}{n}$$

де λ_1 – інтенсивність руйнування окремої копії.

Число місць в черзі до каналу дорівнює k – кількості поколінь ЗС. Тривалість обслуговування заявок каналом апроксимується показовим законом розподілу вірогідності з інтенсивністю μ за правилом LIFO (останній прийшов – перший обслуговується). Інтенсивність μ визначається таким чином:

$$\mu = \frac{1}{t_v + t_c + t_g + t_l}; \quad (4)$$

$$t_v = \frac{vn}{2},$$

$$t_l = ln;$$

де

t_v – середній час відновлення ЦІФ з справнього покоління;

t_c – середній час відновлення ЦІФ з попереднього стану в наступний;

t_l – середній час перевірки щойно створеного покоління;

v – середній час відновлення ЦІФ з справної копії;

l – середній час перевірки щойно створеної копії.

Описана система є класичною системою масового обслуговування з кінцевою кількістю місць в черзі. Якщо стаціонарну вірогідність стану S_i позначити P_i , тоді наведений на рис. 1 граф станів описуватиметься такою системою рівнянь [13]:

$$P_0 = \frac{1 - \frac{\lambda}{\mu}}{1 - (\frac{\lambda}{\mu})^{k+1}}; \quad (5)$$

$$P_i = (\frac{\lambda}{\mu})^i P_0, \quad (6)$$

$$i = 1, 2, \dots, k.$$

P_k – це ймовірність руйнування k поколінь копій. У зв'язку з тим, що метою резервування є забезпечення безвідмовної роботи об'єкту загалом, в нашому випадку більш цікава ймовірність безвідмовної роботи k поколінь копій:

$$Q_0 = 1 - P_k. \quad (7)$$

Значення величини часу відновлення ЦІФ t_b залежить від стану ЗС і визначається виразом:

$$t_b = \sum_{i=0}^k P_i t_i, \quad (8)$$

де t_i – середній час відновлення ЦІФ в актуальний стан, за умови що ЗС знаходиться в стані S_i . Величина t_i визначається як сумарний час відновлення ЦІФ в усі проміжні стани, тобто:

$$t_i = i(t_v + t_c + t_g + t_l) + t_v + t_c. \quad (9)$$

Якщо у вираз (7) підставити значення t_i та P_i з виразів (5) і (8), то вийде наступна формула для визначення часу відновлення t_b :

$$t_b = \frac{P_0}{\mu} \sum_{i=1}^k i \left(\frac{\lambda}{\mu}\right)^i + t_v + t_c. \quad (10)$$

Результати обчислень за наведеними формулами (5)-(10) для конкретної системи показали, що для забезпечення ймовірності $Q_0 = 0,99999$ необхідно створювати не менш двох поколінь копій, по одній в кожному поколінні, кожні 7 діб. З метою забезпечення надійного зберігання інформації в [10] рекомендується створювати три покоління копій по три копії в кожному. Порівнюючи рекомендоване в [10] з результатами розрахунків можна зробити висновок що використання наведених формул може скоротити час на копіювання та відновлення ЦІФ.

Особистий досвід автора свідчить що частіше шукають незруйновану інформацію, яка була півроку або рік тому знищена власником за відсутністю потреби в ній. Якщо інформація не є внесеною в ЦІФ, то цей випадок не має відношення до задачі що розглядається і адміністратор обчислювальної системи не несе відповідальності за її збереження. Проте, враховуючи велику можливість таких випадків, є сенс зберігати покоління копій за півроку або рік, але то вже інша задача.

Висновки

В роботі вирішена задача визначення моментів часу створення резервних копій стану окремого централізованого інформаційного фонду. Формули (5)-(10) дозволяють визначити оптимальний період запам'ятовування t , кількість копій n та поколінь копій k , які необхідні для забезпечення Q_0 – заданого рівня ймовірності безвідмовної роботи ЦІФ.

Список посилань

1. Резервное_копирование [Электронный ресурс] /Материал из Википедии. – 2015. – Режим доступа:https://ru.wikipedia.org/wiki/Резервное_копирование.
2. Уваров А.С. Windows Server 2003. Резервное копирование [Электронный ресурс] / Технический блог специалистов ООО"Интерфейс". Записки IT специалиста. – 2015. – Режим доступа:http://interface31.ru/tech_it/2010/01/windows-server-2003-rezervnoe-kopirovanie.html.
3. Горловой А. Эволюция и перспективы резервного копирования данных [Электронный ресурс] / «Экспресс-Электроника». – 2003. – №06. – Режим доступа: <http://habrahabr.ru/company/veeam/blog/242983>.
4. Преимущества нового метода резервного копирования виртуальных машин перед классической схема [Электронный ресурс] /Блог компании «Veeam Software». – 2015. – Режим доступа:<http://citforum.ru/hardware/articles/backup>.
5. Пискунов И. Обоснование необходимости приобретения систем резервного копирования и восстановления данных [Электронный ресурс] / Аналитический центр Anti-Malware.ru. – 2015. – Режим доступа:<http://www.anti-malware.ru/node/14024>.
6. Что такое Cloud4u backup[Электронный ресурс] / Сайт компании Cloud4u. – 2015. – Режим доступа:http://www.cloud4u.ru/cloud-ervices/backup/?gclid=CKfM2I2ossMCFWTHtAod_iUApq.
7. Резервное копирование и восстановление данных [Электронный ресурс] /Компания РЕД-НОД. – 2015. – Режим доступа:<https://rednod.ru/backup>.
8. Кульба В.В. Анализ стратегий резервирования информационных массивов в АСУ. – В кн.:Сборник трудов Института проблем управления. – М.:1977, Вып.14. – С.20-32.
9. Кульба В.В., Пелихов В.П., Шелков А.Б. Стратегии резервирования информационных массивов. – В кн.: Сборник трудов Института проблем управления. – М.:1978, Вып.16. – С.26-42.
10. Каган Б.М., Мкртумян И.Б. Основы эксплуатации ЭВМ. – М.:Энергоиздат, 1983. – 376 с.
11. Балина Е.И., Бродецкий Г.Л. Эффективность организации запоминания промежуточных результатов при ненадежном запоминающем устройстве и ограничениях на его использование // Известия АН СССР. Техн. Кибернетика. – 1985. – №1. – С. 80-85.
12. Бродецкий Г.Л. Об одной задаче периодического запоминания результатов. – К.; Кибернетика. – 1974. – №5. – С. 70-74.
13. Клейнрок Л. Теория массового обслуживания. – М.: Машиностроение. – 1979. – 431 с.