

Вестник Евразийской науки / The Eurasian Scientific Journal <https://esj.today>

2018, №6, Том 10 / 2018, No 6, Vol 10 <https://esj.today/issue-6-2018.html>

URL статьи: <https://esj.today/PDF/05ITVN618.pdf>

Статья поступила в редакцию 25.10.2018; опубликована 14.12.2018

Ссылка для цитирования этой статьи:

Сизов В.А. Разработка моделей повышения эффективности сохранности данных в распределенной вычислительной среде на основе динамического резервирования данных // Вестник Евразийской науки, 2018 №6, <https://esj.today/PDF/05ITVN618.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

For citation:

Sizov V.A. (2018). Development of models for improving the efficiency of data security in a distributed computing environment based on dynamic data backup. *The Eurasian Scientific Journal*, [online] 6(10). Available at: <https://esj.today/PDF/05ITVN618.pdf> (in Russian)

УДК 004

Сизов Валерий Александрович

ФГБОУ ВО «Российский экономический университет имени Г.В. Плеханова», Москва, Россия

Профессор кафедры «Прикладной информатики и информационной безопасности»

Доктор технических наук, профессор

E-mail: sizovva@gmail.com

РИНЦ: https://elibrary.ru/author_profile.asp?id=772122

Разработка моделей повышения эффективности сохранности данных в распределенной вычислительной среде на основе динамического резервирования данных

Аннотация. В статье рассматривается задача повышения сохранности данных в распределенной вычислительной среде за счет комплексного использования методов резервирования и восстановления. Рассматривается распределенная вычислительная среда, построенная на основе вычислительной сети, которая используются в качестве организационно-технической формы управления информационными потоками в больших информационных системах. Для этого определяется понятие динамического резерва, в состав которого входят как сами данные, так их копии и/или предыстории. Формализована задача определения оптимального содержания динамического резерва и его размещения по узлам вычислительной сети. Разработаны модели динамического резервирования данных в распределенной вычислительной среде по критерию равномерного распределения выигрыша по узлам вычислительной сети. Обоснована оптимизационная модель синтеза динамического резерва для однородных кластеров вычислительных сетей. Для упрощения решения предложенной оптимизационной задачи, рассмотрена ее декомпозиция на множество задач «о ранце», для решения которых можно использовать известные алгоритмы. Использование динамического резерва позволяет повысить эффективность сохранности данных в распределенной вычислительной среде за счет комплексного учета требований пользователей и технических параметров вычислительной сети.

Ключевые слова: распределенная вычислительная среда; вычислительная сеть; резервирование данных; восстановление данных; копии; предыстории; моделирование; оптимизация

Введение

Создание новых информационных технологий, построенных по типу «вычислительного облака» (cloud computing) связано с широким внедрением вычислительных сетей (ВС), распределенных баз данных и современных систем передачи информации¹.

Это обусловлено качественными изменениями в структуре современных организационно-технических систем управления, широким внедрением различных типов персональных компьютеров, гаджетов с легким, в том числе мобильным подключением к глобальным ВС, интеллектуальных терминалов, резким снижением затрат на функционирование распределенных систем с одновременным повышением качества передачи информации, развитием промышленного Интернета вещей, а также с устойчивым ростом спроса на высокую оперативность и устойчивость обработки информации.

Распределенная вычислительная среда – «вычислительное облако» (PBC), построенная на основе ВС, в ряде случаев используются в качестве организационно-технической формы управления процессами, протекающими в реальном масштабе времени (управления информационными потоками в больших информационных системах, таких как, энергосистемы регионального уровня и выше, управление аэропортами, управление движением транспортных средств железных дорог, системами информатизации в интересах Министерства обороны и др.).

Такие системы предъявляют повышенные требования к оперативности обработки данных в реальных условиях функционирования, а также их сохранности в условиях действия дестабилизирующих факторов [1, 2]. Решение проблемы сохранности данных в таких системах является актуальным и требует учета ряда условий, таких как: параметры используемой системы распределённой обработки данных, жёсткие временные ограничения получения ответа на запрос при принятии решения, требования к качеству выходной информации и др.

Формализация процесса динамического резервирования данных в распределенной вычислительной среде

Под выходной информацией в PBC понимается информация, получаемая в результате выполнения функций системы распределённой обработки данных и выдаваемая на объект её деятельности, пользователям или в другие системы. Под качеством выходной информации понимается совокупность свойств информации, обуславливающих её пригодность удовлетворять потребности пользователей при решении стоящих перед ним задач [3].

Одним из основных системных методов повышения качества выходной информации в PBC, направленных на улучшение вероятностно-временных характеристик функционирования систем, являются резервирование и восстановление данных [4-6].

Особенности функционирования распределенной вычислительной среды позволяют при решении задач повышения качества выходной информации методами резервирования и восстановления, кроме традиционных средств введения избыточности, ввести и достаточно эффективно использовать избыточность, которую можно определить, как виртуально-восстановительную [7]. Она учитывает свойство достоверности информации не только как свойство безошибочности (т. е. соответствия данных, полученных потребителем, данным, сформированным у источника), но и как свойство актуальности (сохранения достаточной

¹ <http://nerohelp.info/1104-ct-r.html>.

степени соответствия реальному состоянию объективной реальности на момент использования данных об этом состоянии).

Суть динамического резервирования данных состоит в том, что в большинстве случаев задачи, решаемые территориально-распределенными органами управления, позволяют заранее выделить необходимые наборы данных (промежуточные массивы) и распределить их по узлам вычислительной сети для немедленного использования в будущем. В состав динамического резерва могут входить как сами данные, так их копии и/или предыстории.

Таким образом, в ряде случаев допускается

$$|X(t) - X(t - \tau)| \leq \delta, \delta > 0, \forall \tau, \tau \in [0, t], \quad (1)$$

где $X(t)$, $X(t - \tau)$ – соответствующие значения какого-либо информационного элемента в момент времени t и $(t - \tau)$, δ – абсолютная граница допустимых пределов отклонений значений информационных элементов. Выполнение неравенства (1) обуславливает актуальность информационного элемента на интервале $T = [0, t]$. При этом под степенью виртуальности резерва понимается вероятность неактуальности всех составляющих его элементов на интервале T .

Одной из основных задач динамического резервирования данных в ВС, решаемых на этапе предпроектного анализа, является задача определения оптимального содержания динамического резерва и его размещения по узлам ВС.

Пусть M – число узлов ВС, N – число информационных элементов системы; a_n – относительная частота корректировок n -ого информационного элемента (число корректировок на заданный период времени T), t_{nm}^n – время передачи n -ого информационного элемента из m -ого в m' -ый узел ВС, b_n – величина объёма n -ого информационного элемента, $W = \llbracket \lambda_{nm} \rrbracket$ – матрица взаимосвязей источников информации (узлов ВС) и информационных элементов,

$$\lambda_{nm} = \begin{cases} 1, \text{ если источником } n\text{-го информационного элемента} \\ \text{является } m\text{-ый узел ВС} \\ 0 \text{ в противном случае.} \end{cases}$$

$C = \llbracket c_{nm} \rrbracket$ – матрица выигрышей, где c_{nm} – взвешенные оценки выигрыша, получаемого пользователем в результате размещения n -ого информационного элемента в m -ом узле ВС, $c_{nm} \in [0, 1], \forall n, n = \overline{1, N}, m = \overline{1, M}$. Способы построения матрицы C определяются конкретными условиями применения РВС. В частности, оценка выигрыша, получаемого пользователем в результате размещения n -ого информационного элемента в m -ом узле ВС может складываться из объективных оценок (например, времени передачи n -ого информационного элемента в m -ый узел из других узлов ВС) и субъективной оценки негативизма по отношению к необходимости передачи n -ого информационного элемента в m -ый узел ВС.

Тогда, используя следующие переменные величины:

$$x_{nm} = \begin{cases} 1, \text{ если } n\text{-ый информационный элемент} \\ \text{размещается в } m\text{-ом узле ВС,} \\ 0 \text{ в противном случае.} \end{cases}$$

степень виртуальности резерва $P(x_{nm})$ можно определить по следующей формуле

$$P(x_{nm}) = 1 - \prod_{n=1}^N (1 - a_n \sum_{m=1}^M \sum_{m'=1}^M x_{nm} \lambda_{nm'} t_{nm'}^n).$$

В случаях, когда распределенная вычислительная среда состоит из однородных элементов [9], в качестве основных критериев синтеза динамического резерва целесообразно использовать максиминный критерий равномерного распределения выигрыша по узлам ВС, критерий минимума степени виртуальности динамического резерва и др.

Модели динамического резервирования данных в распределенной вычислительной среде

Оптимизационная модель синтеза динамического резерва по критерию равномерного распределения выигрыша по узлам ВС имеет следующий вид.

Найти:

$$\max \min_m \sum_{n=1}^N C_{nm} X_{nm}, \quad (2)$$

при ограничениях:

- на степень виртуальности резерва

$$1 - \prod_{n=1}^N \left(1 - a_n \sum_{m=1}^M \sum_{m'=1}^M x_{nm} \lambda_{nm'} t_{nm'}^n \right) \leq P^* \quad (3)$$

где P^* – максимально допустимая степень виртуальности резерва;

- на относительное время корректировок информационных элементов

$$\sum_{n=1}^N \sum_{m=1}^M \sum_{m'=1}^M x_{nm} a_n \lambda_{nm'} t_{nm'}^n \leq T_{\text{кор}}^* \quad (4)$$

где $T_{\text{кор}}^*$ – максимально допустимое относительное время корректировки информационного элемента (на интервале T);

- на объем внешней памяти m -ого узла ВС

$$\sum_{n=1}^N x_{nm} b_n \leq B_m^*, \forall m, m = \overline{1, M} \quad (5)$$

где B_m^* – максимально допустимый объем памяти m -ого узла для хранения информации;

- на отсутствие дублирования информационного элемента в узлах ВС

$$\sum_{n=1}^N x_{nm} \leq 1, \forall m, m = \overline{1, M} \quad (6)$$

Утверждение 1.

Решение задачи 2-6 является допустимым при $P^* \Rightarrow 1$ и выполнении ограничений 4-6.

Доказательство.

Пусть $i_n = \sum_{m=1}^M \sum_{m'=1}^M x_{nm} a_n i_{nm'} t_{nm'}^n$. Тогда ограничения 3, 4 будут иметь следующий вид

$$\begin{cases} \sum_{n=1}^N 1 - \prod_{n=1}^N (1 - i_n) \leq P^* \\ i_n \leq t_{\text{кор}}^* \end{cases} \quad (7)$$

Прологарифмировав первое неравенство системы 7 и разложив функцию логарифма в степенной ряд последовательно, получим:

$$\begin{cases} \sum_{n=1}^N \ln(1 - i_n) \leq \ln(1 - P^*) \\ \sum_{n=1}^N i_n \leq t_{\text{кор}}^* \end{cases} \quad \begin{cases} \sum_{n=1}^N i_n \leq \ln(1 - P^*) - i \\ \sum_{n=1}^N i_n \leq t_{\text{кор}}^* \end{cases}$$

где i – остаточный член ряда.

Очевидно, что при $P^* \Rightarrow 1$ ограничение 4 является более жестким. Предложенный анализ жесткости ограничений может быть использован для снижения размерности задач с конкретными исходными данными.

Современные РВС организационного типа, функционирующие на базе ВС критичны к объёму информации, передаваемой по каналам связи. Поэтому наиболее актуальной можно считать задачу нахождения

$$\max_{x_{nm}} \min_m \sum_{n=1}^N C_{nm} X_{nm}$$

при ограничении 4. Данная задача сводится к задаче максимизации путём введения дополнительных переменных $y_m = \{0,1\}$. Она имеет следующий вид.

Найти:

$$\max_{\{x_{nm}, y_m\}} \sum_{m=1}^M y_m \sum_{n=1}^N C_{nm} X_{nm}$$

при ограничениях:

$$\begin{aligned} \sum_{m=1}^M y_m \sum_{n=1}^N C_{nm} X_{nm} &< \sum_{n=1}^N C_{nm} X_{nm}, \forall m, m = \overline{1, M}, \\ \sum_{m=1}^M y_m &= 1, \\ \sum_{n=1}^N \sum_{m=1}^M \sum_{m'=1}^M x_{nm} a_n i_{nm} / t_{nm'}^n &\leq T_{\text{кор}}^*. \end{aligned}$$

Пусть $t_{mm'}^n = \text{const}, \forall m, m = \overline{1, M}, \forall m', m = \overline{1, M}, m \neq m^*, \forall n, n = \overline{1, N}$,

Тогда решение задачи определения оптимального содержания динамического резерва и его размещения в ВС сводится к решению M многомерных задач о ранце [9], которые формулируются следующим образом.

Найти:

$$\sum_{n=1}^N C_{nm^*} X_{nm^*}, \forall m^*, m^* = \overline{1, M}$$

при ограничениях:

$$\sum_{m=1}^M \sum_{n=1}^N x_{nm} a_n \leq D, D = \frac{T_{\text{кор}}^*}{t_{mm}^n},$$
$$\sum_{n=1}^N (c_{nm^*} + a_n) + \sum_{n=1}^N (a_n - c_{nm^*}) x_{nm'} + \sum_{m=1}^M \sum_{n=1}^N x_{nm} a_n \leq D,$$
$$\forall m', m' = 1, M, m' \neq m^*.$$

Заключение

Повышение эффективности сохранности данных в РВС основывается на резервировании и восстановлении данных. При решении задачи сохранности данных методами резервирования и восстановления используется динамический резерв, в состав которого входят как сами данные, так их копии и/или предыстории. В результате, формализована задача определения оптимального содержания динамического резерва и его размещения по узлам ВС, решаемая на этапе предпроектного анализа. В случаях, когда РВС состоят из однородных элементов, например, сетевых кластеров в качестве основных критериев синтеза динамического резерва предлагается использовать максиминный критерий равномерного распределения выигрыша по узлам ВС, критерий минимума степени виртуальности резерва и др.

ЛИТЕРАТУРА

1. Сизов В.А. Мониторинг репутационных рисков компаний в условиях информационного противоборства. Интернет-журнал Науковедение. том 9, №5, 2017 <https://naukovedenie.ru/PDF/88TVN517.pdf>.
2. Сизов В.А. Методология разработки моделей системы информационного противоборства. Инжиниринг предприятий и управление знаниями (ИП&УЗ-2016): сборник научных трудов XIX научно-практической конференции. 26-27 апреля 2016 г. / под науч. ред. Ю. Ф. Тельнова. – Москва: ФГБОУ ВО «РЭУ им. Г.В. Плеханова», 2016. – 420 с. (ISBN 978-5-7307-1109-9).
3. Петриченко Е.А. Информация и философские категории «явление и сущность» // Современные проблемы науки и образования. – 2014. – № 1: журнал. – 2014. – № 1.
4. Таненбаум Э., ван Стеен М. Распределенные системы. Принципы и парадигмы. СПб.: Питер, 2003.
5. Казарин О.В. Безопасность программного обеспечения компьютерных систем. М.: МГУЛ, 2003.
6. В.В. Кульба, С.К. Сомов. Репликация, резервирование и схемы восстановления информации в ненадежных распределенных системах. Вестник РГГУ. Серия «Экономика. Управление. Право» № 3(9), 2017 г., с. 28-39.
7. Сизов В.А. Модели и методы виртуально-восстановительного резервирования данных автоматизированных информационно-управляющих систем в условиях чрезвычайных ситуаций // Журнал «Автоматика и телемеханика», №7, 1998 г.
8. Климанов В.П., Со Хтет Зо. Математическая модель сетевого кластера типа пирамидальная решетка / В.П. Климанов, Со Хтет Зо // Вестник МГТУ «СТАНКИН», № 3(34), 2015.
9. Габидулин Э.М., Кшевецкий А.С., Колыбельников А.И. Защита информации: учебное пособие – М.: МФТИ, 2011. – 225 с. – ISBN 978-5-7417-0377-9.

Sizov Valerii Aleksandrovich

Plekhanov Russian university of economics, Moscow, Russia
E-mail: sizovva@gmail.com

Development of models for improving the efficiency of data security in a distributed computing environment based on dynamic data backup

Abstract. The article deals with the problem of improving data integrity in a distributed computing environment through the integrated use of backup and recovery methods. We consider a distributed computing environment built on the basis of a computer network, which is used as an organizational-technical form of information flow management in large information systems. For this, the notion of a dynamic reserve is defined, which includes both the data itself and their copies and/or history. The task of determining the optimal content of a dynamic reserve and its placement on the nodes of a computer network is formalized. Models of dynamic data backup in a distributed computing environment are developed according to the criterion of uniform distribution of gain across nodes of the computer network. The optimization model of dynamic reserve synthesis for homogeneous clusters of computer networks is substantiated. To simplify the solution of the proposed optimization problem, its decomposition into a set of “about the knapsack” problems was considered, for solving which one can use known algorithms. The use of a dynamic reserve allows to increase the efficiency of data integrity in a distributed computing environment due to the integrated consideration of user requirements and technical parameters of the computer network.

Keywords: distributed computing environment; computing network; data backup; data recovery; copies; history; modeling; optimization