

ОПТИМИЗАЦИЯ РЕЗЕРВИРОВАНИЯ ИНФОРМАЦИИ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ ОБРАБОТКИ ДАННЫХ РЕАЛЬНОГО ВРЕМЕНИ

Е.А. Микрин, С.К. Сомов

Рассмотрены методы повышения надежности работы распределенных систем обработки данных, работающих в реальном масштабе времени. Перечислены особенности и характеристики таких систем по сравнению с системами обработки данных, работающими в обычном режиме. Получены аналитические выражения для расчета коэффициента готовности таких систем. Сформулированы задачи оптимизации резервирования данных в распределенных системах обработки данных реального времени.

Ключевые слова: распределенная система обработки данных, реальное время, оптимальное резервирование, коэффициент готовности.

ВВЕДЕНИЕ

Распределенные системы обработки данных различного назначения (корпоративные системы, системы специального назначения, сервисные системы общего доступа) — это сложные аппаратно-программные комплексы. Компоненты таких систем могут находиться на больших расстояниях друг от друга и объединяются в единую систему с помощью специального системного программного обеспечения и каналов связи. Распределенные системы обработки данных (РСОД) строятся, как правило, на основе глобальных, корпоративных или вычислительных сетей специального назначения.

В процессе работы таких систем возможны различные инциденты, которые могут привести к появлению ошибочных данных, к ошибочным результатам решения задач и обработки запросов системой, а в крайнем случае и к невозможности нормального функционирования РСОД, т. е. к потере системой работоспособности, ее отказу.

Для повышения безопасности работы систем обработки данных вводят различного рода избыточность [1, 2]. В частности, задача обеспечения сохранности используемых системой данных решается главным образом применением информационной избыточности (созданием и хранением резервных данных). Основным методом обеспече-

ния сохранности данных служит метод резервирования, который предполагает использование идентичной избыточности (создание и хранение копий данных) и/или неидентичной (создание и хранение предысторий данных).

Для РСОД, базирующихся на вычислительных сетях и работающих в жестких условиях реального масштаба времени, характерны некоторые особенности, основная из которых состоит в наличии существенных ограничений на время обработки запросов к данным системы и на время решения задач в системе. В результате причиной отказа системы может стать не только выход из строя ее аппаратных элементов, но и превышение временем решения задачи или обработки запроса некоторого установленного лимита.

Для повышения безопасности работы РСОД реального времени различных классов и назначения применяется метод резервирования используемых системой массивов данных.

1. ПОСТАНОВКА ЗАДАЧИ

К РСОД, работающих в режиме реального времени, предъявляются жесткие требования ко времени реакции системы (обработка запросов или решение задач в системе). Поэтому коэффициент готовности системы K_T служит основным показателем качества функционирования таких систем.

Согласно классическому определению [1, 3], коэффициент готовности системы — это вероятность того, что система окажется работоспособной в произвольный момент времени, кроме планируемых периодов, в течение которых работа системы по назначению не предусматривается.

Для достижения максимального значения коэффициента готовности системы обработки данных, работающей в реальном времени, необходимо обеспечить минимум времени нахождения ее в неработоспособном состоянии. Один из методов, позволяющий минимизировать время ее нахождения в неработоспособном состоянии, заключается в резервировании массивов данных, с которыми она работает.

Для повышения безопасности функционирования систем обработки данных (СОД) различных классов и назначения в настоящее время широко применяются три основные стратегии резервирования информации, которые учитывают особенности обновления и использования массивов данных [2, 4, 5].

Стратегия I заключается в использовании некоторого числа копий основного массива данных. В случае, если основной массив разрушен, то для продолжения решения задачи система переходит к использованию первой его копии, если же и она разрушена, то используется следующая копия и т. д.

Стратегия II основывается на учете особенностей обновления текущих массивов данных и заключается в том, что вместо копий текущего массива используются его предыстории (предыдущие версии массива и журналы их изменений). В случае разрушения текущего массива запускается специальная программа, которая восстанавливает разрушенный массив с использованием его предыдущей версии и записей из журнала изменений. Если восстановление закончилось неудачно, то текущий массив восстанавливается из следующей предыстории и т. д.

Стратегия III это смешанная стратегия, использующая возможности первых двух стратегий резервирования, т. е. она использует как копии, так и предыстории текущего массива данных. Сначала используются копии массива в соответствии со стратегией I, а затем, в случае неудачи восстановления из копий, используются предыстории массива в соответствии со стратегией II.

Под отказом РСОД реального времени, будем понимать превышение временем решения задачи (обработки запроса) системой некоторой заданной пороговой величины t_n . Время, необходимое для восстановления разрушенного резерва массива данных, используемого в РСОД реального вре-

мени, обозначим как t_b . В соответствии с приведенным определением коэффициент готовности

$$K_r = E(t^*)/x,$$

где x — среднее время решения задачи, т. е. время между моментом начала решения задачи и моментом времени, когда система готова приступить к решению следующей задачи; $E(t^*)$ — среднее время работоспособного состояния системы, значение которого определяется далее (см. формулу (1)).

Время x определяется по формуле:

$$x = E(T_v) + (1 - \rho_v)t_b,$$

где $(1 - \rho_v)$ — вероятность того, что при решении задачи будут разрушены резервируемые массивы данных; $E(T_v)$ — среднее время решения задачи без учета затрат времени на восстановление разрушенных массивов данных.

Значение среднего времени $E(T_v)$ зависит от того, какая из трех $v = I, II, III$, стратегий резервирования массивов данных применяется в системе.

Время пребывания запроса в РСОД реального времени состоит из времени ω ожидания запроса в очереди на обработку и времени T собственно обработки запроса в системе. Отказ системы при обработке запроса наступает, если $T + \omega > t_n$. Известны работы, в которых определяются дисциплины диспетчеризации обработки запросов, обеспечивающие минимальное среднее время ожидания запроса в очереди. В данной работе основное внимание посвящено выбору оптимальных методов резервирования массивов данных, которые позволяют снизить вероятность отказа системы уже на этапе обработки запросов, тем самым сокращая время обработки запроса в системе.

Пусть имеется ситуация, когда $0 \leq \omega \leq t_n$. Так как время ω ожидания запроса в очереди на обработку случайная величина, то время T , в течение которого возможна обработка запроса без отказа системы, также случайная величина: $T = t_n - \omega$.

Допустим, что для случайной величины T задана функция распределения $F_T(t) = P\{T < t\}$. Тогда, в соответствии с приведенным выше определением отказа, время t^* работоспособного состояния системы можно представить в виде зависимости от значения T (рис. 1):

$$t^* = \begin{cases} x & \text{при } T > x, \\ T & \text{при } T \leq x. \end{cases}$$

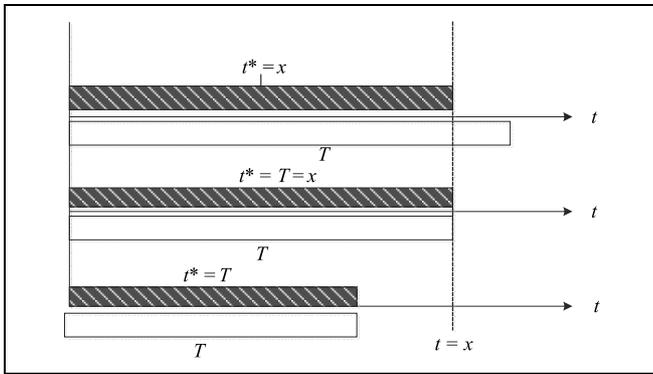


Рис. 1. Зависимость времени t^* работоспособного состояния системы от времени T , в течение которого возможна обработка запроса

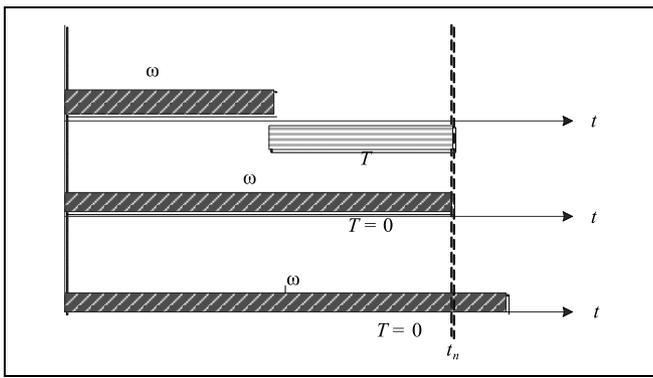


Рис. 2. Зависимость времени T от времени ω ожидания запроса в очереди на обработку

Среднее время $E(t^*)$ работоспособного состояния системы

$$E(t^*) = xP\{T > x\} + \int_0^x tf_T(t)dt, \quad (1)$$

где $P\{T > x\} = 1 - F_T(x)$; $f_T(t) = dF_T(t)/dt$.

Предположим, что рассматриваемую нами систему РСОД реального времени можно отнести к классу систем массового обслуживания типа М/М/1. На вход системы поступает пуассоновский поток запросов с интенсивностью λ . Время обслуживания запросов распределено по показательному закону с параметром μ . В таких системах время ожидания в очереди на обслуживание описывается функцией распределения [6]

$$W(t) = P\{\omega < t\} = 1 - \beta e^{-\mu(1-\beta)t},$$

где $\beta = \lambda x = \lambda \mu^{-1}$ — коэффициент использования системы.

Допустим, что нам задано значение t_n . Тогда (рис. 2)

$$T = \begin{cases} t_n - \omega & \text{при } \omega < t_n, \\ 0 & \text{при } \omega \geq t_n. \end{cases}$$

Так как при $0 \leq \omega \leq t_n$ выполняется $T = t_n - \omega$, то при $0 < t \leq t_n$ событию $\omega < t$ соответствует неравенство $T > t_n - t$. В этом случае справедливо выражение:

$$P\{\omega < t\} = P\{T > t_n - t\} \text{ при } 0 < t \leq t_n.$$

Следовательно,

$$P\{T > t_n - t\} = \begin{cases} 0 & \text{при } t \leq 0, \\ P\{\omega < t\} & \text{при } 0 < t \leq t_n, \\ 1 & \text{при } t > t_n. \end{cases}$$

Так как $P\{T < t_n - t\} = 1 - P\{T > t_n - t\}$, то, заменив $t_n - t$ на τ , получим:

$$P\{T < \tau\} = \begin{cases} 1 & \text{при } \tau \geq t_n, \\ 1 - P\{\omega < t_n - \tau\} & \text{при } 0 \leq \tau < t_n, \\ 0 & \text{при } \tau < 0. \end{cases}$$

Перейдя от τ к t , окончательно получим:

$$F_T(t) = P\{T < t\} = \begin{cases} 1 & \text{при } t \geq t_n, \\ \beta e^{-\mu(1-\beta)(t_n-t)} & \text{при } 0 \leq t < t_n, \\ 0 & \text{при } t < 0. \end{cases}$$

Отсюда для $P\{T > x\} = 1 - P\{T < x\}$:

$$P\{T > x\} = \begin{cases} 0 & \text{при } t \geq t_n, \\ 1 - \beta e^{-\mu(1-\beta)(t_n-x)} & \text{при } 0 \leq x < t_n, \\ 1 & \text{при } x < 0. \end{cases}$$

В соответствии с формулой (1) и с учетом сделанных предположений получим, что среднее время $E(t^*)$ работоспособного состояния РСОД реального времени определяется по формуле:

$$E(t^*) = \begin{cases} x[1 - \beta e^{-\mu(1-\beta)(t_n-x)}] + We^{-\mu(1-\beta)t_n}[1 - \beta e^{-(1-\beta)x}] & \text{при } 0 \leq x < t_n, \\ 0 & \text{при } x \geq t_n, \end{cases}$$

где $W = \beta x / (1 - \beta)$ — среднее время ожидания запроса в очереди на обработку.

На основе приведенных результатов получим:

$$K_r = B(\lambda, x, t_n) \text{ при } 0 \leq x < t_n,$$

$$\text{где } B(\lambda, x, t_n) = 1 - \beta e^{-(1-\beta)(t_n/x-1)} + \frac{\beta}{1-\beta} e^{-(1-\beta)t_n/x} \times \\ \times [1 - \beta e^{(1-\beta)}].$$

Анализ функции $B(\lambda, x, t_n)$ показывает, что [5]:

- при $\beta \rightarrow 0$ ($x \rightarrow 0$ или $\lambda \rightarrow 0$) $K_r \rightarrow 1$;
- $K_r \rightarrow 0$ при $\beta \rightarrow 1$;
- $K_r \rightarrow 1$ при $t_n \rightarrow \infty$;
- при $t_b \gg E(T_v)$ для стратегий резервирования I, II и III выполняется соотношение $K_r^I > K_r^{III} > K_r^{II}$.

2. КОЭФФИЦИЕНТ ГОТОВНОСТИ РАСПРЕДЕЛЕННОЙ СИСТЕМЫ ОБРАБОТКИ ДАННЫХ РЕАЛЬНОГО ВРЕМЕНИ

Одна из особенностей работы РСОД реального времени состоит в том, что при децентрализованном хранении ресурсов системы для обработки запросов можно применять различные дисциплины обработки. Основные из них [4, 5]:

1) запрос для его обработки адресуется в ближайший (согласно некоторому критерию) узел с необходимыми ресурсами системы;

2) запрос адресуется для обработки одновременно в K узлов с ресурсами системы;

3) запрос, полученный в j -м узле, последовательно передается по содержащим ресурсы системы узлам пути длиной K , начинающегося в j -м узле, до тех пор, пока в некотором узле этого пути запрос не будет успешно обработан системой либо не будут пройдены все K узлов этого пути;

4) из j -го узла, в котором получен запрос, этот запрос поочередно направляется в K ближайших узлов с ресурсами системы до тех пор, пока либо в некотором узле он не будет успешно обработан, либо не будут опрошены все K узлов.

Для дисциплины 1 обработки запросов в РСОД реального времени определим значение коэффициента K_r готовности системы по отношению к обработке всех возникающих в сети запросов к РСОД и значение коэффициента готовности K_{rj} системы по отношению к обработке запросов, выданных в j -м узле сети.

В соответствии с дисциплиной 1 информационный запрос для обработки адресуется в ближайший узел с ресурсами системы. В этом случае

$$K_{rj} = \sum_{i \in J_p} \psi_{ji} r_{ji} B(\lambda_i, x_i, t_{ni}) r_{ij} = \sum_{i \in J_p} \psi_{ji} r_{ji} B_i r_{ij}$$

Здесь:

$$\psi_{ji} = \begin{cases} 1, & \text{если запрос из узла } j \text{ адресуется в узел } i; \\ 0 & \text{в противном случае;} \end{cases}$$

J_p — множество индексов узлов сети с ресурсами РСОД реального времени;

$$t_{ni} = \begin{cases} t_n - 2t_3 & \text{при } i \neq j; \\ t_n & \text{при } i = j; \end{cases}$$

t_3 — среднее время задержки сообщений в сети;

$$\lambda_i = \sum_{j=1}^N (V_j + U_j \psi_{ji}), \text{ где } V_j \text{ — интенсивность запросов на модификацию ресурсов системы, а } U_j \text{ — интенсивность информационных запросов к системе, возникающих в узле } j.$$

Коэффициент готовности K_r системы по отношению к обработке всех запросов, возникающих в сети, определяется по формуле:

$$K_r = \sum_{j=1}^N (V_j + U_j) K_{rj} / \sum_{j=1}^N (V_j + U_j). \quad (2)$$

Аналогичным образом получаются формулы, приведенные в табл. 1, для определения K_{rj} при

Таблица 1

Формулы расчета характеристик системы

Дисциплина обработки запросов	Характеристики K_{rj} и t_{ni}
1	$K_{rj} = \sum_{i \in J_p} \psi_{ji} r_{ji} B_i r_{ij}$ $t_{ni} = \begin{cases} t_n - 2t_3 & \text{при } i \neq j, \\ t_n & \text{при } i = j \end{cases}$
2	$K_{rj} = 1 - \prod_{i \in N_j} (1 - B_i r_{ji} r_{ij})$ $t_{ni} = \begin{cases} t_n - 2t_3 & \text{при } i \neq j, \\ t_n & \text{при } i = j \end{cases}$
3	$K_{rj} = \sum_{m=1}^K r_{ij_1} r_{i_m j} B_{i_m} \prod_{\alpha=2}^m (1 - P_{i_{\alpha-1}}) r_{i_{\alpha-1} i_\alpha}$ $t_{ni_m} = t_n - (m+1-y_j)t_3 - \sum_{\alpha=1}^{m-1} (W_{i_\alpha} + x_{i_\alpha})$ $i_m, i_\alpha \in N_j$
4	$K_{rj} = \sum_{m=1}^K r_{ji_m} r_{i_m j} B_{i_m} \prod_{\alpha=1}^{m-1} (1 - r_{ji_\alpha} P_{i_\alpha} r_{i_\alpha j})$ $t_{ni_m} = t_n - (m-y_j)2t_3 - \sum_{\alpha=1}^{m-1} (W_{i_\alpha} + x_{i_\alpha})$ $i_m, i_\alpha \in N_j$



использовании дисциплин 2, 3 и 4 обработки запросов [3].

В формулах табл. 1 приняты обозначения:

N_j — множество индексов узлов сети, в которые можно адресовать для обработки информационные запросы, возникшие в узле j ;

$$y_j = \begin{cases} 1, & \text{если } j \in N_j; \\ 0, & \text{если } j \notin N_j; \end{cases}$$

$$W_{i_a} = \beta_{i_a} x_{i_a} / (1 - \beta_{i_a});$$

$$K = |N_j|.$$

3. АНАЛИЗ ЭФФЕКТИВНОСТИ ДИСЦИПЛИН ОБРАБОТКИ ЗАПРОСОВ

Проведем качественную оценку эффективности рассматриваемых дисциплин обработки запросов в РСОД реального времени в смысле величин K_{rj} и K_r в случае, когда система построена на основе однородной полносвязной вычислительной сети, содержащей равномерно распределенный резерв массивов данных системы.

Учитывая полносвязность и однородность вычислительной сети и учитывая формулы (см. табл. 1), можно показать, что справедливы неравенства [4]:

$$K_{rj}^2 > K_{rj}^1, \quad K_{rj}^2 > K_{rj}^3, \quad K_{rj}^2 > K_{rj}^4,$$

$$K_{rj}^3 > K_{rj}^1, \quad K_{rj}^4 > K_{rj}^1,$$

где индексы 1, 2, 3 и 4 соответствуют дисциплинам 1, 2, 3 и 4 обработки запросов.

Справедливость неравенства $K_{rj}^4 > K_{rj}^3$ следует из соотношения $(1 - P)r < r(1 - r)P$, которое выполняется, так как $0 < r < 1$. В итоге справедливо соотношение

$$K_{rj}^2 > K_{rj}^4 > K_{rj}^3 > K_{rj}^1,$$

из которого также следует и справедливость неравенства:

$$K_r^2 > K_r^4 > K_r^3 > K_r^1.$$

Сформулируем полученные результаты в виде утверждения.

Утверждение. *Применение дисциплины 2 обработки запросов по сравнению с применением дисциплин 1, 3 и 4 обеспечивает наибольшее значение коэффициента готовности РСОД реального времени, построенной на базе однородной полносвязной вычислительной сети.*

В вычислительных сетях возможны разнообразные варианты размещения резервов РСОД реального времени, обеспечивающие различные значения стоимости и надежности функционирования системы. В связи с этим возникает проблема поиска оптимального размещения резерва по узлам вычислительной сети.

4. ЗАДАЧА ОПТИМАЛЬНОГО РЕЗЕРВИРОВАНИЯ В РАСПРЕДЕЛЕННОЙ СИСТЕМЕ ОБРАБОТКИ ДАННЫХ РЕАЛЬНОГО ВРЕМЕНИ

Задача оптимального резервирования массивов данных, используемых в РСОД реального времени, согласно критерию максимума коэффициента готовности системы формулируется следующим образом: необходимо найти такое размещение X резерва по узлам вычислительной сети, чтобы затраты ресурсов системы на обработку поступающих запросов не превышали некоторого лимита при максимално возможном значении коэффициента K_r готовности системы, т. е.:

$$K_r \rightarrow \max$$

при $ZP(X) \leq \bar{S}$,

$$x_k \leq \bar{x}_k, \quad k = \overline{1, N},$$

где $X = (x_k)$; x_k — объем резерва, размещенного в k -м узле сети.

Значение коэффициента K_r определяется по формуле (2), а затраты $ZP(X)$ — по формулам, полученным в работе [4] и приведенным в табл. 2.

Таблица 2

Зависимость затрат ресурсов системы от дисциплины обработки запросов

Дисциплина	Затраты ZP
1	$ZP = \sum_{j=1}^N \sum_{k=1}^N (U_j y_{jk} + V_j y_k) ZP_{jk}(x_k)$
2	$ZP = \sum_{j=1}^N \sum_{k=1}^N (U_j y_{jk} + V_j y_k) ZP_{jk}(x_k)$
3	$ZP = \sum_{j=1}^N \sum_{k=1}^N (U_j + V_j y_k) ZP_{jk}(x_k)$
4	$Z = \sum_{j=1}^N \sum_{k=1}^N \left\{ U_j \prod_{i=1}^{n(j,k)-1} [1 - r_{ji} P_{ji}(x_{ji}) r_{ji}] + V_j y_k \right\} Z_{jk}(x_k)$ $j_i \in N_j, \quad i = \overline{1, K}$

Возможна также постановка обратной задачи оптимального распределения резерва ресурсов РСОД реального времени по узлам сети, в которой в качестве критерия оптимизации принят минимум затрат $ZP(X)$, а на значение коэффициента готовности системы наложены ограничения, т. е.:

$$ZP(X) \rightarrow \min$$

при

$$K_{ij} \geq \bar{K}, \quad j \in J_3,$$

$$x_k \leq \bar{x}_k, \quad k = \overline{1, N}.$$

Здесь J_3 — множество индексов узлов сети, в которых возникают запросы к системе.

ЗАКЛЮЧЕНИЕ

Кратко рассмотрены методы повышения надежности функционирования распределенных систем обработки данных, работающих на базе вычислительных сетей в режиме реального времени, путем резервирования информационных и программных ресурсов системы. Рассмотрены особенности и характеристики таких систем по сравнению с системами обработки данных, работающими в обычном режиме.

Для систем, построенных на базе вычислительной сетей, получены аналитические выражения для определения основного показателя надежности их функционирования — коэффициента готовности — с учетом жестких ограничений на время получения ответа на запрос. При этом учитывается возможность применения различных дисциплин

обработки запросов к системе. Сформулированы задачи оптимизации резервирования массивов данных в системе по критерию максимума коэффициента ее готовности и минимума затрат на обработку запросов. Приведенные результаты могут быть полезны при проектировании распределенных систем обработки данных, работающих в рамках глобальных, корпоративных или вычислительных сетей специального назначения.

ЛИТЕРАТУРА

1. Козлов Б.А., Ушаков И.А. Справочник по расчету надежности аппаратуры радиоэлектроники и автоматики. — М.: Советское радио, 1975. — 472 с.
2. Кульба В.В., Мамиконов А.Г., Шелков А.Б. Резервирование программных модулей и информационных массивов в АСУ // Автоматика и телемеханика. — 1980. — № 8. — С. 133–141.
3. Надежность в технике. Термины и определения. ГОСТ 13377–75. — М.: Изд-во стандартов, 1975. — 28 с.
4. Кульба В.В., Сомов С.К., Шелков А.Б. Резервирование данных в сетях ЭВМ. — Казань: Изд-во Казанского ун-та, 1987. — 175 с.
5. Сомов С.К. Резервирование программных модулей и информационных массивов в сетях ЭВМ: дисс. ... канд. техн. наук. — М.: ИПУ РАН, 1983. — 217 с.
6. Клейнрок Л. Вычислительные системы с очередями — М.: Мир, 1979. — 600 с.

Статья представлена к публикации членом редколлегии В.В. Кульбой.

Микрин Евгений Анатольевич — академик РАН, ген. конструктор, ПАО «Ракетно-космическая корпорация «Энергия» им. С.П. Королева», г. Королев, [✉ Eugeny.Mikrin@rsce.ru](mailto:Eugeny.Mikrin@rsce.ru),

Сомов Сергей Константинович — канд. техн. наук, вед. инженер, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, [✉ ssomov2016@ipu.ru](mailto:ssomov2016@ipu.ru).

Новая книга

Финягина В.И. Анатолий Григорьевич Бутковский: ученый и человек / Под ред. академика РАН С.Н. Васильева. — М.: ФИЗМАТЛИТ, 2016. — 216 с. + 34 с. цв. вклеек. — ISBN 978-5-9221-1677-0.

А.Г. Бутковский — выдающийся ученый в области теории и систем управления и в первую очередь в области оптимального управления системами с распределенными параметрами. Он был уникален по широте своих научных и общегуманитарных интересов — от математических проблем управления с приложениями в физике, металлургии, теплотехнике, химических технологиях, авиапромышленности и других областях науки и техники до методологических и философских проблем образования, науки и культуры.

Книга содержит обзор научных достижений А.Г. Бутковского, воспоминания родных, друзей и коллег, а также его неопубликованные рукописи, завещанные читателям к использованию в научной и научно-исторической работе.

Для широкого круга читателей, интересующихся историей развития науки.