

УДК 004.652.4

UDC 004.652.4

МЕТОДИКИ ОПТИМАЛЬНОГО ВЫБОРА КОНФИГУРАЦИИ И СТРУКТУРИРОВАНИЯ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ СБОРА И ХРАНЕНИЯ ХРОНОЛОГИЧЕСКИХ ДАННЫХ С НЕСКОЛЬКИМИ ЦЕНТРАМИ ОБРАБОТКИ ДАННЫХ

THE TECHNIQUES FOR OPTIMAL CONFIGURATION SELECTION AND STRUCTURING OF DISTRIBUTED INFORMATION SYSTEMS FOR COLLECTING AND STORING HISTORICAL DATA FROM MULTIPLE DATA CENTERS

Лаптев Владимир Николаевич
к.т.н., доцент
ФГБОУ «Кубанский государственный аграрный университет», 350044, Россия, г. Краснодар, ул. Калинина 13, E-mail: mail@kubsau.ru

Laptev Vladimir Nikolaevich
Cand.Tech.Sci., associate professor
Kuban State Agricultural university, 350044, Russia, Krasnodar, Kalinin st., 13, E-mail: mail@kubsau.ru

Сопильняк Юрий Николаевич
к. пед. н.
ФКБОУ «Краснодарский университет МВД РФ», Краснодар, Россия, 350005, Россия, г. Краснодар, ул. Ярославского 128

Sopilnyak Yuri Nikolaevich
Cand.Ped.Sci.
Krasnodar University of Ministry of internal Affairs of the Russian Federation, 350005 Russia, Krasnodar, Yaroslavskyst st. 128. E-mail: post@krdu-mvd.ru

Дьяченко Роман Александрович
к.т.н., доцент

Dyachenko Roman Aleksandrovich
Cand.Tech.Sci., associate professor

Багдасарян Рафаэль Хачикович
аспирант
ФГБОУ «Кубанский государственный технологический университет», 350072, Россия, г. Краснодар, ул.Московская 2

Bagdasaryan Rafael Khachikovich
postgraduate student
Kuban State Technological University, 350072 Russia, Krasnodar, ., Moskovskaya st, 2

В статье рассматриваются методики оптимального выбора конфигурации и структурирования распределенной информационной системы сбора и хранения хронологических данных с несколькими центрами обработки данных

The article discusses the methodology of optimal configuration selection and structuring of distributed information systems for collecting and storing historical data from multiple data centers

Ключевые слова: СИСТЕМА СБОРА И ХРАНЕНИЯ ДАННЫХ, ОПТИМАЛЬНЫЙ ВЫБОР, ЦЕНТР ОБРАБОТКИ ДАННЫХ

Keywords: COLLECTION AND STORAGE, OPTIMAL CHOICE, DATA PROCESSING CENTER

Введение

В настоящее время широко применяются информационные системы сбора и хранения хронологических данных.

В частности, такие системы применяются в сфере ЖКХ, информационных системах газораспределения, информационных системах энергокластеров и в других областях.

Построение таких систем сопровождается сложными процессами проектирования, оценки работоспособности решения, целесообразности его применения для каждого конкретного случая, а также управленческими

решениями, связанными с выбором из предложенных возможных конфигураций системы.

Описанные сложности создают необходимость применения проверенных методик, предоставляющих инструменты для проектирования и внедрения таких систем, позволяющих получить наиболее приемлемую конфигурацию конечной системы, удовлетворяющую технические, временные и финансовые ограничения.

Таким образом, разработка методики оптимального выбора конфигурации и структурирования распределенной информационной системы сбора и хранения хронологических данных с несколькими центрами обработки данных, являются важной научно-практической задачей [1-3].

1. Методика распределения центров сбора данных по центрам обработки данных для однотипных каналов связи

Задача распределения центров сбора данных по центрам обработки данных для однотипных каналов связи сводится к распределению центров сбора данных по центрам обработки данных в условиях минимизации стоимости подключения и ограничениях на время сбора, передачи, хранения данных и надежность.

Таким образом, задача распределения центров сбора данных (ЦСД) по центрам обработки данных (ЦОД) формулируется в виде задачи минимизации затрат

$$\min \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij} \cdot c_{ij} \right) \quad (1)$$

при ограничениях:

- 1) а однократность включения j -ого ЦСД к ЦОД:

$$\forall j : \sum_{i=1}^n x_{ij} = 1 ; \quad (2)$$

2) на обязательность использования всех центров обработки данных:

$$v_i : \sum_{j=1}^m x_{ij} > 0; \tag{3}$$

3) на время сбора, передачи и записи темпоральной информации в хранилище данных [3]:

$$v_j : \sum_{i=1}^n x_{ij} \cdot t_{ij} \leq t_{\max} \quad . \tag{4}$$

где t_{\max} - максимально допустимое время передачи информации от ЦСД к ЦОД. При поддержке ЦОД параллельной обработки запросов;

4) на надежность канала связи:

$$v_j : \sum_{i=1}^n p_{ij} \cdot t_{ij} \geq p_{\min} \quad . \tag{5}$$

где p_{\min} - требуемое значение надежности канала передачи учетной информации.

Исходными данными для решения задачи являются матрицы:

1) матрица затрат на размещение центров сбора и обработки данных

$$C = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ c_{21} & \dots & c_{2n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \dots & c_{mn} \end{pmatrix} . \tag{6}$$

где c_{ij} - затраты на подключение j -ого ЦСД к i -му ЦОД (включая затраты на проектирование, прокладку канала связи, пуско-наладочные работы, оплату персонала), $i \in I, I = 1, \dots, n$ - множество индексов (номеров) ЦОД, $j \in J, J = 1, \dots, m$ - множество индексов (номеров) ЦСД;

2) матрица времени сбора, передачи и записи данных

$$T = \begin{pmatrix} t_{11} & \dots & t_{1n} \\ t_{21} & \dots & t_{2n} \\ \vdots & \ddots & \vdots \\ t_{m1} & \dots & t_{mn} \end{pmatrix} , \tag{7}$$

где t_{ij} - время сбора, передачи и записи данных при передаче учетной информации от j -ого ЦСД к i -му ЦОД;

3) матрица вероятностей безотказной работы каналов связи

$$P = \begin{pmatrix} p_{11} & \dots & p_{1n} \\ p_{21} & \dots & p_{2n} \\ \vdots & \ddots & \vdots \\ p_{m1} & \dots & p_{mn} \end{pmatrix}, \quad (8)$$

где p_{ij} - вероятность безотказной работы канала связи между j -м ЦСД к i -м ЦОД.

При формировании матрицы затрат на подключение j -ого ЦСД к i -му ЦОД учитываются следующие показатели: A^1 - матрица стоимостей проектных работ;

A^2 - матрица стоимостей коммуникационного оборудования, устанавливаемого на ЦСД; A^3 - матрица стоимостей коммуникационного оборудования, устанавливаемого на ЦОД; A^4 - матрица стоимостей оборудования канала связи между j -м ЦСД и i -м ЦОД; A^5 - матрица стоимостей технических работ по установке коммуникационного оборудования и прокладке канала связи и A^6 - матрица стоимостей пуско-наладочных работ. Капитальные затраты на подключение j -ого ЦСД к i -му ЦОД вычисляются следующим образом:

$$c_{ij} = A_{ij}^1 + A_{ij}^2 + A_{ij}^3 + A_{ij}^4 + A_{ij}^5 + A_{ij}^6. \quad (9)$$

Капитальные затраты на внедрение системы сбора и хранения

данных:

$$A = \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij} \cdot (A_{ij}^1 + A_{ij}^2 + A_{ij}^3 + A_{ij}^4 + A_{ij}^5 + A_{ij}^6) \right) = \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij} \cdot c_{ij} \right). \quad (10)$$

Решением поставленной задачи минимизации является бинарная матрица:

$$(11)$$

$$X = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{pmatrix},$$

где $x_{ij} = 1$ - если j -й ЦСД подключен к i -му ЦОД и $x_{ij} = 0$ в противном случае.

Поставленная задача минимизации может быть решена как методом перебора всех допустимых значений переменной X , так и, в случае большой размерности, другими методами нелинейной целочисленной оптимизации.

Методика, позволяющая решить задачу (1), состоит из следующих этапов:

- 1) формирование множества I (на основе схемы электроснабжения энергокластера);
- 2) формирование множества J (аналогично на основе схемы электроснабжения);
- 3) сбор стоимостной информации, расчет и формирование матрицы C ;
- 4) формирование матрицы T (в случае если отсутствует возможность проведения реальных расчетов, можно использовать имитационную модель, описанную в главе 1);
- 5) определение максимально допустимого времени передачи информации от ЦСД к ЦОД t_{max} ;
- 6) формирование матрицы P (описание одного из методов приведено в главе 3);
- 7) определение требуемого значения надежности канала передачи учетной информации p_{min} ;
- 8) решение задачи (4.1) методами целочисленной оптимизации.

Системно-аналитическое описание методики по технологии *SADT* представлено на *IDEF*-диаграмме 1.

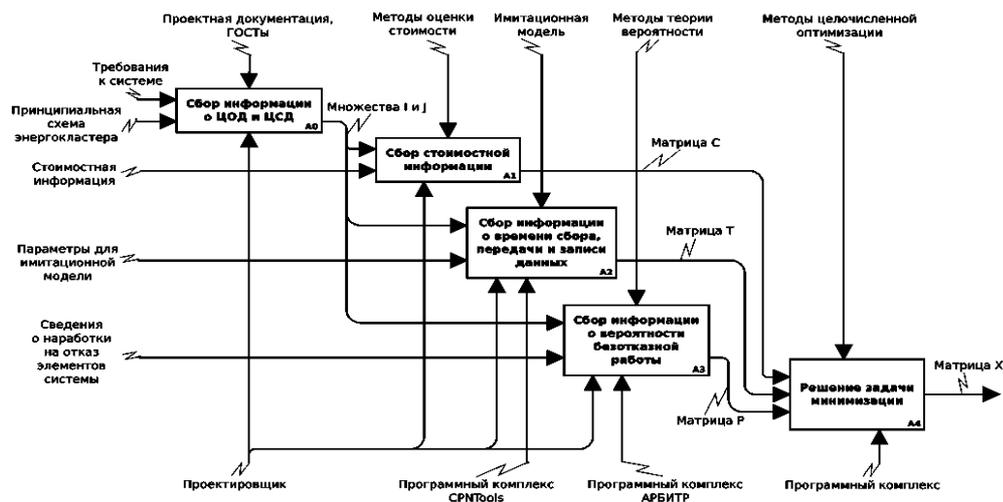


Рисунок 1 - IDEF-0 диаграмма методики конфигурации распределенной информационной системы с несколькими центрами обработки данных

На основе данной методики было разработано программное обеспечение, позволяющее автоматизировать процесс сбора данных, расчет матриц C, T, P и решения оптимизационной задачи (1).

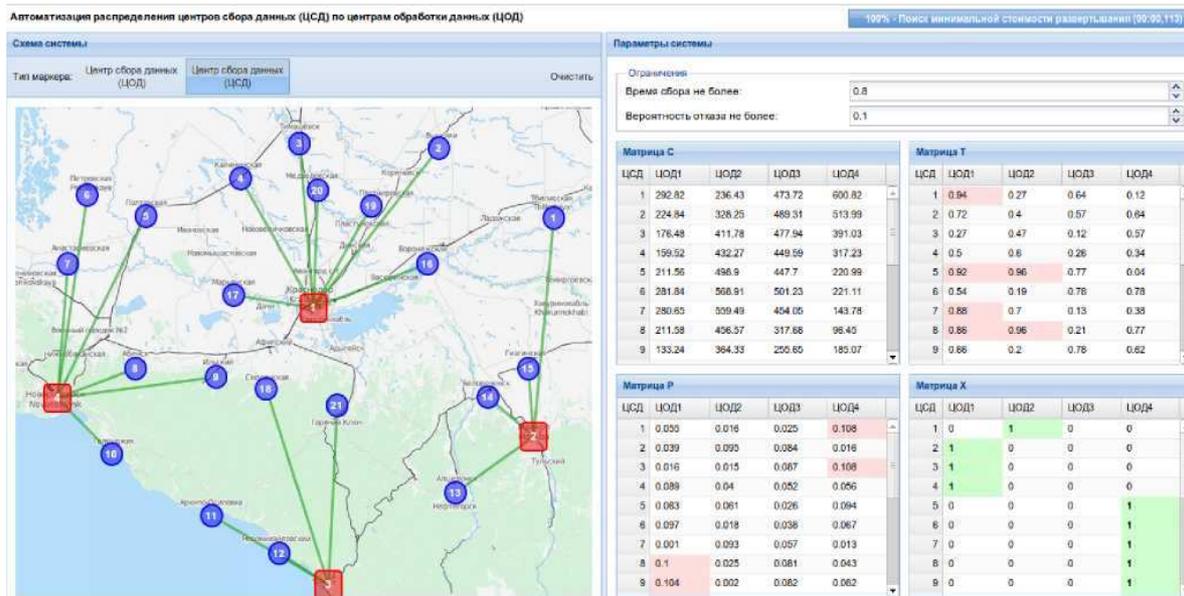


Рисунок 2 - Интерфейс программы вычисления оптимального распределения ЦСД к ЦОД

2. Методика распределения центров сбора данных по центрам обработки данных с использованием нескольких типов каналов связи

Рассмотрим задачу распределения центров сбора данных по центрам обработки данных в условиях минимизации стоимости подключения и ограничениях на время сбора, передачи, хранения данных и надежность, но с возможностью применения различных типов каналов связи.

В таком случае задача распределения центров сбора данных по центрам обработки данных формулируется в виде задачи минимизации затрат

$$\min \sum_{k=1}^l \left(\sum_{i=1}^n \left(\sum_{j=1}^m x_{ijk} \cdot c_{ijk} \right) \right) \quad (12)$$

при ограничениях:

- 1) на однократность включения j -ого ЦСД к ЦОД:

$$\forall j : \sum_{i=1}^n x_{ijk} = 1; \quad (13)$$

- 2) на обязательность использования всех центров обработки данных:

$$\forall i : \sum_{j=1}^m x_{ijk} > 0; \quad (14)$$

- 3) на время сбора, передачи и записи темпоральной информации в хранилище данных [3]:

$$\forall j : \sum_{i=1}^n x_{ijk} \cdot t_{ijk} \leq t_{max}; \quad (15)$$

где t_{max} - максимально допустимое время передачи информации от ЦСД к ЦОД. При поддержке ЦОД параллельной обработки запросов;

- 4) на надежность канала связи:

$$\forall j : \sum_{i=1}^n p_{ijk} \cdot t_{ijk} \geq p_{min}; \quad (16)$$

где p_{min} - требуемое значение надежности канала передачи учетной информации.

Исходными данными для решения задачи являются матрицы:

1) матрица затрат на размещение центров сбора и обработки данных C_{ijk} , где c_{ijk} - затраты на подключение j -ого ЦСД к i -му ЦОД каналом связи k (включая затраты на проектирование, прокладку канала связи, пуско-наладочные работы, оплату персонала), $i \in I, I = 1, \dots, n$ - множество индексов (номеров) ЦОД, $j \in J, J = 1, \dots, m$ - множество индексов (номеров) ЦСД, $k \in K, K = 1, \dots, l$ - множество доступных каналов связи;

2) матрица времени сбора, передачи и записи данных T_{ijk} , где t_{ijk} - время сбора, передачи и записи данных при передаче учетной информации от j -ого ЦСД к i -му ЦОД k -м каналом связи;

3) матрица вероятностей безотказной работы каналов связи P_{ijk} где p_{ijk} - вероятность безотказной работы канала связи между j -м ЦСД к i -м ЦОД на k -м канале связи.

Каждая из матриц C , T и P имеет вид трехмерной матрицы в случае проектирования системы сбора и хранения хронологических данных с применением различных каналов связи. Для этого дополнительно вводится измерение k , $K = 1, \dots, n$ - множество каналов связи между ЦОД и ЦСД.

Решением поставленной задачи минимизации является тринарная матрица X_{ijk} где $x_{ijk} = 1$ - если j -й ЦСД подключен к i -му ЦОД k -м каналом связи и $x_{ijk} = 0$ в противном случае.

Поставленная задача минимизации, также как и в случае использования одного типа канала связи, может быть также решена методом перебора всех допустимых значений переменной Хлибо, в случае большой размерности, другими методами нелинейной целочисленной оптимизации.

Методика, позволяющая решить задачу (12), состоит из следующих этапов:

1) формирование множества I (на основе схемы электроснабжения

энергокластера);

2) формирование множества J (аналогично на основе схемы электроснабжения);

3) формирование множества K (на основе анализа допустимости применения конкретного типа канала связи);

4) сбор стоимостной информации, расчет и формирование матрицы C ;

5) формирование матрицы T (в случае если отсутствует возможность проведения реальных расчетов, можно использовать имитационную модель, описанную в главе 1);

6) определение максимально допустимого времени передачи информации от ЦСД к ЦОД t_{max} ;

7) формирование матрицы P (описание одного из методов приведено в главе 3);

8) определение требуемого значения надежности канала передачи учетной информации p_{min} ;

9) решение задачи (12) методами целочисленной оптимизации.

Системно-аналитическое описание методики по технологии *SADT* представлено на *IDEF-0* диаграмме (рисунок 3).

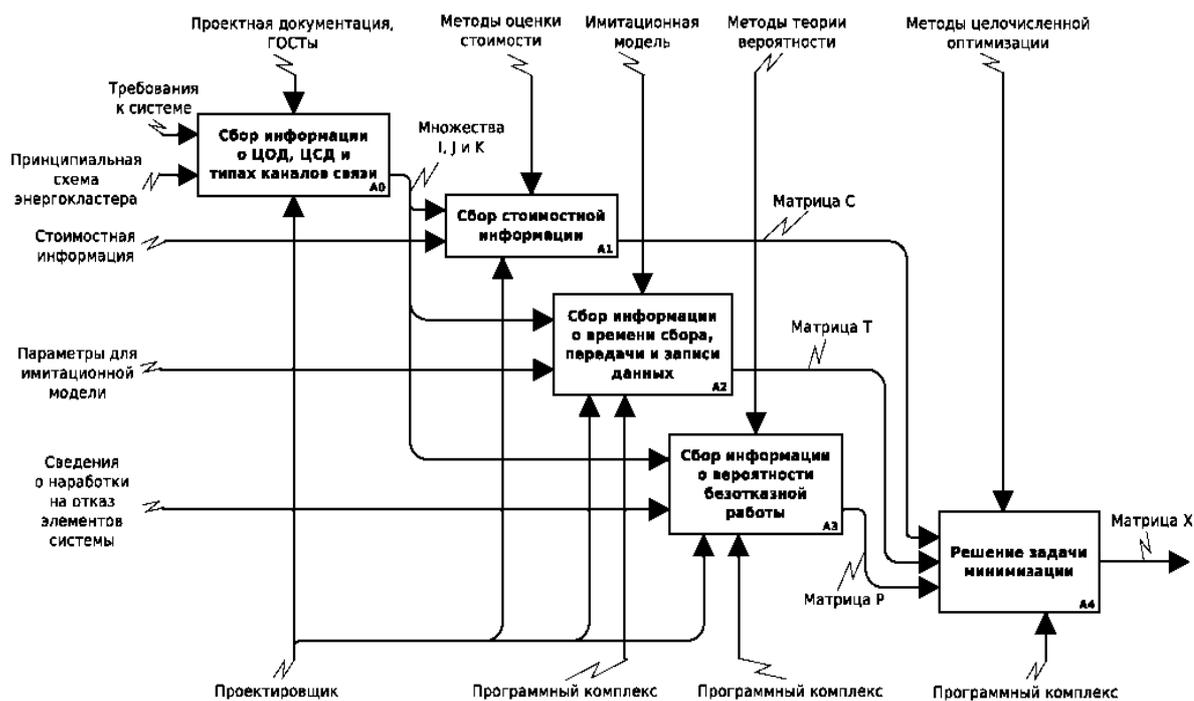


Рисунок 3 - IDEF-0 диаграмма методики конфигурации распределенной информационной системы с несколькими центрами обработки данных с применением различных типов каналов связи

Задача структурирования данных информационной системы сбора и хранения данных решена с помощью известных методик.

Для оценки положительного экономического эффекта от применения предложенной методики были рассмотрены два варианта внедрения системы сбора и хранения данных. Первый вариант использует метод распределения центров сбора данных по центрам обработки данных, использующий принцип подключения к ближайшему доступному узлу. Второй вариант использует для расчета разработанную методику оценки затрат внедрения системы сбора и хранения данных. Результаты проведенных экспериментов представлены в таблице 1.

Таблица 1 - Оценка положительного экономического эффекта от применения предложенной методики

Объект эксперимента	Эффект (снижение стоимости)
Система учета электроэнергии	-14.3%
Система учета потребления газа	-2.9%

Полученные методики позволяют проектировать информационные системы сбора и хранения хронологических данных, применяемых в системах ЖКХ, информационных системах газораспределения, информационных системах энергокластеров и в других областях. Разработанное вспомогательное программное обеспечение позволяет автоматизировать вычисление конфигурации проектируемой системы при известной матрицы затрат на размещение центров сбора и обработки данных, матрицы времени сбора, передачи и записи данных и матрицы вероятностей безотказной работы каналов связи.

Список литературы

1. Дьяченко, Р. Моделирование систем сбора и передачи данных с применением цветных сетей Петри [Электронный ресурс] / Р.А. Дьяченко, А.В. Фишер, В.В. Богданов // Фундаментальные исследования. – 2013.- №11(6).-с. 1122-1126. –Режим доступа: http://rae.ru/fs/?section=content&op=show_article&article_id=10002528
2. Лаптев В.Н. вопросу повышения надежности системы сбора и хранения хронологических данных [Электронный ресурс] / Ю.Н. Сопильняк, Р.А. Дьяченко, Р.Х. Багдасарян // Политематический сетевой электронный научный журнал КубГАУ. – Краснодар: КубГАУ, 2014. №101(07), - 16с. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/162.pdf>.
3. Фишер, А.В. Организация хранения хронологических данных в базах данных систем мониторинга и прогнозирования [Электронный ресурс] / А.В. Фишер, Р.А. Дьяченко, И.С. Лоба // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ).-2012.-№79(5).- Режим доступа: <http://ej.kubagro.ru/archive.asp?n=79>

References

1. D'jachenko, R. Modelirovanie sistem sbora i peredachi dannyh s primeneniem cvetnyh setej Petri [Jelektronnyj resurs] / R.A. D'jachenko, A.V. Fisher, V.V. Bogdanov // Fundamental'nye issledovanija. – 2013.- №11(6).-с. 1122-1126. –Rezhim dostupa: http://rae.ru/fs/?section=content&op=show_article&article_id=10002528
2. Laptev V.N. Voprosu povyshenija nadezhnosti sistemy sbora i hranenija hronologicheskikh dannyh [Jelektronnyj resurs] / Ju.N. Sopil'njak, R.A. D'jachenko, R.H. Bagdasarjan // Politematicheskij setевой jelektronnyj nauchnyj zhurnal KubGAU. – Krasnodar: KubGAU, 2014. №101(07), - 16s. – Rezhim dostupa: <http://ej.kubagro.ru/2014/07/pdf/162.pdf>.
3. Fisher, A.V. Organizacija hranenija hronologicheskikh dannyh v bazah dannyh sistem monitoringa i prognozirovanija [Jelektronnyj resurs] / A.V. Fisher, R.A. D'jachenko, I.S. Loba // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU).-2012.-№79(5).- Rezhim dostupa: <http://ej.kubagro.ru/archive.asp?n=79>