

МЕТОДИКА ОЦЕНКИ ЭФФЕКТИВНОСТИ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ ХРАНИЛИЩ

**А.Ю. Иванов, доктор технических наук, профессор;
В.С. Горшков. Санкт-Петербургский университет ГПС МЧС России**

Представлен подход к оценке эффективности распределенных информационных хранилищ, планируемых к применению в автоматизированных системах МЧС России. Предложены показатели эффективности, разработаны модельная и процедурная составляющие для расчета их значений, что в совокупности формирует методику оценки эффективности.

Ключевые слова: распределенное информационное хранилище, оценка эффективности, показатели эффективности, модели

EFFICIENCY ESTIMATION METHOD OF DISTRIBUTED DATA WAREHOUSES

A.Y. Ivanov; V.S. Gorshkov.
Saint-Petersburg university of State fire service of EMERCOM of Russia

This article represents approach to estimation of distributed data warehouses efficiency planned for using in automated systems of EMERCOM of Russia. Efficiency coefficients are suggested, model and procedure components for calculation of their values are developed, and as a result we have efficiency estimation method.

Key words: distributed data warehouse, efficiency estimation, efficiency coefficients, models

Эффект от внедрения информационных систем в практику автоматизированного управления силами и средствами МЧС России определяется возможностью решения должностными лицами органов управления задач, связанных с аналитической обработкой данных. Этот класс задач базируется на концепции оперативной аналитической обработки (ОАОД) и интеллектуального анализа данных (ИАД). Под ОАОД понимают технологии комплексного многомерного анализа данных для поддержки принятия решений, а ИАД основан на статистических методах и служит для выявления заранее неизвестных закономерностей. Основу ОАОД и ИАД составляют информационные хранилища (ИХ).

Используемые в настоящее время ИХ не способны обеспечить требования по оперативности и надежности выполнения пользовательских запросов в распределенных системах в силу своего централизованного построения. В то же время ориентация на аналитическую обработку накапливаемых данных определяет необходимость их использования в практике автоматизированного управления силами и средствами в чрезвычайных ситуациях (ЧС). Выход из сложившейся проблемной ситуации видится на пути организации *распределенных информационных хранилищ* (РИХ) автоматизированных систем (АС) МЧС России. Имеющийся научный и практический опыт построения и использования распределенных баз данных создает определенные предпосылки к формированию аналогичного класса информационных хранилищ.

Оценка эффективности РИХ может проводиться на основе известных подходов к оценке сложных систем: внешнего и внутреннего. При *внешнем подходе* РИХ рассматривается как элемент соответствующей АС, а ее эффективность оценивается через эффективность последней. С позиций *внутреннего подхода* РИХ представляется как самостоятельная система, а к её выходу (процессу выполнения запросов) со стороны АС предъявлены требования, удовлетворение которых является целью функционирования хранилища.

Наиболее существенными свойствами, которые отражают аспект практической применимости РИХ, следует считать его объем и реактивность, как временные параметры процесса реализации запросов. В этом случае внутренняя оценка эффективности сводится к разработке моделей и/или расчетных соотношений, позволяющих получать значения соответствующих показателей. Недостатком подобной оценки считается известная доля субъективизма, проявляющаяся при задании требований к значениям показателей эффективности, а также отсутствие учета влияния характера построения хранилища на процесс функционирования АС в целом.

Внешняя оценка эффективности ориентирована на получение значений показателей эффективности АС в целом с учетом характера структурного построения РИХ.

Показатели эффективности и модели для получения их значений

Эффективность, как степень соответствия системы цели реализуемой операции, является отражением комплекса свойств, характеризующих систему и/или операцию с разнообразных позиций.

Операцией, которую выполняет любая АС МЧС, является решение функциональных задач (ФЗ) по управлению силами и средствами в ЧС. Сложность процесса управления определяет, что требуемые решения не могут быть выработаны при решении одной задачи. Поэтому в существующих автоматизированных системах принята к использованию концепция комплексов задач для поддержки принятия управленческих решений. Концепция предполагает последовательное решение нескольких ФЗ, связанных по данным. Последнее означает, что результаты решения одних ФЗ выступают как исходные данные (часть исходных данных) для решения других задач.

Решение каждой ФЗ и комплекса задач в целом, как операция, реализуемая АС, характеризуется многими свойствами, к наиболее существенным из которых относятся своевременность (оперативность), надежность, достоверность и безопасность решения [1].

Каждое свойство может быть выражено количественно через один или несколько показателей эффективности. В таком случае интегральная оценка эффективности является результатом нахождения значения одного из показателей, признанного наиболее значимым, при выводе остальных в ограничения, или путем мультипликативной или аддитивной свертки всех показателей [1, 2].

Принятие решений на применение сил и средств МЧС в чрезвычайных ситуациях сопряжено с действиями должностных лиц органов управления в жестких временных рамках. В таких условиях доминирующим свойством процесса решения функциональных задач является его своевременность, которая может быть оценена различными показателями.

Процесс решения функциональных задач в аспекте времени решения относится к классу вероятностных операций, поскольку он подвержен воздействию случайных факторов: нагрузка на автоматизированную систему, количество исходных данных, подлежащих обработке, необходимость выработки нескольких вариантов решения и т.д. При таких условиях время решения ФЗ или комплекса в целом является случайной величиной. Наиболее простым показателем, характеризующим временные параметры процесса решения задач является среднее время решения (\bar{t}_p). Однако этот показатель мало информативен, в первую очередь из-за того, что не учитывает требований, предъявляемых к времени решения. Поэтому более предпочтительным является такой показатель, как вероятность своевременного решения задачи (комплекса задач) – $P(t_p \leq t_{\text{дон}})$, где $t_{\text{дон}}$ – допустимое время решения задачи, задаваемое директивно. Тогда критерием эффективности процесса решения функциональных задач выступает следующее правило:

$$\max P(t_p \leq t_{\text{дон}}).$$

В [3] излагается последовательность перехода от начальных моментов времени пребывания заявки на каждом из узлов сети к центральным моментам времени прохождения заявкой сети в целом. Таким образом, исходными данными для моделирования процесса решения ФЗ необходимо знать вероятностно-временные характеристики процесса обработки запросов к РИХ.

Этот процесс интерпретируется как обслуживание заявок в стохастической сети массового обслуживания (СеМО) [4]. Последняя представляет собой совокупность взаимосвязанных систем массового обслуживания (СМО), каждая из которых соответствует одному из элементов, участвующих в обработке запросов к РБД.

Основным показателем оперативности функционирования подобных сетей является функция распределения времени между поступлением и удовлетворением запроса и/или его моменты.

Интервалы между поступающими запросами к РИХ являются случайными величинами. Это обусловлено неопределенностью поведения объектов и субъектов управления и случайным характером порождаемых ими процессов. Время выполнения запросов также случайно и зависит от целого ряда факторов, таких как приоритет и тип запроса, место размещения требуемых для его выполнения фрагментов данных относительно инициатора запроса, объем требуемых фрагментов, число записей, релевантных условию запроса, загрузка и работоспособность серверов и коммуникационной системы.

Такое положение приводит к необходимости рассмотрения процесса обработки запросов к РИХ как случайного и исследования его в рамках теории вероятностей и теории массового обслуживания.

Стохастическая СеМО определяется следующей совокупностью характеристик [5-7]:

- 1) множеством СМО $\{S_1, S_2, \dots, S_n\}$, образующих сеть;
- 2) числом каналов K_1, K_2, \dots, K_n в системах S_1, S_2, \dots, S_n , соответственно;
- 3) матрицей траекторий движения заявок $R = \|r_{ij}\|$, где r_{ij} – номер СМО, на которую переходит заявка, продвигающаяся по i -му пути на j -й фазе обслуживания при детерминированной процедуре маршрутизации, или матрицей вероятностей перехода заявок из одной СМО в другую $P = \|p_{ij}\|$, где p_{ij} – вероятность того, что заявка, покидающая S_i , поступает в S_j ;
- 4) числом заявок, циркулирующих в замкнутой сети (Z);
- 5) интенсивностью источников заявок в разомкнутой сети $A = \{\lambda_i^{(k)}\}$, где i – тип заявки; k – категория срочности;
- 6) законами распределения времени $F_1(t), F_2(t), \dots, F_n(t)$ и дисциплинами обслуживания заявок в системах S_1, S_2, \dots, S_n .

Системы S_1, S_2, \dots, S_n и связи между ними определяют структуру сети. Интенсивность источников заявок $\lambda_{ij}^{(k)}$, интенсивность обслуживания $\mu_{ij}^{(k)}$, длины очередей и режим работы приборов характеризуют нагрузку и производительность СеМО.

Элементами, участвующими в обработке запросов к РИХ и представляющими различные СМО, являются (рис. 2):

- серверы базы данных, на которых размещаются фрагменты данных (СМО 1);
- серверы приложений, на которых происходит решение функциональных задач (СМО 2);
- автоматизированные рабочие места пользователей – должностных лиц органов управления (СМО 3);
- каналы передачи данных между ЦУКС, которые ведут информационный обмен (СМО 4).

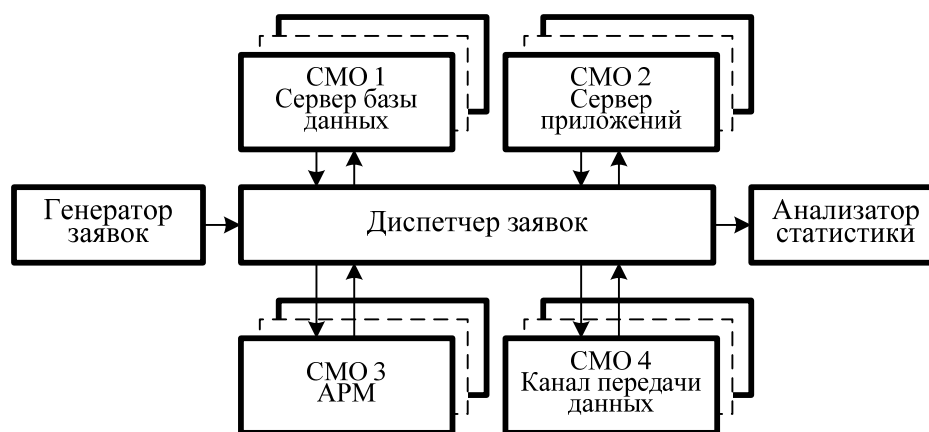


Рис. 2. Представление процесса обработки запросов к РИХ в виде обслуживания заявок в СеМО

Поиск реального закона распределения промежутков времени между моментами поступления двух соседних заявок к РИХ затруднен, так как при проектировании, как правило, известна лишь интенсивность решения задач управления.

Суммарный поток запросов от большого числа источников с произвольным случайным распределением подпотоков, доли которых в общем потоке примерно равны, стремится к простейшему. Следовательно, допущение о простейшем входном потоке или об экспоненциальном законе распределения длительности интервалов является справедливым.

При выборе законов распределения времени обслуживания заявок целесообразным представляется ориентация на типовые законы распределения времени выполнения запросов и решения задач. Это обусловлено тем, что запросы характеризуются случайными значениями длительности выполнения, зависящими от типа запросов, и хорошо определяются непрерывными распределениями.

Модель, представленная в виде СеМО, может быть исследована аналитическим или имитационным методами. Сравнительный анализ показывает, что метод имитационного моделирования при исследовании разработанной модели является более предпочтительным по следующим причинам:

- 1) большая размерность задачи, выступающая критическим фактором при построении аналитических моделей;
- 2) нарушение свойств стационарности, ординарности и отсутствия последействия входного потока в процессе обслуживания заявок в СМО, образующих СеМО, приводящее к тому, что входной поток заявок на очередной фазе обслуживания в сети (за исключением начальной) отличается от простейшего.

Процедура оценки эффективности

Основными особенностями АС МЧС России, связанными со спецификой их построения и функционирования, и подлежащими учету при оценке эффективности построения, выступают:

- многофункциональность АС, определяющая то, что запросы к РИХ имеют различную значимость и, соответственно, характеризуются разным уровнем требований к своевременности их обработки;
- сложность компьютерных сетей, составляющих техническую основу АС, обусловленная большим числом входящих в ее состав элементов, многообразием связей между ними и сложностью самого процесса обработки запросов;
- участие в процессе обработки запросов различных видов обеспечения (технического, программного, информационного и др.), которые могут существенно влиять на процесс обработки запросов и его показатели;
- функционирование АС в условиях воздействия случайных факторов;
- возможность возникновения различных ошибок, связанных как с локальным, так и с

распределенным характером обработки данных в АС.

Представленные модели позволяют сформировать процедуру их применения в целях получения значений показателей своевременности решения функциональных задач при использовании РИХ как информационного базиса. Эта процедура представляет собой последовательность этапов.

1. Определение исходных данных.

В качестве исходных данных выступают следующие:

– структура компьютерной сети АС, определяемая числом узлов сети (пунктов управления), составом узлов сети (числом серверов и АРМ), топологией локальных вычислительных сетей узлов, связностью узлов каналами передачи данных;

– вариант построения РИХ, определяющий состав, характеристики, связи и размещение фрагментов данных в сети, а также порядок обновления вторичных копий и данных обобщения;

– характеристики потока запросов, включающие в себя интенсивности и места возникновения запросов, множество требуемых для выполнения запросов фрагментов данных и порядок доступа к ним;

– характеристики пропускной способности каналов передачи данных и быстродействия элементов компьютерной сети;

– характеристики систем управления базами данных РИХ в узлах сети, определяющие дисциплину и технологию обслуживания запросов;

– требования к значениям показателей своевременности решения функциональных задач.

Источниками перечисленных исходных данных являются:

– техническое задание на разработку АС;

– описание организации информационной базы;

– паспортные данные технических и программных средств, предполагаемых к использованию в АС;

2. Настройка имитационной модели процесса обработки запросов на конкретный вариант компьютерной сети и размещения фрагментов распределенного информационного хранилища.

На этом этапе формируется структура СеМО, для чего указывается число систем массового обслуживания по их типам, и организуются связи между ними. Также задаются характеристики систем массового обслуживания, и уточняется процедура продвижения заявок, имитирующих запросы каждого типа, по сети массового обслуживания.

Дополнительно определяется количество модельных испытаний для обеспечения требуемой точности и надежности результатов имитационного моделирования.

3. Проведение эксперимента на имитационной модели.

Целью настоящего этапа является получение вероятностно-временных характеристик процесса обработки запросов к РИХ. В качестве таковых могут выступать:

– среднее время реализации запросов различных типов;

– функция распределения времени обработки запросов;

– другие числовые характеристики случайной величины – времени обработки запросов, в частности – начальные моменты четырех низших порядков.

Этот этап носит ярко выраженный технологический характер и сводится к выполнению на ЭВМ программы, описывающей имитационную модель.

4. Формирование графового представления процесса решения функциональных задач, как взаимосвязанной совокупности процессов обработки запросов к РИХ.

Этап связан с анализом алгоритмов решения функциональных задач. Указывается последовательность реализации запросов, отвечающих содержанию задач, и переходные вероятности между запросами (см. рис. 1). Значения переходных вероятностей определяются пропорционально интенсивностям инициации соответствующих запросов.

5. Определение временных параметров процесса решения функциональных задач.

Для этого выполняются следующие виды работ:

- проводится вычисление числовых характеристик случайной величины – времени решения, таких как ее четыре центральных момента;
- методом аппроксимации отыскивается вид функции распределения этой случайной величины;
- по найденной функции распределения и директивно задаваемому допустимому времени решения каждой задачи находится значение искомой величины – вероятности своевременного решения $P(t_p \leq t_{don})$.

6. Расчет значений показателей эффективности решения комплекса задач.

Выполнение этого этапа может осуществляться различными способами. Если функциональные задачи связаны по данным, то вычисления проводятся по аналогии с этапами 4 и 5 приводимой последовательности.

Если задачи не зависимы по данным, то необходимо перейти от векторного показателя к скалярному. Обычно такой переход осуществляется с помощью функций полезности и одного из методов свертки [1].

Представленные модели и описанная процедура в совокупности составляют методику оценки эффективности распределенных информационных хранилищ АС МЧС России

Методика может быть использована в интересах заказчика, разработчиков и конечных пользователей. Применение методики заказчиком позволяет оценить соответствие ожидаемых (реальных) показателей своевременности решения функциональных задач требованиям технического задания. Разработчики АС могут использовать методику для сравнения альтернативных вариантов построения РИХ при увязке последних с техническим и программным обеспечением, а также для выбора ограниченного числа вариантов, подлежащих детальной проработке. Дополнительно разработчики и ДЛ имеют возможность с помощью методики выявлять так называемые «узкие» места в проектах РИХ и определять направления их оптимизации.

Использование методики возможно на этапе эксплуатации АС при проведении реконфигурации компьютерной сети и РИХ по причине изменения информационных потребностей должностных лиц органов управления МЧС и возникновения по этой причине новых информационных потоков. Однако на данной стадии жизненного цикла математическое моделирование следует сочетать с натурными экспериментами с применением средств прототипирования и тестирования РИХ.

Литература

1. Волков П.И., Иванов А.Ю., Иванов Е.В. Построение критерия эффективности систем автоматизации управления. СПб.: ВАС, 1989. 74 с.
2. Штойер Р. Многокритериальная оптимизация. Теория, вычисления и приложения. М.: Радио и связь, 1992. 270 с.
3. Иванов А.Ю. Мобильные распределенные базы данных автоматизированных информационно-управляющих систем МЧС России: монография / под ред. В.С. Артамонова СПб.: Санкт-Петербургский университет Государственной противопожарной службы МЧС России, 2008. 152 с.
4. Клейнрок Л. Теория массового обслуживания. М.: Машиностроение, 1979. 432 с.
5. Иванов Е.В. Имитационное моделирование средств и комплексов связи и автоматизации. СПб.: ВАС, 1992. 206 с.
6. Черненький В.М. Разработка САПР: В 10 кн. Кн. 9. Имитационное моделирование: Практ. пособие / под ред. А.В. Петрова М.: Высш. шк., 1990. 87 с.
7. Альянах И.Н. Моделирование вычислительных систем. Л.: Машиностроение. Ленингр. отд-ние, 1988. 223 с.