

УДК 004.9

## ПРОГРАММНЫЙ КОМПЛЕКС РАСЧЕТА НАГРУЗКИ НА ОСНОВЕ ДАННЫХ МОНИТОРИНГА

**Н.Б. Баканова**, заведующая сектором ИППИ РАН, к.т.н.; nina@iitp.ru

**Д.В. Волчков**, аспирант ИППИ РАН

**Ключевые слова:** информационные системы, организационное управление, мониторинг систем, каналы связи, расчет сетевых нагрузок.

**Введение.** Ключевой вопрос создания распределенных информационных систем организационного управления – реализация информационного взаимодействия с территориально удаленными подразделениями и подведомственными организациями [5]. В организационном управлении составляющими информационных потоков являются: документы, которые могут обладать значительным объемом; указания по исполнению поручений, содержащихся в документах; информация о выполнении поручений.

В соответствии со спецификой прикладные задачи организационного управления требуют оперативного обмена данными между узлами распределенной сети и стабильного выполнения интерактивных режимов поддержки управленческой деятельности.

При проектировании ведомственных сетей передачи данных (ПД), объединяющих удаленные организации и подразделения, используются современные каналы связи, обладающие, как правило, достаточно высокой пропускной способностью. Однако нагрузка на различные сегменты сети, создаваемая прикладными системами, существенно зависит от функциональных задач подразделений и, соответственно, меняется при реструктуризации организационного управления. Кроме того, помимо трафика прикладных информационных систем, те же каналы используются для передачи данных IP-телефонии, видеоконференций, FTP (File Transfer Protocol), голосовой почты и т.п., что нередко приводит к неравномерной загрузке каналов связи и, как следствие, к нестабильности (задержкам и отказам) при работе интерактивных режимов прикладных систем.

Обеспечить стабильность информационных режимов можно путем создания отдельных подканалов ПД прикладных систем с возможностью корректировки их характеристик в со-

ответствии нагрузками на сегменты сети. Для расчета характеристик целесообразно использовать реальные данные, которые накапливаются в подсистемах мониторинга функционирования, входящих в комплекс сопровождения жизненного цикла прикладных систем организационного управления.

**Постановка задачи.** Схема информационной системы организационного управления крупной иерархической организации представлена на рис. 1. На схеме показаны внешние и внутренние информационные потоки, поступающие в узлы информационной системы (УИС). Центральным узлом является головная организация, к периферийным узлам относятся отраслевые службы, удаленные подразделения и другие подведомственные организации, входящие в структуру организационного управления.

В головной организации (центральный УИС) располагается сервер с центральной базой данных (БД), в периферийных УИС находятся локальные серверы БД.

В общем случае требования к пропускным способностям каналов между серверами распределенной системы различны для всех сегментов сети и определяются объемами информации,

передаваемыми между отдельными территориально удаленными узлами. Требуется рассчитать характеристики участков сети, обеспечивающие прохождение информационных потоков через данный узел.

Расчет сетевой нагрузки на основе данных мониторинга позволит прогнозировать и обоснованно скорректировать характеристики пропускной способности функционирующих каналов связи при модификации структуры организационного управления или изменении информационной нагрузки на периферийные УИС. По результатам расчетов может возникнуть необходимость корректировки требований к пропускной способности канала связи организации.

**Разработка программного комплекса анализа нагрузки на УИС.** Для корректировки требований к пропускной способности каналов ПД и прогнозирования их изменений при модификации структуры потоков разработан программный комплекс анализа и расчета параметров сети на основе данных подсистемы мониторинга функционирования прикладных задач. В программный комплекс включены модуль сбора и передачи информации о нагрузке и модуль анализа данных мониторинга.

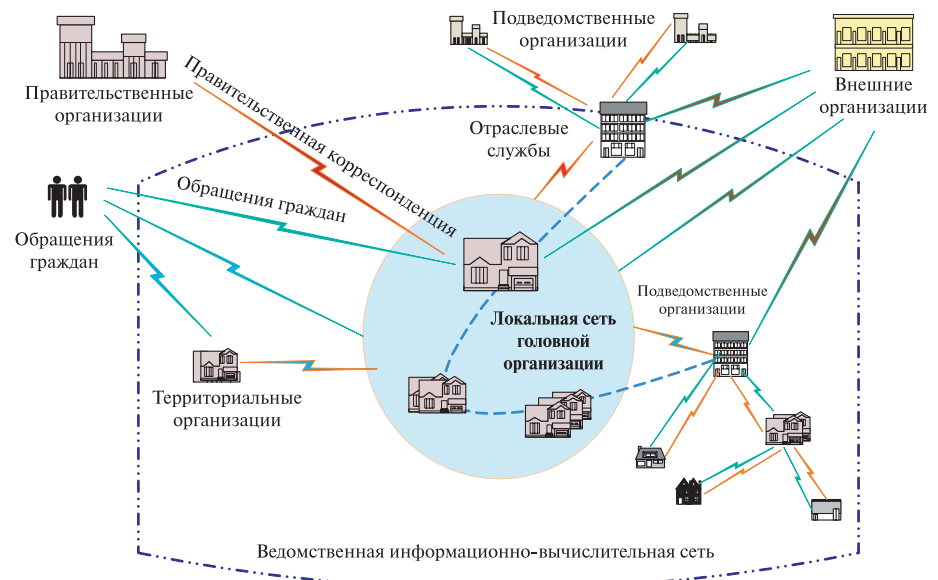


Рис. 1

Исходными данными для расчетов является информация о работе, накопленная в БД мониторинга информационной системы. БД содержит следующую информацию о работе: адрес рабочей станции клиента, идентификатор запрашиваемого объекта (карточка документа, поисковый запрос, справочник и т.д.), время работы с объектом, объемы исходящего и входящего трафика.

Иллюстрация принципов работы программного комплекса, представленная на рис. 2, включает следующие элементы:

- центральный сервер, УИС головной организации;
- локальные серверы, рабочие станции, каналы ПД и маршрутизаторы, расположенные в каждом периферийном УИС;
- выделенные VPN-каналы, объединяющие все УИС.

На рис. 2 показано, что подсистема мониторинга включает программное обеспечение (ПО) центрального УИС и ПО периферийных УИС. В ПО центрального УИС включен модуль анализа данных мониторинга, в ПО периферийных УИС – модуль сбора и передачи информации о нагрузке.

Для обеспечения работы программного комплекса периферийные УИС содержат следующие компоненты:

- локальная БД мониторинга, в которой хранятся данные мониторинга функционирования прикладных задач периферийного узла;
- ПО мониторинга периферийного УИС, включающего модуль сбора и передачи информации о нагрузке.

**Модуль сбора и передачи информации о нагрузке.** Модуль собирает данные мониторинга со всех локальных серверов периферийных УИС и с помощью сервисов протокола FTP передает их на центральный сервер для последующего анализа и обработки. В качестве коммуникативного формата выбран текстовый формат XML (Extensible Markup Language).

Основной идеей разработанного комплекса является пересылка данных сеансов работы с информационной системой, трансформированных из внутреннего представления БД в XML файлы [2]. Модуль интегрирован в ПО мониторинга периферийных УИС, не являющихся окончательными узлами в иерархии. Модуль осуществляет сбор и отправку информации о нагрузке на сегмент сети между данным узлом и ближайшими по иерархии нижестоящими УИС.

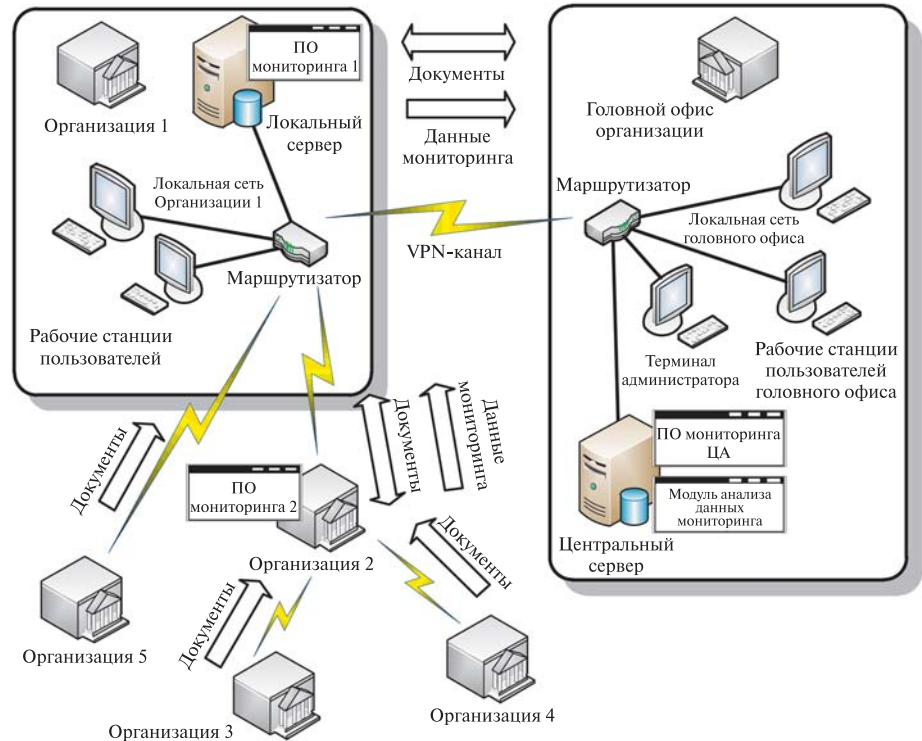


Рис. 2

Для обеспечения работы программного комплекса в центральном УИС содержится:

- центральная БД мониторинга, в которой хранятся данные мониторинга, собранные со всех локальных серверов;
- ПО мониторинга центрального аппарата, в которое входит модуль анализа данных, предназначенный для расчета, анализа и прогнозирования изменения основных параметров сети ПД.

В ПО мониторинга центрального УИС также интегрирован модуль сбора информации о нагрузке, поступающей со всех периферийных УИС.

**Модуль анализа данных мониторинга.** Модуль предназначен для выполнения функций, необходимых для расчета характеристик канала на основе данных мониторинга:

- вычисление среднеквадратичных отклонений времени обслуживания заявок и среднего времени отклика системы;
- расчет и прогнозирование изменений основных параметров сети;
- отображение выходных данных в виде сводок, справок, специализированных интерфейсов.

Параметры для любого сегмента сети рассчитываются с помощью математического аппарата теории массового обслуживания [4]. При расчете любой узел сети рассматривался в виде про-

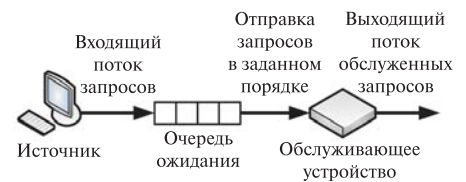


Рис. 3

стейшей системы (рис. 3) с очередь к обслуживающему устройству (ОУ).

Цель расчетов – получение информации о скорости поступления запросов в очередь, времени обслуживания запросов в ОУ на входе в систему, общем количестве ожидающих запросов, времени ожидания запросов в системе. При этом важно знать средние значения параметров и диапазон их изменений, оцениваемый значениями среднеквадратичных отклонений.

Будем предполагать, что рассматриваемая на рис. 1 распределенная информационная система удовлетворяет следующим условиям: изменение скорости поступления запросов соответствует закону Пуассона, в модели имеется одно ОУ. Для расчета параметров системы с одним ОУ используются следующие формулы [6]:

$$\rho = \lambda T_s; \tag{1}$$

$$q = \frac{\rho}{1 - \rho}; \tag{2}$$

$$\omega = \frac{\rho^2}{1 - \rho}; \tag{3}$$

$$q = \omega + \rho; \tag{4}$$

$$T_q = \frac{T_s}{1-\rho}; \tag{5}$$

$$T_w = \frac{\rho T_s}{1-\rho}; \tag{6}$$

$$\sigma_q = \frac{\sqrt{\rho}}{1-\rho}; \tag{7}$$

$$\sigma_{T_q} = \frac{T_q}{1-\rho}; \tag{8}$$

$$m_{T_q}(r) = T_q \ln \frac{100}{100-r}; \tag{9}$$

$$m_{T_w}(r) = T_w \ln \frac{100\rho}{100-r}, \tag{10}$$

где  $\lambda$  – средняя скорость поступления запросов;  $w$  – среднее число запросов, находящихся в очереди;  $T_w$  – среднее время ожидания запросов в очереди;  $T_s$  – среднее время обработки запросов ОУ;  $q$  – среднее число запросов, находящихся во всей системе (ожидающие в очереди и обрабатываемые ОУ);  $T_q$  – среднее время пребывания запросов в ОУ (среднее время ответа);  $\sigma_q$  и  $\sigma_{T_q}$  – среднеквадратичные отклонения;  $\rho$  – утилизация (степень загрузки) ОУ, определяемая долей общего времени, в течение которой ОУ было занято. Выражения вида  $m_x(r)$  означают, что  $x$  меньше или равно  $m_x(r)$  в  $r$  процентах случаев. Они предназначены для оценки параметров системы, необходимых для обеспечения заданных значений среднего времени ответа системы  $T_q$  и среднего времени обслуживания запросов  $T_w$ .

При вхождении в режим насыщения системы размер очереди резко возрастает, стремясь к бесконечности при  $\rho = 1$ . Исходя из этого, допустимой будем считать скорость поступления данных на ОУ, не превышающую 70–90% от теоретического максимума. Норма поступления заявок не должна быть больше общего уровня обслуживания, т.е.  $\lambda < \rho$  или  $\lambda/\rho < 1$ . Величина  $\lambda/\rho$  характеризует интенсивность работы системы [2].

Поскольку в информационной системе используются запросы разного типа с точки зрения времени обработки и объема данных (запрос данных об исполнении документа, поиск в БД, обращение к справочникам и т.п.), то при расчете нагрузок на ОУ учитываются заявки  $M$ -типов  $\{1, 2, \dots, M\}$ , поступающие с интенсивностями  $\{I_1, I_2, \dots, I_M\}$ . Загрузка ОУ потоком  $i$ -типа составляет

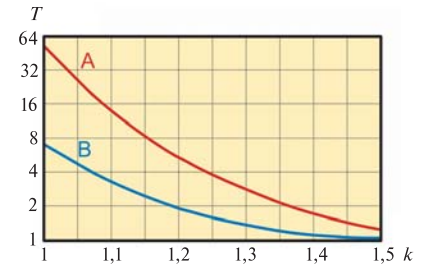
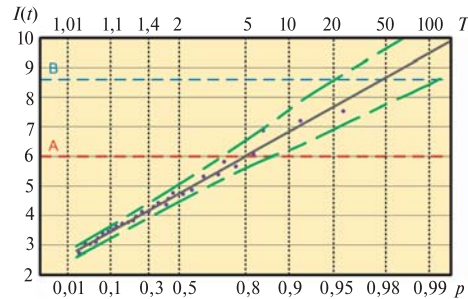


Рис. 4

$\rho_i = \lambda_i T_{Si}$ , а общая загрузка  $\rho = \sum_{i=1}^M \rho_i$ .

Другие характеристики обслуживания  $q_i, w_i, T_{qi}, T_{wi}$  определяются для каждого  $i$ -потока в отдельности по формулам (2), (3), (5) и (6).

Средние значения времени ожидания и обслуживания по одной заявке из суммарного потока в системе связаны со средними количествами заявок в очереди соотношениями:

$$\bar{T}_w = \sum_{i=1}^M \frac{\lambda_i T_{wi}}{\lambda} = \frac{1}{\lambda} \sum_{i=1}^M \omega_i = \frac{\bar{\omega}}{\lambda}; \tag{11}$$

$$\bar{T}_q = \frac{1}{\lambda} \sum_{i=1}^M q_i = \frac{\bar{q}}{\lambda}, \tag{12}$$

где  $\lambda = \sum_{i=1}^M \lambda_i$ ,  $\lambda_i / \lambda$  – вероятность того, что поступила заявка  $i$ -типа;  $\bar{\omega}$  – средняя длина очереди заявок всех типов;  $\bar{q}$  – среднее число заявок всех типов.

Приведенный подход позволяет рассчитать средние значения параметров загрузки сегментов сети данными прикладных систем организационного управления и лежит в основе разработанного программного комплекса.

Кроме того, для распределенных систем организационного управления важно определить, как часто будут происходить перегрузки канала после подключения дополнительных рабочих мест или при увеличении инфопотоков, насколько необходимо расширить пропускную способность канала для обеспечения приемлемой отказоустойчивости канала. Для определения пропускной способности канала, обеспечивающей приемлемый уровень отказоустойчивости, реализован модуль прогнозирования нагрузки на сеть. Модуль выполняет расчеты на основе данных мониторинга.

Для прогнозирования величины периода отказов канала  $T$  и вероятности  $p$  при том или ином объеме передаваемой информации  $I(t)$  в модуле использован математический аппарат статистики экстремальных значений [1, 3]. Результаты вычислений отображаются в виде графиков. На рис. 4 проиллюстрирован

пример выдачи результата вычислений:  $A$  – интенсивность трафика 6 Мбит/с;  $B$  – 8,5 Мбит/с.

Разработанный модуль позволяет определить, какое влияние окажет на отказоустойчивость сети изменение объема передаваемой информации в  $k$  раз. На основании анализа данных можно принять решение о необходимой пропускной способности канала, удовлетворяющей заданным требованиям к значениям периода возможной перегрузки канала.

**Заключение.** Разработанный программный комплекс позволяет оценить: характеристики каналов связи крупной управленческой организации со сложной иерархической структурой, в которой функционирует распределенная информационная система; периодичность пиковых нагрузок на канал, соединяющий удаленные организации; прогнозировать изменения параметров сети при подключении новых пользователей, изменении объемов документопотоков и т.п. Исходными данными для анализа служит информация, накопленная в БД мониторинга информационной системы. При относительной простоте вычислений данный подход позволяет получить достаточно точные характеристики каналов связи.

Программный комплекс может быть использован при проектировании информационных систем организационного управления, а также включен в комплекс обеспечения жизненного цикла систем для отслеживания таких изменений нагрузки каналов связи, как расширение функциональности подсистемы мониторинга.

ЛИТЕРАТУРА

1. Гумбель Э. Статистика экстремальных значений / Пер. с англ. В.Ю. Татарского под ред. Д.М. Чибисова. – М.: Мир, 1965. – 450 с.
2. Кульгин М.В. Технологии корпоративных сетей. Энциклопедия. – СПб: Изд – во «Питер», 2000. – 704 с.

3. **Бахвалов Л.А., Репин Д.С., Филаретов Г.Ф.** Алгоритм анализа трафика в корпоративных компьютерных сетях на основе статистики экстремальных значений. // Программные продукты и системы. – 2008. – №3(83). – С. 8–10.
4. **Баканова Н.Б., Атанасова Т.** Моделирование интегрированных ресурсов в распределенных информационных сетях// Электросвязь. – 2008 – № 3. – С. 41–44.
5. **Волчков Д.В., Сурпин В.П.** Расчет, анализ и прогнозирование изменения основных параметров корпоративной сети для распределенной информационной системы документального обеспечения управления. // Информационные технологии и системы (ИТиС'11): Сборник трудов конференции. – М.: ИППИ РАН, 2011. – С. 276–281.
6. **Шнепс М.А.** Системы распределения информации: Справ. пособие. – М.: Связь, 1979. – 344 с.

*Получено 25.0212.*