

**В. А. Богатырев**, д-р техн. наук, проф., e-mail: vladimir.bogatyrev@gmail.com,  
**С. А. Паршутина**, аспирант, e-mail: svetlana.parshutina@gmail.com,  
Санкт-Петербургский национальный исследовательский университет  
информационных технологий, механики и оптики

## Анализ влияния кратности резервирования многопутевых передач на вероятность их своевременного обслуживания

*Исследованы возможности повышения вероятности своевременного безошибочного обслуживания в системах многопутевой передачи данных в результате резервированного выполнения критичных к задержкам запросов. На основе результатов имитационного моделирования проанализирована эффективность резервированных многопутевых передач в распределенных компьютерных системах, в которых реализуется доставка копий пересылаемых по многим путям пакетов в один адресуемый узел и/или многопутевое распределение запросов к одному или нескольким узлам обработки, готовым к их обслуживанию.*

**Ключевые слова:** резервирование, многопутевая передача, распределение запросов, критичность к времени пребывания

### Введение

Ключевой проблемой проектирования распределенных компьютерных систем и сетей является обеспечение надежности [1–3], безопасности [4–6] и производительности обработки, передачи и хранения данных при взаимодействии узлов через сеть [7–9]. Для реализации надежного взаимодействия узлов распределенной компьютерной системы необходимо обеспечить не только ее структурную надежность и отказоустойчивость за счет резервирования путей, обеспечивающих связанность ее узлов, но и функциональную надежность (устойчивость функционирования) процессов межмашинного обмена через сеть в условиях сбоев, отказов и внешних деструктивных воздействий. Особенно остро проблема обеспечения надежности взаимодействия стоит для систем реального времени, в которых решаются задачи с высокими требованиями к своевременности выполнения и нередко отсутствует возможность повторных передач. В таких системах формируются жесткие требования к задержкам обслуживания запросов; несоблюдение этих требований приводит к нарушениям условий своевременности выполнения запросов при межмашинном обмене.

Повышение качества обслуживания потоков при взаимодействии компьютерных узлов через сеть может достигаться на основе многопутевых передач (в том числе на основе многопутевой маршрутизации) [10–15].

Под маршрутом (путем) понимается последовательность узлов, задействованных при передаче пакетов от узла-источника к узлу-приемнику (адресату). При многопутевой передаче предусматривается наличие двух и более маршрутов между конечными узлами, при этом маршруты могут быть непересекающимися (dis-joint) или могут иметь общие узлы и линии связи. Многопутевая маршрутизация эффективна при высоких нагрузках в сети [16], и значительная часть разрабатываемых моделей нацелена на решение задачи балансировки нагрузки [17–19] для предотвращения и устранения перегрузок и оптимального использования сетевых ресурсов. Превалирует подход, когда распределение потоков данных по разным маршрутам происходит в зависимости от пропускной способности и загрузки линий связи, при этом часто решают задачу минимизации максимальной загрузки каналов сети [20], в программно-конфигурируемых сетях — минимизации очередей в маршрутизаторах [16]. Кроме того, получили распространение модели с выравниванием нагрузки в сети, учитывающие надежность путей на основе критериев: коэффициента готовности каналов [21]; вероятностей их отказов, известных априори [22].

Многопутевая маршрутизация также широко используется для повышения надежности и отказоустойчивости передач в условиях отказов узлов и линий связи, когда один маршрут выступает в роли основного, а остальные — в роли резервных (горячего резерва) [23]. При

отказе основного пути происходит перенаправление потоков по альтернативным маршрутам без потери времени на поиск нового пути.

Распределение пакетов одного потока или потоков от одного узла-источника по разным маршрутам способствует рациональному использованию сетевых ресурсов, предупреждению недоставки пакетов вследствие возникновения перегрузок в сети и потери времени на поиск новых маршрутов в случае отказа основного и, в результате, увеличению вероятности и уменьшению времени передачи пакетов адресату (адресатам).

Известные подходы к повышению надежности систем с многопутевой маршрутизацией связаны со структурным резервированием путей, в то время как функциональная надежность взаимодействия компьютерных систем может быть повышена за счет резервированных передач, когда при использовании существующей совокупности маршрутов реализуется резервирование пересылаемых пакетов (создание их копий), передаваемых по разным физическим путям.

В рамках использования механизмов многопутевых передач в целях повышения надежности взаимосвязи компьютерных систем через сеть в данной статье исследуются возможности повышения вероятности своевременного обслуживания критичных к задержкам запросов на основе их резервированного обслуживания.

Как показано в работах [24–26], снизить среднее время ожидания и повысить вероятность своевременности выполнения критичных к задержкам запросов удастся при их резервированном обслуживании. Исследование многоканальных систем массового обслуживания (СМО) с общей очередью, в которых копии запросов, выполняемых в разных каналах, создаются только при наличии свободных каналов, проведено в работах [27, 28]. Такая дисциплина обслуживания не ориентирована на принципиальное для систем реального времени повышение вероятности своевременного выполнения всех критичных к задержкам запросов. Резервированное обслуживание запросов с заданием числа создаваемых копий в зависимости от их критичности к задержкам ожидания в очередях для одноуровневых кластерных систем, представляемых группой одноканальных СМО с локальными очередями, предложено и исследовано в работах [24–26]. Резервированное обслуживание критичных к ожиданию запросов считается успешным,

если своевременно выполнена хотя бы одна копия запроса. Применение резервированного обслуживания к системам передачи данных с резервированием каналов исследовано в работе [7]. В работах [29, 30] предложено развитие концепции резервирования процессов обработки и передачи данных применительно к многоуровневым системам, предполагающим многоэтапное (последовательное) выполнение копий запросов в узлах разных уровней. Резервированное многоэтапное обслуживание критичных к выполнению запросов считается успешным, если хотя бы для одной копии суммарное (накопленное) время ожидания в последовательно обслуживающих ее узлах (представляемых одноканальными СМО) не превышает заданную предельно допустимую задержку. Для одно- и многоуровневых кластеров эффективность резервированных процессов обслуживания удастся повысить при приоритезации копий запросов в зависимости от их критичности к ожиданию в очередях [30] и при уничтожении неактуальных копий [31, 32] ("просроченных" копий, находящихся в очередях, когда одна из них уже обслужена).

При многопутевой резервированной передаче предлагается создание копий пакетов при их отправке по разным маршрутам с кратностью резервирования (копирования), зависящей от критичности пакетов к допустимым задержкам [24–26, 30–32].

В системах распределенной обработки данных возможна организация взаимодействия, включающего передачу данных (сообщений, пакетов) и распределение (перераспределение) запросов [12, 15, 33, 34]. Таким образом, возможны два направления использования резервированных многопутевых передач:

- $S_1$  — для резервированной передачи пакетов данных по нескольким путям в один узел-приемник, при этом пересылаемый пакет считается переданным успешно, если за заданное время хотя бы одна его копия доставлена адресату;
- $S_2$  — для резервированного распределения запросов по нескольким путям к одному или более (согласно заданному числу) узлам-приемникам из некоторого множества возможных адресатов, при этом запрос считается успешно выполненным, если за заданное время хотя бы одна его копия доставлена в один или в заданное число узлов, готовых к ее выполнению.

**Цель работы** — исследование возможностей повышения надежности и своевременности взаимодействия компьютерных систем при многопутевой резервированной передаче в зависимости от кратности резервирования формируемых копий пакетов при их отправке в узел-приемник (узлы-приемники) по разным маршрутам.

Следует подчеркнуть, что резервированное обслуживание запросов (в том числе на передачу данных через сеть) связано с разрешением технического противоречия, так как, с одной стороны, обслуживание части или всех копий запросов приводит к увеличению загрузки, а значит, к росту времени ожидания копий запросов в очередях, но, с другой стороны (в силу стохастичности задержек передачи по разным путям), существует вероятность того, что некоторая копия будет обслужена значительно быстрее остальных. Если при межмашинном обмене достаточно своевременности обслуживания хотя бы одной копии запроса (пакета), выполнение этого условия приводит к повышению вероятности своевременного обслуживания запросов и уменьшению среднего времени пребывания запроса в сети. Резервированная пересылка пакетов через сеть в условиях возможных ошибок передач усиливает эффект повышения вероятности безошибочности и, следовательно, своевременности передач через сеть.

Разрешение указанного технического противоречия связано с поиском оптимальной кратности резервирования передач в зависимости от интенсивности потоков запросов и их критичности к задержкам обслуживания (к предельно допустимому времени пребывания запросов в сети).

Для достижения поставленной цели впервые решаются задачи анализа эффективности резервированных многопутевых передач в распределенных компьютерных системах, в которых реализуется доставка копий пересылаемых по многим путям пакетов в один адресуемый узел и/или многопутевое распределение запросов к одному или нескольким узлам обработки, готовым к их обслуживанию.

Метод решения задач исследования основан на применении имитационного моделирования.

### **Имитационные модели многопутевых резервированных передач**

Предлагаемые модели разработаны в среде имитационного моделирования AnyLogic 7

Professional в рамках дискретно-событийного подхода, при котором моделируемые процессы представляют в виде последовательности операций, производимых над некими сущностями (например, заявками, требованиями, запросами) с использованием специальных компонентов модели из библиотеки "Process Modeling Library" — "блоков" ("blocks").

Процессы многопутевых резервированных передач данных по одному адресу и распределения запросов по адресам входящих в группу серверов в виде потоковых диаграмм (flowcharts) показаны соответственно на рис. 1, *а* и *б*, при этом выделены этапы:

- создания запросов и их копий в источнике ("блоки" 'source', 'req\_clone') с подготовкой к отправке, включающей упаковку формируемых пакетов в кадр используемого протокола и их передачу в канал ('buffer', 'delay');
- выбора пути следования запроса к узлу-получателю из пяти доступных маршрутов, организованных в виде виртуальных каналов, в "блоке" 'path1\_5';
- обслуживания запроса в коммуникационных узлах (маршрутизаторах) 'router1', 'router2', ... с возможностью его удаления из системы в "блоках" 'loss1', 'loss2', ..., 'loss5\_N';
- обслуживания запроса узлом-получателем ('server', 'server1'—'server5') с возможностью его удаления из системы в "блоках" 'P0', 'P0\_1'—'P0\_5', 'timeout', 'timeout1'—'timeout5', 'relevant', 'relevant1'—'relevant5';
- выхода запроса из системы ('serviced', 'serviced1'—'serviced5').

Модели на рис. 1 представляют ситуации, в которых имеется пять вариантов (маршрутов) следования запросов от источника до адресуемого сервера (рис. 1, *а*) или пяти серверов кластера (рис. 1, *б*) с четырьмя промежуточными узлами на каждом пути. Маршруты не пересекаются, т. е. не имеют общих узлов.

Задаются следующие параметры модели: интенсивность потока запросов  $\Lambda$ ,  $c^{-1}$ ; кратность (число создаваемых копий) резервирования передач  $k$ ; число пересылаемых запросов  $N$ , шт.; максимальное (предельно допустимое) время ожидания  $t_0$ ,  $c$ ; пропускные способности линий связи  $BW$ , бит/с; протяженности сегментов линий связи  $L$ , м; задержки в коммуникационных узлах (маршрутизаторах)  $RD$ ,  $c$ ; задержки в узлах назначения (серверах)  $SD$ ,  $c$ ; вероятности (или условия) потерь запросов при передаче  $LR$ ; вероятности (интенсивности)

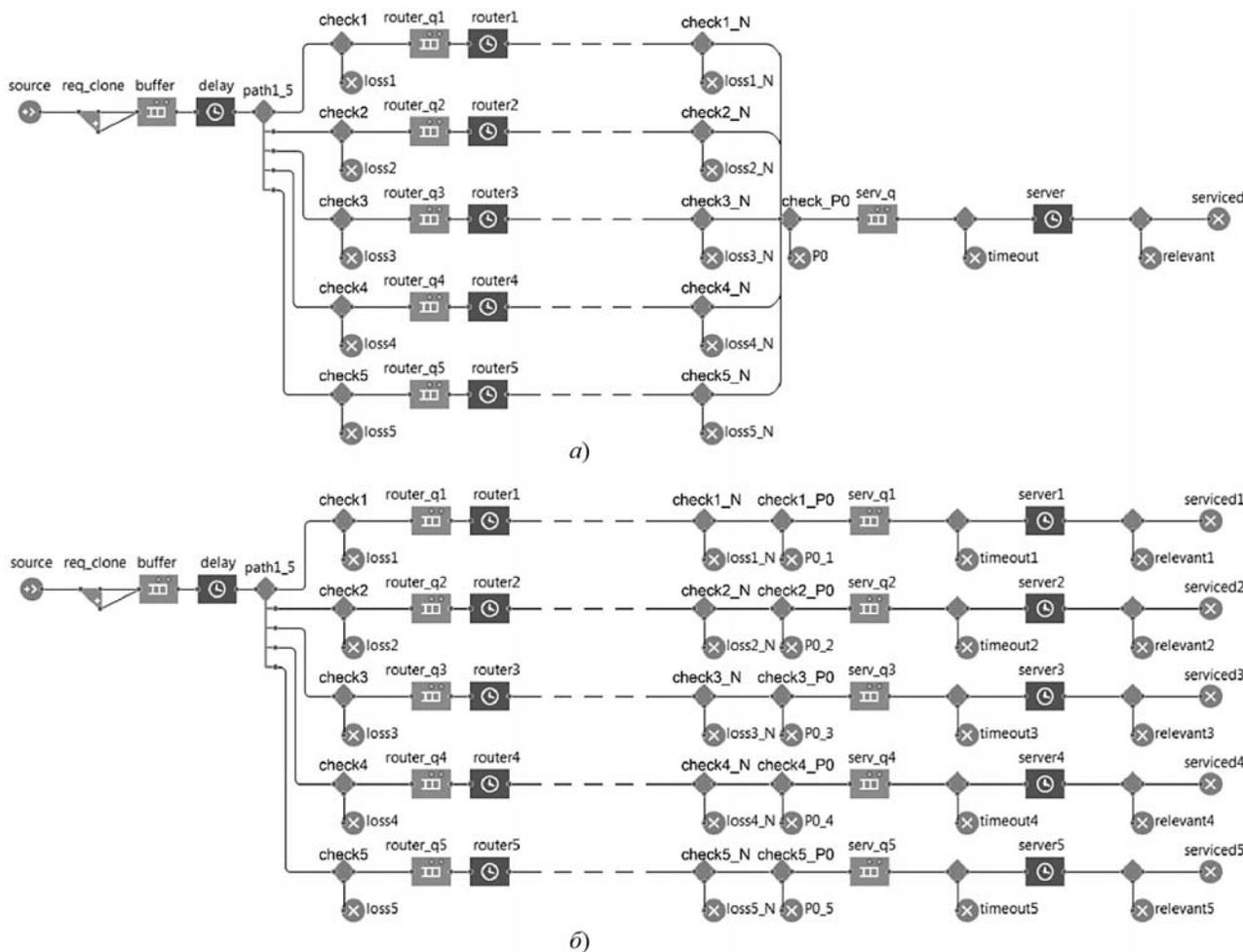


Рис. 1. Фрагмент модели сети с многопутевой передачей данных по одному адресу (а) и многопутевым распределением запросов (б)

битовых ошибок  $BER$ ; вероятности готовности узлов назначения (серверов) к обслуживанию запросов  $P_0$ ; вероятность (или алгоритм) обнаружения битовых ошибок в запросах коммуникационными узлами  $P_1$ ; размеры запросов (пакетов данных)  $RS$ , бит. В модели также задается: максимальное число и размер буферов выходного порта источника, входных портов промежуточных узлов и узлов назначения. Максимальное число буферов определяется соответственно емкостью (максимальной длиной) очередей — "блока" 'buffer', "блоков" 'router\_q1', 'router\_q2', ... —  $RQ$  (шт.) и "блоков" 'serv\_q', 'serv\_q1'—'serv\_q5' —  $SQ$  (шт.). Для простоты будем считать, что размеры буферов совпадают с максимальным размером пересылаемых пакетов (кадров).

Перемещение запросов от одного "блока" к другому с точки зрения модельного времени происходит мгновенно, за исключением "блоков" типа  $Delay$ , которые запросы покидают по истечении заданного времени или при выпол-

нении некоторого условия. Примем, что задержка запросов при передаче по линиям связи либо отсутствует, либо учитывается в очередях входных интерфейсов следующих узлов. В "блоке" 'source' скорость создания запросов определяется через "интенсивность прибытия" (rate), или экспоненциально распределенное число запросов в единицу времени; "время между прибытиями" (interarrival time), заданное константой или функцией, например, вероятностного распределения, по умолчанию — экспоненциального; расписание (schedule), в том числе загружаемое из базы данных.

Каждому запросу присваивается порядковый номер  $reqNumber$  от 1 до  $N$  — максимального числа поступающих в систему запросов. Начало его пребывания в системе  $startTime$  — текущее время, размер в битах  $reqSize$  (например, размер кадра формата Ethernet DIX (II) в компьютерных сетях, построенных на основе технологий семейства Ethernet, от 512 до 12 144 бит). При кратности резервирования

передач  $k > 1$  в "блоке" 'req\_clone' создается  $(k - 1)$  идентичных копий запроса. Запросы и их копии помещаются в "очередь" 'buffer', после чего по одному попадают в "блок" 'delay', в котором задерживаются на время  $reqSize/BW$ . Далее запросы распределяются по маршрутам ('path1\_5') на основе алгоритма кругового обслуживания (Round Robin) и либо удаляются из системы с некоторой вероятностью или по срабатыванию некоторого условия ('loss1', 'loss2', ..., 'loss5\_N'), либо попадают в очереди промежуточных узлов 'router\_q1', 'router\_q2', ... и ожидают обслуживания.

Поток запросов — однородный; дисциплина обслуживания — беспriorитетная; выбор запросов осуществляется в порядке их поступления в очереди 'router\_q1', 'router\_q2', ..., 'serv\_q', 'serv\_q1'—'serv\_q5' (FIFO).

Задержка обслуживания в "блоках" 'router1', 'router2', ..., 'server', 'server1'—'server5':  $RD = SD = t_{prg} + buffTime + procTime$ , где  $t_{prg}$  — время распространения сигнала;  $buffTime$  — время поступления запроса в принимающий буфер;  $procTime$  — время обработки запроса (задержка коммутации).

Время распространения сигнала  $t_{prg} = L/S$ , где  $L$ , м — протяженность сегментов линий связи;  $S$ , м/с — скорость распространения электромагнитных волн, которая зависит от физической среды и колеблется от 0,6 до 0,9 скорости света в вакууме [35], для витой пары  $\approx 0,7$ .

Время буферизации  $buffTime = (reqSize + 64)/BW + 0,96 \cdot 10^{-6}$ , включая преамбулу кадра в 64 бит и время межкадрового интервала  $0,96 \cdot 10^{-6}$  с,  $BW$  — пропускная способность сегмента линии связи, бит/с.

Достигшие узла назначения запросы либо обслуживаются ('serviced', 'serviced1'—'serviced5'), либо досрочно покидают систему в случае их неактуальности или неготовности серверов ('P0', 'P0\_1'—'P0\_5'). Запрос теряет актуальность, если время его пребывания в системе (разница между текущим модельным временем и  $startTime$ ) превысило максимальное время ожидания ("блоки" 'timeout', 'timeout1'—'timeout5') или уже была обслужена ранее другая его копия ("блоки" 'relevant', 'relevant1'—'relevant5').

Максимальное время ожидания  $t_0$  задается исходя из требований вычислительного процесса (конкретного приложения/программы).

После выхода каждого запроса из системы осуществляется проверка условия окончания эксперимента: сумма доставленных в узлы на-

значения и досрочно покинувших систему запросов достигает значения  $kN$ .

Результаты эксперимента — статистика по числу успешно пересланных запросов, досрочно покинувших систему запросов, минимальному, среднему и максимальному времени пребывания запросов в системе.

Исследование влияния кратности резервирования передач на вероятность своевременного обслуживания запросов при учете интенсивности потоков запросов (трафика) и их критичности к времени пребывания в сети (ограничений на предельно допустимое время пребывания  $t_0$ ) предполагает анализ результатов многочисленных имитационных экспериментов. Инструментальные средства поддержки проведения серии имитационных экспериментов с варьирующимися значениями параметров модели были предложены в работе [36].

### Исследование многопутевых резервированных передач, критичных к времени ожидания

Проанализируем эффективность резервированного взаимодействия в инфокоммуникационной системе в случае резервированной передачи данных по нескольким путям в один узел-приемник ( $S_1$ ) и в случае резервированного распределения запросов по нескольким путям к одному из нескольких возможных приемников ( $S_2$ ) [12—15, 34].

При проведении имитационных экспериментов предположим, что используются одинаковые маршрутизаторы, серверы и сегменты линий связи. Линии связи имеют пропускную способность  $BW = 100$  Мбит/с. Время распространения сигнала  $t_{prg} = 100/0,7 \cdot 2,99 \cdot 10^8 \approx 5 \cdot 10^{-7}$  с, где  $2,99 \cdot 10^8$  м/с — скорость света в вакууме;  $N = 5 \cdot 10^4$  шт.;  $RS = 4096$  бит;  $LR = 0$ ;  $BER = 0$ ;  $P_0 = 1$ ;  $P_1 = 1$ . Задержки обслуживания в маршрутизаторах  $RD$  и серверах  $SD$  заданы с помощью встроенной функции экспоненциального распределения 'exponential' в AnyLogic с параметром  $\lambda = 6,5$  и минимальным значением аргумента соответственно  $RD_{min} = 2,5 \cdot 10^{-6}$  с и  $SD_{min} = 10^{-6}$  с.

На рис. 2 показана установленная экспериментально зависимость среднего времени пребывания запросов в системе  $T$ , с от интенсивности потока запросов  $\Lambda$ , с<sup>-1</sup> при  $t_0 = 15$  с для кратностей резервирования передач  $k = 1, 2, \dots, 5$  (кривые 1—5 соответственно). Рис. 2, а пред-

ставляет вариант резервированного взаимодействия  $S_1$ , а рис. 2, б — вариант  $S_2$ .

Как видно из рис. 2, существует граница интенсивности потока запросов, ниже которой резервирование процессов передачи (рис. 2, а) и распределения запросов (рис. 2, б) позволяет снизить среднее время пребывания пакетов (запросов) в системе.

Рассмотрим случай многопутевой резервированной передачи данных ( $S_1$ ) и многопутевого резервированного распределения запросов ( $S_2$ ) с учетом возможных потерь при обслуживании из-за превышения максимального времени ожидания  $t_0$ .

На рис. 3 представлены результаты моделирования при  $t_0 = 2,1 \cdot 10^{-3}$  с и кратности резервирования передач  $k = 1, 2, \dots, 5$  (кривые 1—5 соответственно).

Как видно из рис. 2 и рис. 3, изменение максимального времени ожидания может оказывать влияние на диапазон значений интенсив-

ности потока запросов, при котором резервирование передач целесообразно.

На рис. 4 (см. вторую сторону обложки) показано существование границы эффективности (целесообразности) многопутевой резервированной передачи пакетов (а) и многопутевого резервированного распределения запросов через сеть (б), когда резервирование передач приводит к увеличению вероятности своевременности доставки запросов в узлы назначения  $P$  по сравнению с вариантом без резервирования. При увеличении  $t_0$  повышается доля успешно переданных пакетов. По мере возрастания интенсивности потока запросов увеличивается нагрузка в системе и снижается вероятность успешных передач. Установлено, что при уменьшении вероятности доставки  $P$  до значений 0,05 и ниже существенно возрастает вариативность среднего времени пребывания запросов в системе. Таким образом, область целесообразности многопутевого резервиро-

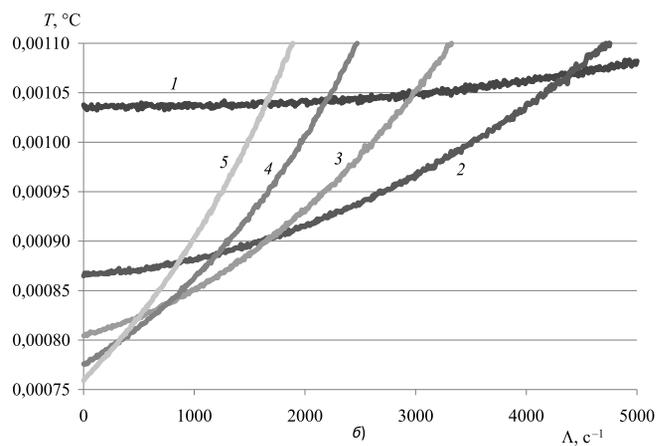
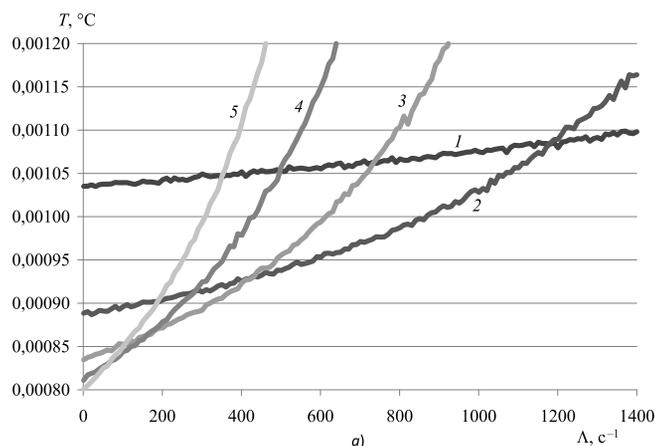


Рис. 2. Среднее время пребывания в системе в случае многопутевой резервированной передачи пакетов (а) и многопутевого резервированного распределения запросов (б) при  $t_0 = 15$  с

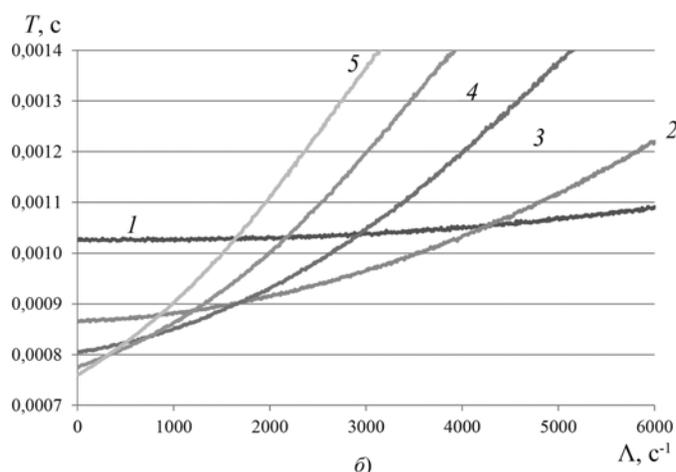
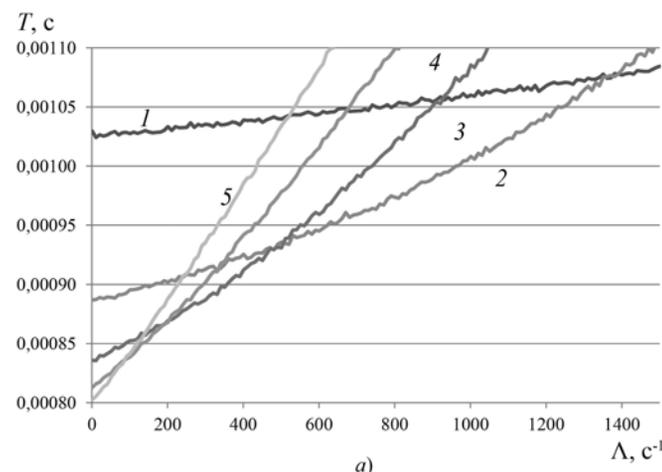


Рис. 3. Среднее время пребывания в системе в случае многопутевой резервированной передачи пакетов (а) и многопутевого резервированного распределения запросов (б) при  $t_0 = 2,1 \cdot 10^{-3}$  с

ванного обслуживания запросов существует и для передач пакетов по одному адресу ( $S_1$ ) и для распределения запросов через сеть ( $S_2$ ).

На рис. 5 (см. вторую сторону обложки) показана зависимость среднего времени пребывания запросов в системе и вероятности их своевременного распределения (доставки к серверам) от их размера: 512 бит (кривые 7–9), 12 144 бит (кривые 1–3) и переменного размера (512, 1024, 2048, 4096, 8192 или 12 144 бит; кривые 4–6). В последнем случае размер пакета (кадра Ethernet) задается согласно равномерному распределению. При этом  $t_0 = 2,1 \cdot 10^{-3}$  с, кратность резервирования передач  $k = 1$  (кривые 1, 4, 7),  $k = 2$  (кривые 2, 5, 8) и  $k = 3$  (кривые 3, 6, 9). Здесь  $k = 1$  — случай без резервирования.

### Анализ резервированного распределения запросов через сеть с учетом ошибок передач и неготовности серверов

Рассмотрим многопутевое резервированное распределение запросов через сеть с учетом их возможных потерь из-за ошибок передач ( $BER$ ), сбоев и иных ошибок, например, программного и аппаратного обеспечения ( $LR$ ), неготовности узлов назначения ( $P_0$ ). Будем предполагать, что возникающие битовые ошибки обнаруживаются промежуточными узлами с вероятностью  $P_1$ , при этом искаженные пакеты всегда обнаруживаются и отбрасываются узлами назначения.

Размер запроса  $RS$  варьируется.

На рис. 6 (см. вторую сторону обложки) представлена вероятность доставки запросов к серверам кластера  $P$  при  $LR = 10^{-4}$  (кривые 1–3 — рис. 6, а и 1–6 — рис. 6, б),  $LR = 10^{-3}$  (кривые 4–6 — рис. 6, а),  $BER = 10^{-7}$  (кривые 1–3 — рис. 6, а и 1–6 — рис. 6, б),  $BER = 10^{-5}$  (кривые 4–6 — рис. 6, а),  $P_0 = 0,99$  (кривые 1–3 — рис. 6, а и 1–6 — рис. 6, б),  $P_0 = 0,95$  (кривые 4–6 — рис. 6, а), кратности резервирования  $k = 1$  соответствуют кривые 1, 4,  $k = 2$  — кривые 2, 5,  $k = 3$  — кривые 3, 6.

Зависимости на рис. 6, а (см. вторую сторону обложки) построены в предположении идеальности контроля по обнаружению битовых ошибок ( $P_1 = 1$ ). Результаты имитационных экспериментов, приведенные на рис. 6, б, учитывают возможность необнаружения ошибки средствами контроля, при этом вероятности обнаружения ошибки  $P_1 = 0,9$  соответствуют кривые

1–3, а случаю отсутствия контроля, когда  $P_1 = 0$ , соответствуют кривые 4–6. Как видно из рис. 6, а, резервированное многопутевое распределение запросов через сеть в условиях ненадежной доставки позволяет повысить вероятность своевременного распределения запросов. При возрастании интенсивности потока запросов вероятность своевременности резервированного распределения снижается. В то же самое время из-за потерь вследствие ошибок передач и неготовности серверов (кривые 4–6) уменьшается длительность ожидания в очередях узлов и, таким образом, время пребывания запросов в системе, что приводит к повышению вероятности успешных передач в определенном диапазоне интенсивности потока запросов.

Необнаружение битовых ошибок (кривые 4–6), по сравнению со случаем с высокой вероятностью нахождения искажений в пересылаемых пакетах в промежуточных узлах (кривые 1–3), ведет к повышению нагрузки, времени пребывания запросов в системе, уменьшению вероятности своевременного распределения запросов через сеть (рис. 6, б, см. вторую сторону обложки).

На рис. 7 показана вероятность своевременного распределения запросов  $P$  при  $t_0 = 2,1 \cdot 10^{-3}$  с,  $LR = 10^{-3}$ ,  $BER = 10^{-5}$ ,  $P_0 = 0,95$ ,  $P_1 = 1$  и интенсивности потока запросов  $\Lambda = 3951, 3776, 2951, 2126, 326$  с $^{-1}$  в зависимости от кратности резервирования передач  $k$ .

При решении задачи максимизации доли своевременно и безошибочно распределенных через сеть критичных к задержкам запросов и/или при ограничениях повторных передач представленная зависимость указывает на су-

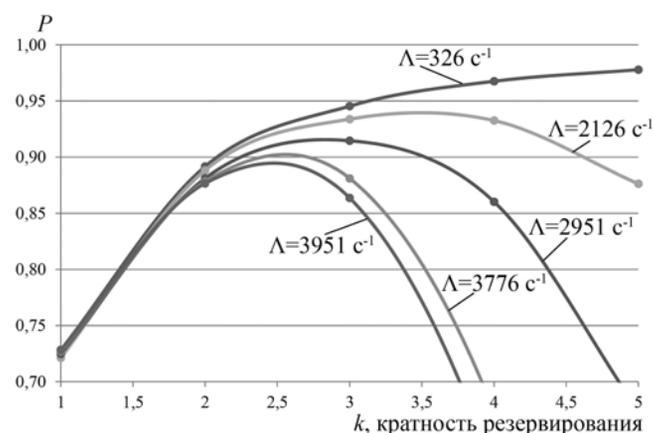


Рис. 7. Вероятность доставки запросов к серверам кластера в зависимости от кратности резервирования передач  $k$

существование оптимальной кратности резервирования запросов в зависимости от интенсивности потока запросов.

### Заключение

Исследованы возможности повышения вероятности своевременного безошибочного обслуживания и снижения средних задержек в системах многопутевой передачи данных в результате резервированного выполнения критичных к задержкам запросов.

На основе результатов имитационного моделирования проанализирована эффективность резервированных многопутевых передач в распределенных компьютерных системах, в которых реализуется доставка копий пересылаемых по многим путям пакетов в один адресуемый узел и/или многопутевое распределение запросов к одному или нескольким узлам обработки, готовым к их обслуживанию. Модели разработаны в рамках дискретно-событийного подхода в среде имитационного моделирования AnyLogic 7 Professional.

Показано существование области эффективности резервированной многопутевой передачи данных (пакетов) и резервированного многопутевого распределения запросов в зависимости от их критичности к задержкам и загруженности системы.

Показано существование оптимальной кратности резервирования при многопутевом распределении запросов, обеспечивающей максимизацию вероятности своевременно и безошибочно распределенных через сеть критичных к задержкам запросов или запросов, распределяемых в условиях, исключающих возможность повторных передач.

Установлена зависимость оптимальной кратности резервирования при многопутевом распределении запросов через сеть от интенсивности потока запросов и их критичности к задержкам.

### Список литературы

1. **Kopetz H.** Real-Time Systems: Design Principles for Distributed Embedded Applications. Springer, 2011. P. 396.
2. **Sorin D.** Fault Tolerant Computer Architecture. Morgan & Claypool, 2009. 103 p.
3. **Шубинский И. Б.** Надежные отказоустойчивые информационные системы. Методы синтеза. Ульяновск: Областная типография "Печатный двор", 2016. 544 с.
4. **Советов Б. Я., Колбанев М. О., Татарникова Т. М.** Технологии инфокоммуникации и их роль в обеспечении

информационной безопасности // Геополитика и безопасность. 2014. № 1(25). С. 69–77.

5. **Верзун Н. А., Колбанев М. О., Татарникова Т. М.** Технологическая платформа четвертой промышленной революции // Геополитика и безопасность. 2016. № 2 (34). С. 73–78.

6. **Gatchin Y. A., Zharinov I. O., Korobeynikov A. G., Zharinov O. O.** Theoretical estimation of Grassmann's transformation resolution in avionics color coding systems // Modern Applied Science. 2015. Vol. 9, N. 5. P. 197–210.

7. **Богатырев В. А., Богатырев С. В.** Резервированная передача данных через агрегированные каналы в сети реального времени // Известия высших учебных заведений. Приборостроение. 2016. Т. 59. № 9. С. 735–740.

8. **Богатырев В. А.** Надежность и эффективность резервированных компьютерных сетей // Информационные технологии. 2006. № 9. С. 25–30.

9. **Богатырев В. А.** Комбинаторно-вероятностная оценка надежности и отказоустойчивости кластерных систем // Приборы и системы. Управление, контроль, диагностика. 2006. № 6. С. 21–26.

10. **Новиков С. Н.** Классификация методов маршрутизации в мультисервисных сетях связи // Вестник СибГУТИ. 2013. № 1. С. 57–67.

11. **Шувалов В. П., Варакина И. Ю.** Классификация методов многопутевой маршрутизации // Т-Comm #1. 2014. С. 29–32.

12. **Богатырев В. А., Паршутина С. А.** Модели многопутевой отказоустойчивой маршрутизации при распределении запросов через сеть // Вестник компьютерных и информационных технологий. 2015. № 12. С. 23–28.

13. **Богатырев В. А., Паршутина С. А.** Многопутевое резервированное распределение через сеть критичных к задержкам запросов // Вестник компьютерных и информационных технологий. 2016. № 10 (148). С. 41–46.

14. **Bogatyrev V. A., Parshutina S. A.** Efficiency of Redundant Multipath Transmission of Requests Through the Network to Destination Servers // Communications in Computer and Information Science, IET. 2016. Vol. 678. P. 290–301.

15. **Bogatyrev V. A., Parshutina S. A.** Redundant Distribution of Requests Through the Network by Transferring Them Over Multiple Paths // Communications in Computer and Information Science, IET. 2016. Vol. 601. P. 199–207.

16. **Лемешко А. В., Вавенко Т. В.** Усовершенствование потоковой модели многопутевой маршрутизации на основе балансировки нагрузки // Проблемы телекоммуникаций. 2012. № 1 (6). С. 12–29.

17. **Gunnar A., Johansson M.** Robust load balancing under traffic uncertainty — tractable models and efficient algorithms // Telecommun Syst. 2011. 48:93–107.

18. **Banner R., Orda A.** Multipath Routing Algorithms for Congestion Minimization, CCIT Report No. 429, Department of Electrical Engineering, Technion, Haifa, Israel, 2004. URL: <http://www.ee.technion.ac.il/people/ron/Congestion.pdf>.

19. **Жуков И. А., Кулаков Ю. А., Шпак И. Ю.** Адаптивная многопутевая маршрутизация // Проблемы информатизации та управління. 2009. № 3 (27). С. 73–78.

20. **Merindol P.** Improving Load Balancing with Multipath Routing / P. Merindol, J. Pansiot, S. Cateloin // Proc. of the 17-th International Conference on Computer Communications and Networks, IEEE ICCCN 2008. 2008. P. 54–61.

21. **Шувалов В. П., Селянина И. Ю.** Методика обеспечения отказоустойчивости в мультисервисных сетях связи // Проблемы информатики. 2012. № 2 (14). С. 55–62.

22. **Rajeev V., Muthukrishnan C. R.** Reliable backup routing in fault tolerant real-time networks // Proceedings. Ninth IEEE International Conference on Networks, ICON, 2001.

23. **Будылдина Н. В., Трибунский Д. С., Шувалов В. П.** Оптимизация сетей с многопротокольной коммутацией по меткам. М.: Горячая линия-Телеком, 2010. 144 с.

24. **Богатырев В. А., Богатырев А. В.** Оптимизация резервированного распределения запросов в кластерных системах реального времени // Информационные технологии. 2015. Т. 21, № 7. С. 495–502.

25. Bogatyrev V. A., Bogatyrev A. V. Functional Reliability of a Real-Time Redundant Computational Process in Cluster Architecture Systems // Automatic Control and Computer Sciences, IET. 2015. Vol. 49, N. 1. P. 46–56.

26. Богатырев В. А., Богатырев А. В. Надежность функционирования кластерных систем реального времени с фрагментацией и резервированным обслуживанием запросов // Информационные технологии. 2016. Т. 22, № 6. С. 409–416.

27. Дудин А. Н., Сунь Б. Многолинейная система MAP/RH/N с управляемым широкополосным обслуживанием ненадежными приборами // Автоматика и вычислит. техника. 2009. Т. 43, № 5. С. 32–43.

28. Дудин А. Н., Сунь Б. Многолинейная ненадежная система с управляемым широкополосным обслуживанием // Автоматика и телемеханика. 2009. Т. 70, № 12. С. 147–160.

29. Богатырев В. А., Богатырев С. В. Своевременность обслуживания в многоуровневых кластерных системах с поэтапным уничтожением просроченных запросов // Вестник компьютерных и информационных технологий. 2018. № 2 (164). С. 28–35.

30. Богатырев В. А., Богатырев С. В. Резервированное обслуживание запросов, критичных к задержкам ожидания, в двухуровневых системах // Известия высших учебных заведений. Приборостроение. 2017. Т. 60, № 10. С. 945–950.

31. Богатырев В. А., Богатырев С. В. Резервированное обслуживание в группе одноканальных систем с назначением различных приоритетов копиям запроса // Известия высших учебных заведений. Приборостроение. 2017. Т. 60, № 11. С. 1033–1039.

32. Богатырев В. А., Богатырев С. В. Резервированное обслуживание в кластерах с уничтожением неактуальных запросов // Вестник компьютерных и информационных технологий. 2017. № 1 (151). С. 21–28.

33. Богатырев В. А., Богатырев А. В. Модель резервированного обслуживания запросов реального времени в компьютерном кластере // Информационные технологии. 2016. Т. 22, № 5. С. 348–355.

34. Богатырев В. А., Богатырев С. В. Надежность мультикластерных систем с перераспределением потоков запросов // Известия высших учебных заведений. Приборостроение. 2017. Т. 60, № 2. С. 171–177.

35. Олифер В., Олифер Н. Компьютерные сети. Принципы, технологии, протоколы. СПб.: Питер, 2016. 992 с.

36. Паршутина С. А. Организация имитационных экспериментов при проектировании распределенных компьютерных систем с резервированной передачей данных // Научно-технический вестник информационных технологий, механики и оптики. 2017, Т. 17, № 5. С. 841–849.

V. A. Bogatyrev, D. Sc., Professor, e-mail: vladimir.bogatyrev@gmail.com,

S. A. Parshutina, Postgraduate, e-mail: svetlana.parshutina@gmail.com,

St. Petersburg National Research University of Information Technologies, Mechanics and Optics

## The Analysis of How the Redundancy Degree of Multipath Transmissions Affects the Probability of their Timely Service

*The possibility of using redundant processing of delay-sensitive requests to increase the probability of timely faultless service and reduce delays in multipath data transmission systems is investigated. Simulation is used to evaluate the effectiveness of redundant multipath transmissions: the process of sending multiple copies of packets along different paths in a given network, in order to deliver them to one destination node or to one of several service nodes which are ready to execute incoming requests. The models are developed based on the discrete-event approach in AnyLogic 7 Professional simulation environment.*

*The degree of redundancy defines the number of copies of packages which are distributed among different routes in the network and depends on the sensitivity of these packets to time delays.*

*In distributed computer systems, the redundant multipath data transmission and the redundant multipath distribution of requests are considered the ways of increasing the efficiency of communication between computational processes which exchange messages that are critical to service delays over the network.*

*The redundant data transmission means that data packets are cloned and copies of those packets travel along several paths to one destination node. Data are regarded to be transmitted successfully if at least one copy of each packet has been delivered to that node in time, i.e. within a given period of time. The redundant distribution of requests over the network implies that copies of requests are forwarded along several routes to one or more destination nodes, which have been selected out of all possible ones in some way. A request is regarded to be distributed successfully if at least one copy of it has been delivered to all selected nodes within the specified time and those nodes have been ready for serving incoming requests. The effectiveness of redundant multipath transmissions — when the time of delivery of packets is limited — is determined by the probability of timely faultless service.*

*The existence of the scope of efficiency for the redundant multipath data transmissions and the redundant multipath distribution of requests, with considering their sensitivity to time delays, is shown.*

*It is found that there exists the optimal degree of redundancy that ensures the maximization of the probability of timely and faultless distribution of the multiplied requests over a given network when those requests are critical to time delays and/or can be transmitted only once.*

*It was also discovered that the optimal redundancy degree of multipath transmissions depends on the intensity of the flow of requests and their sensitivity to time delays.*

**Keywords:** redundancy, multipath transmission, distribution of requests, sensitivity to time delays

DOI: 10.17587/it.24.772-781

## References

1. **Kopetz H.** Real-Time Systems: Design Principles for Distributed Embedded Applications, Springer, 2011, 396 p.
2. **Sorin D.** Fault Tolerant Computer Architecture, Morgan & Claypool 2009, 103 p.
3. **Shubinskij I. B.** *Nadezhnye otkazoustojchivye informacionnye sistemy. Metody sinteza* (Reliable fault-tolerant information systems. Synthesis method), Ul'yanovsk, Oblastnaya tipografiya "Pechatnyj dvor", 2016, 544 p. (in Russian).
4. **Sovetov B. Ya., Kolbanyov M. O., Tatarnikova T. M.** Tekhnologii infokommunikacii i ih rol' v obespechenii informacionnoj bezopasnosti (Infocommunication technologies and their role in information security), *Geopolitika i bezopasnost'*, 2014, no. 1(25), pp. 69–77 (in Russian).
5. **Verzun N. A., Kolbanev M. O., Tatarnikova T. M.** Tekhnologicheskaya platforma chetvertoj promyshlennoj revolyucii (Technology platform of the fourth industrial revolution), *Geopolitika i Bezopasnost'*, 2016, no. 2 (34), pp. 73–78 (in Russian).
6. **Gatchin Y. A., Zharinov I. O., Korobeynikov A. G., Zharinov O. O.** Theoretical estimation of Grassmann's transformation resolution in avionics color coding systems, *Modern Applied Science*, 2015, vol. 9, no. 5, pp. 197–210.
7. **Bogatyrev V. A., Bogatyrev S. V.** Rezervirovannaya poredacha dannyh cherez agregirovannye kanaly v seti real'nogo vremeni (Redundant data transmission through aggregated channels in real-time network), *Izvestiya vuzov. Priborostroenie*, 2016, vol. 59, no. 9, pp. 735–740 (in Russian).
8. **Bogatyrev V. A.** Nadezhnost' i ehffektivnost' rezervirovannyh komp'yuternyh setej (Reliability and efficiency of redundant computer networks), *Informacionnye Tekhnologii*, 2006, no. 9, pp. 25–30 (in Russian).
9. **Bogatyrev V. A.** Kombinatorno-veroyatnostnaya ocenka nadezhnosti i otkazoustojchivosti klasternyh sistem (Combinatorial probabilistic evaluation of reliability and fault tolerance of cluster systems), *Pribory i sistemy. Upravlenie, kontrol', diagnostika*, 2006, no. 6, pp. 21–26 (in Russian).
10. **Novikov S. N.** Klassifikaciya metodov marshrutizacii v mul'tiservisnyh setyah svyazi (Classification of routing methods in multiservice communication networks), *Vestnik SibGUTI*, no. 1, 2013, pp. 57–67 (in Russian).
11. **Shuvalov V. P., Varaksina I. Yu.** Klassifikaciya metodov mnogoputevoj marshrutizacii (Classification of multipath routing methods), *T-Comm*, 2014, no. 1, pp. 29–32 (in Russian).
12. **Bogatyrev V. A., Parshutina S. A.** Modeli mnogoputevoj otkazoustojchivoj marshrutizacii pri raspredelenii zaprosov cherez set' (Multipath fault-tolerant routing models for distributing queries through the network), *Vestnik Komp'yuternyh i Informacionnyh Tekhnologij*, 2015, no. 12, pp. 23–28 (in Russian).
13. **Bogatyrev V. A., Parshutina S. A.** Mnogoputevoe rezervirovannoe raspredelenie cherez set' kritichnyh k zaderzhkam zaprosov (Multipath redundant distribution of delay-sensitive queries through the network), *Vestnik Komp'yuternyh i Informacionnyh Tekhnologij*, 2016, no. 10 (148), pp. 41–46 (in Russian).
14. **Bogatyrev V. A., Parshutina S. A.** Efficiency of Redundant Multipath Transmission of Requests Through the Network to Destination Servers, *Communications in Computer and Information Science, IET* – 2016, vol. 678, pp. 290–301.
15. **Bogatyrev V. A., Parshutina S. A.** Redundant Distribution of Requests Through the Network by Transferring Them Over Multiple Paths, *Communications in Computer and Information Science, IET* – 2016, vol. 601, pp. 199–207.
16. **Lemeshko A. V., Vavenko T. V.** Usovshenstvovanie potokovoj modeli mnogoputevoj marshrutizacii na osnove balansirovki nagruzki (Improving the threading model of multipath routing based on load balancing), *Problemy Telekomunikacij*, 2012, no. 1 (6), pp. 12–29 (in Russian).
17. **Gunnar A., Johansson M.** Robust load balancing under traffic uncertainty – tractable models and efficient algorithms, *Telecommun Syst.*, 2011, vol. 48, pp. 93–107.
18. **Banner R., Orda A.** Multipath Routing Algorithms for Congestion Minimization, CCIT Report No. 429, Department of Electrical Engineering, Technion, Haifa, Israel, 2004, available at: <http://www.ee.technion.ac.il/people/ron/Congestion.pdf>.
19. **Zhukov I. A., Kulakov Yu. A., Shpak I. Yu.** Adaptivnaya mnogoputevaya marshrutizaciya (Adaptive multipath routing), *Problemy Informaticzii ta Upravlinnya*, 2009, no. 3 (27), pp. 73–78 (in Russian).
20. **Merindol P., Pansiot J., Catelein S.** Improving Load Balancing with Multipath Routing, *Proc. of the 17-th International Conference on Computer Communications and Networks, IEEE ICCCN 2008*, 2008, pp. 54–61.
21. **Shuvalov V. P., Selyanina I. Yu.** Metodika obespecheniya otkazoustojchivosti v mul'tiservisnyh setyah svyazi (A method of providing fault tolerance in multi-service networks), *Problemy Informatiki*, 2012, no. 2(14), pp. 55–62 (in Russian).
22. **Rajeev V., Muthukrishnan C. R.** Reliable backup routing in fault tolerant real-time networks, *Proceedings. Ninth IEEE International Conference on Networks, ICON*, 2001.
23. **Budyldina N. V., Tribunskij D. S., Shuvalov V. P.** Optimizaciya setej s mnogoprotokol'noj kommutaciej po metkam. (Optimization of networks with Multiprotocol label switching), Moscow, Goryachaya liniya-Telekom, 2010, 144 p. (in Russian).
24. **Bogatyrev V. A., Bogatyrev A. V.** Optimizaciya rezervirovannogo raspredeleniya zaprosov v klasternyh sistemah real'nogo vremeni (Optimization of redundant query distribution in real-time cluster systems), *Informacionnye Tekhnologii*, 2015, vol. 21, no. 7, pp. 495–502 (in Russian).
25. **Bogatyrev V. A., Bogatyrev A. V.** Functional Reliability of a Real-Time Redundant Computational Process in Cluster Architecture Systems, *Automatic Control and Computer Sciences, IET*, 2015, vol. 49, no. 1, pp. 46–56.
26. **Bogatyrev V. A., Bogatyrev A. V.** Nadezhnost' funkcionirovaniya klasternyh sistem real'nogo vremeni s fragmentaciej i rezervirovannyh obsluzhivaniem zaprosov (Reliable operation of real-time cluster systems with fragmentation and redundant service requests), *Informacionnye Tekhnologii*, 2016, vol. 22, no. 6, pp. 409–416 (in Russian).
27. **Dudin A. N., Sun' B.** Mnogolinejnaya sistema MAP/PH/N s upravlyaemym shirokoveshchatel'nym obsluzhivaniem nenadezhnymi priborami (Multi-line MAP/PH/N system with managed broadcast service by unreliable devices), *Avtomatika i Vychislit. Tekhnika*, 2009, vol. 43, no. 5, pp. 32–43.
28. **Dudin A. N., Sun' B.** Mnogolinejnaya nenadezhnaya sistema s upravlyaemym shirokoveshchatel'nym obsluzhivaniem (Multi-line unreliable system with managed broadcast service), *Avtomatika i Telemekhanika*, 2009, vol. 70, no. 12, pp. 147–160 (in Russian).
29. **Bogatyrev V. A., Bogatyrev S. V.** Svoevremennost' obsluzhivaniya v mnogourovnevnyh klasternyh sistemah s poehtapnym unichtozheniem prosrochennyh zaprosov (Timeliness of service in multi-level cluster systems with phased elimination of overdue requests), *Vestnik Komp'yuternyh i Informacionnyh Tekhnologij*, 2018, no. 2(164), pp. 28–35 (in Russian).
30. **Bogatyrev V. A., Bogatyrev S. V.** Rezervirovannoe obsluzhivanie zaprosov, kritichnyh k zaderzhkam ozhidaniya, v dvuhurovnevnyh sistemah (Redundant service for latency-critical requests in two-tier systems), *Izvestiya vuzov. Priborostroenie*, 2017, vol. 60, no. 10, pp. 945–950 (in Russian).
31. **Bogatyrev V. A., Bogatyrev S. V.** Rezervirovannoe obsluzhivanie v grupe odnokanal'nyh sistem s naznacheniem razlichnyh prioritetov kopiyam zaprosa (Redundant service in a group of single-channel systems with different priorities assigned to query copies), *Izvestiya vuzov. Priborostroenie*, 2017, vol. 60, no. 11, pp. 1033–1039 (in Russian).
32. **Bogatyrev V. A., Bogatyrev S. V.** Rezervirovannoe obsluzhivanie v klasterah s unichtozheniem neaktual'nyh zaprosov (Redundant service in clusters with the destruction of irrelevant requests), *Vestnik komp'yuternyh i informacionnyh tekhnologij*, 2017, no. 1(151), pp. 21–28 (in Russian).
33. **Bogatyrev V. A., Bogatyrev A. V.** Model' rezervirovannogo obsluzhivaniya zaprosov real'nogo vremeni v komp'yuternom klasterne (Real-time redundant query service model in a computer cluster), *Informacionnye Tekhnologii*, 2016, vol. 22, no. 5, pp. 348–355 (in Russian).
34. **Bogatyrev V. A., Bogatyrev S. V.** Nadezhnost' mul'tiklasternyh sistem s pereraspredeleniem potokov zaprosov (The reliability of multi-cluster systems with redistribution query), *Izvestiya vuzov. Priborostroenie*, 2017, vol. 60, no. 2, pp. 171–177 (in Russian).
35. **Olifer V., Olifer N.** *Komp'yuternye seti. Principy, tekhnologii, protokoly* (Computer networks. Principles, technologies, protocols), SPb., Piter, 2016. 992 p.
36. **Parshutina S. A.** Organizaciya imitacionnyh eksperimentov pri proektirovanii raspredelyonnyh komp'yuternyh sistem s rezervirovannoj peredachej dannyh (Organization of simulation experiments in the design of distributed computer systems with redundant data transmission), *Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki*, 2017, vol. 17, no. 5, pp. 841–849 (in Russian).