

УДК 004

В. С. Мельник

*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)*

## Многопоточная маршрутизация в автоматизированной системе учета конечных нод распределенного кластера

*Изложено описание подхода к проектированию географически распределенной системы, в которой используется многопоточная маршрутизация. Данный механизм позволяет повысить отказоустойчивость проектируемой системы в целом за счет использования более одного канала связи между конечными нодами.*

*Помимо решения задачи повышения надежности в статье изложена методика, позволяющая выполнять балансировку ресурсов, задействуя все имеющиеся каналы связи.*

*Целью проведенных исследований служит разработка комплексного решения по проектированию инфраструктуры передачи непрерывного потока данных несколькими маршрутами, которые позволяют обеспечить отказоустойчивость и разделить нагрузку между собой, позволяя использовать эти же маршруты для административных целей, например для мониторинга системы и отслеживания критических ситуаций, а также для управления имеющимися нодами.*

*В предлагаемой схеме проектирования предлагается обеспечивать отказоустойчивость географически распределенной системы так, чтобы каждое из цепи сообщений выступало в качестве независимой единицы компьютерной системы. В результатах работы проведен анализ различных вариантов подключения системы с учетом вероятности выхода из строя каждого из независимых звеньев.*

### Проектирование сетевой инфраструктуры, отказоустойчивость, надежность, резервирование, балансировка нагрузки, маршрутизация, распределенные системы, кластерные системы

В настоящей статье рассматривается задача проектирования сетевой инфраструктуры с резервированием каналов связи, соединяющих конечные узлы кластерной системы записи потокового видео. Также решается задача обеспечения многопоточной маршрутизации для алгоритмов искусственного интеллекта, описанных ранее [1]. Такая система должна непрерывно записывать видеопоток, поступающий с вещательного сервера. По закону Российской Федерации (Ст. 34 «О хранении материалов радио- и телепередач») телекомпании должны хранить запись эфира за последний месяц. В статье помимо решения основного стека задач – резервирования и балансировки каналов связи, также предлагается и решение косвенной задачи – запись эфира с вещательных серверов во время проведения технических работ на магистральных линиях связи.

Сам по себе сервис записи потокового видео с серверов должен строиться таким образом, чтобы в случае выхода из строя какого-либо отдельного компонента не произошел крах системы записи в целом [2].

На текущий момент времени стремление повысить функциональность систем значительно опережает текущие темпы развития имеющихся методов повышения надежности. В данной ситуации одним из наиболее действенных средств защиты от краха системы является построение нескольких маршрутов до конечной машины. Стоит отметить, что маршруты должны быть построены таким образом, чтобы при нарушении работоспособности одного из них второй продолжал бы работать.

Имеется два подхода для обеспечения отказоустойчивости системы. Первый – увеличение надежности компонентов конечных элементов. В этом случае при отказе одного компонента система продолжит функционировать. Это – последовательный вариант построения кластерной системы.

Вероятность безотказной работы такой системы, состоящей из  $N$  конечных элементов, легко рассчитывается по формуле

$$P(t) = p_1(t)p_2(t) \dots p_N(t) = \prod_{i=1}^N p_i(t).$$

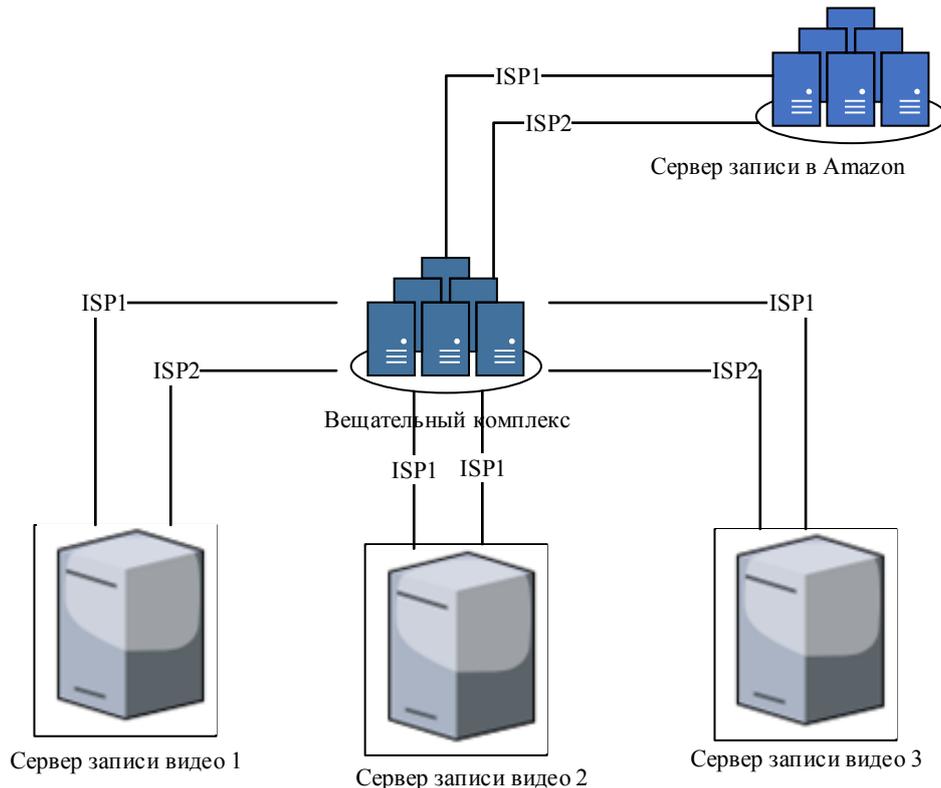


Рис. 1

Второй подход заключается в избыточном резервировании всех критичных компонентов системы таким образом, что при выходе из строя одного из компонентов кластер продолжит функционировать, используя оставшиеся вычислительные мощности. В данном случае работа системы в целом не нарушится. Стоит отметить, что в таком подходе отказ всей системы возможен только в случае выхода из строя всех компонентов. Здесь, как и в первом случае, все узлы кластера подключаются последовательно. Вероятность безотказной работы всего кластера, состоящего из  $k$  элементов, высчитывается по формуле

$$P(t) = 1 - q_1(t)q_2(t) \dots q_i(t) \dots q_k(t) = 1 - \prod_{i=1}^k [1 - q_i(t)],$$

где  $q_i(t)$  – это вероятность выхода из строя  $i$ -го элемента системы. Как можно видеть, чем больше количество независимых элементов кластера, тем выше надежность системы в целом.

Общая схема сетевой инфраструктуры изображена на рис. 1.

На схеме весь вещательный комплекс [2] для простоты обозначен единственной пиктограммой, так как в рамках настоящей задачи нет необходимости излишне детализировать эту часть сети. К каждому удаленному серверу приходит не-

сколько провайдеров. Разумеется, передавать данные в открытом виде небезопасно. В связи с этим через каждый канал связи (ISP) происходит подключение с помощью VPN к вещательному комплексу.

В качестве шлюза везде установлен маршрутизатор от MikroTik. Он позволяет обеспечить необходимую для работоспособности коммутационную схему. Для обеспечения отказоустойчивости и балансировки используется инструмент PCC (Pet Connection Classifier). Данный инструмент позволяет использовать поля, находящиеся в заголовке IP-адреса: с помощью алгоритма хеширования маршрутизатор преобразует выбранные для работы поля в 32-битное значение. Далее это значение делится на указанный пользователем знаменатель и происходит сравнение остатка от деления. В случае, если остаток равен тому или иному значению (указанному в конфигурации устройства), пакет получает дополнительное поле-метку и отправляется в один из каналов VPN [3]. Для работы данного инструмента можно выбрать одно из полей пакета:

- адрес источника;
- адрес назначения;
- порт источника;
- порт назначения.

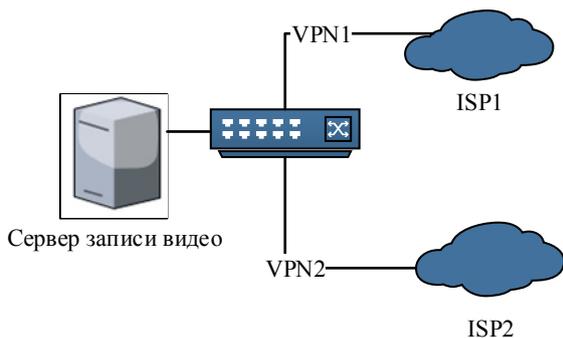


Рис. 2

Иными словами, инструмент позволяет разделить трафик на равные потоки (рис. 2).

В рассматриваемой системе таких потоков всего два, так как вероятность отказа сразу двух каналов связи стремится к нулю. После работы РСС у заголовков пакетов появляется дополнительная метка, которая далее позволит устройству определить дальнейший путь для пакета (рис. 3). В оборудовании от «Mikrotik» эта метка носит наименование Mangle Mark и необходима для добавления к пакету дополнительного поля, которое содержит информацию об очереди, в которой будет обработан пакет. Стоит отметить, что после прохождения очереди и выполнения необходимых действий над пакетом данная метка снимается, а сам пакет принимает исходный вид без каких-либо искажений.

Данный механизм (Mangle Mark) более актуален в случае использования большего количества связующих линий, однако в данной конфигурации системы этого количества более чем достаточно. Инструмент тестируется вместе с реализацией вероятностного алгоритма на большем количестве соединений с привлечением операторов сотовой связи в качестве дополнительных – резервных каналов связи.

Помимо непосредственно инструмента балансировки необходимо дополнительно указать устройству маршруты, с помощью которых отправлять маркированные пакеты. По умолчанию для них выбирается тот же маршрут, по которому пакет пришел. В случае отказа одного из каналов связи маршрутизатор исключит неисправный канал связи из своей таблицы маршрутизации и

будет работать с оставшимися. Для этого и необходимо указать резервный маршрут через альтернативного провайдера. В данном случае используется статическая маршрутизация, которая предполагает нахождение сразу нескольких путей между каждой парой «получатель-источник» заранее. В дальнейшем при работе у маршрутизатора будут одновременно доступны несколько критериев при выборе маршрута для каждого входящего маршрута. Основной – в случае, если оба канала связи работоспособны, и дополнительный – если вся информация отправляется только через одного провайдера.

Многопоточная маршрутизация имеет несколько преимуществ перед обычной – это отсутствие

1) номинальной нагрузки на каждый из каналов связи, что позволяет использовать их под второстепенные задачи (мониторинг оборудования, доступ к видеоматериалам и т. д.);

2) так называемого «холодного» резерва: оба канала связи используются постоянно, что приводит к эффективному разделению имеющихся ресурсов как с технической, так и с финансовой составляющей;

3) времени переключения, которое необходимо для перестройки таблицы маршрутизации устройства. В случае, если таблица маршрутизации велика, ее перестройка может занять некоторое время – тогда пакеты и часть видеопотока будут утеряны.

На основании вышеизложенного стоит подчеркнуть: в настоящей системе используется лавинная многопоточная маршрутизация с использованием всех имеющихся в системе путей. В этом случае пакеты равномерно распределены по всем имеющимся путям. Основа работоспособности системы – методы, которые работают по принципу round robin – по пакетному распределению данных.

При использовании многопоточной маршрутизации основная задача передачи пакетов ложится на пограничный маршрутизатор и заключается в определении маршрутов для передачи маркированных пакетов. Для реализации данной схемы необходимо применить математический аппарат. В данном случае наиболее подходящим



Рис. 3

решением становится аппарат теории графов, так как компьютерную сеть можно представить в виде графа. Основываясь на наработках, полученных в [1], в тестовую версию системы видеозаписи внедрен альтернативный алгоритм, предоставляющий пакетам маршрут, – муравьиный алгоритм.

Муравьиный алгоритм основан на природных процессах. В частности, за основу взята модель поведения колонии муравьев при поиске, когда увеличенным количеством феромона маркируются наиболее удачные пути. Данный алгоритм имеет вероятностную составляющую, что позволяет примерно равными долями отправлять пакеты через имеющиеся каналы связи. Если используются более 2–3 каналов, данный алгоритм позволяет более равномерно распределять нагрузку между ними. Алгоритм определяется следующей формулой:

$$P_{ij} = l_{ij}^q f_{ij}^p / \sum_{k=0}^N l_k^q f_k^p,$$

где  $P_{ij}$  – вероятность перехода по выбранному пути;  $l_{ij}$  – величина, обратная весу перехода;  $f_{ij}$  – количество феромона на переходе;  $q$  – величина, определяющая «жадность» алгоритма;  $p$  – величина, определяющая «стадность» алгоритма;  $p + q = 1$ .

Вначале работает только вероятностная составляющая – по выбранному пути отправляется только один муравей из колонии. В том случае, если он находит источник пищи (поиск маршрута успешен), маршрут помечается большим количеством феромона, тем самым привлекая остальных муравьев. Таким образом и получается, что вся колония будет двигаться по наиболее оптимальному маршруту до момента, пока он таковым остается. Каждое из пройденных ребер (маршрутов) помечается величиной, обратно пропорциональной длине пройденного пути:

$$\Delta_{ij} = \begin{cases} k/L, & (i, j) \in P; \\ 0, & (ij) \notin P, \end{cases}$$

где  $L$  – длина всего пути;  $k$  – изменяемый параметр.

Механизм работы системы автор проверил опытным путем. Эксперимент состоял из двух частей. Первая – проверка работы системы только с одним каналом передачи данных. Вторая – подключение в уже имеющуюся систему нескольких каналов передачи данных один за другим. В таблице наглядно отображены основные параметры, описывающие работоспособность системы в целом. С определенным количеством каналов система максимально нагружалась и эксперимент длился ровно 1 ч. Также фиксированной величиной являются характеристики каналов связи. Каждый обладает пропускной способностью 40 Мбит/с и задержкой передачи в ~10 мс.

Исходя из полученных данных стоит отметить, что оптимальное количество каналов для работы системы – три. Это связано с тем, что можно передавать максимально возможный размер картинки на серверы записи (именно в этом формате происходит изначальное вещание с серверов). Также стоит отметить, что в данном случае показатель загруженности канала составляет всего 35 %, что позволяет использовать каждый из них для административных (управление серверами), сервисных (мониторинг оборудования) целей, а также для передачи данных (выгрузка по запросу необходимого видеоматериала).

В настоящий момент описанный алгоритм реализован автором на программной оболочке маршрутизаторов с помощью использования скриптов, которые позволяют внедрять MikroTik. За основу работы алгоритма берутся интервалы времени, за которые пакет от маршрутизатора-источника попадает на маршрутизатор-приемник. Учитывая, что время передачи меняется, меняется и основной маршрут передачи пакетов, обеспечивая оптимальное время передачи. Разумеется, вся тестовая инфраструктура работает с учетом реальных мощностей установленного оборудования.

Настоящая система имеет централизованную систему управления. Сервер управления располагается в облаке Amazon (AWS) и непрерывно проверяет работоспособность каждого сервера записи и каждого канала до сервера. Результаты

Измеряемый параметр	Результаты испытаний		
	Один канал связи	Два канала связи	Три канала связи
Загруженность каждого из каналов	~98 %	~70 %	~35 %
Потерянные пакеты, шт	534	232	45
Максимальный размер передаваемого изображения, pix	960 × 576	1280 × 720	1920 × 1080
Усредненное количество ошибок на сервере записи	438	217	29

работы собираются в лог-файл, и при необходимости их можно использовать в дальнейшем. Все серверы системы объединены в единую внутреннюю VPN-сеть для защиты передаваемой информации. Расположение центра управления было выбрано не случайно, так как при использовании данного облака вероятность выхода из строя машины крайне мала. Связано это с тем, что серверы, на которых располагается машина, установлены в серверной высокой доступности и время простоя в данных серверных стремится к нулю.

Проверка работоспособности каналов происходит с помощью аг-протокола, чтобы избежать дополнительной нагрузки на инфраструктуру. Каждый маршрутизатор на стороне сервера записи имеет свой внутренний уникальный адрес. В случае недоступности адреса система мониторинга понимает, что один из каналов связи стал недоступен, и также извещает группу инженеров о нарушении работоспособности системы.

Работоспособность сервера записи проверяется скриптом: управляющий сервер с заданной периодичностью пытается подключиться к серверу записи с помощью ssh. В случае, если получен положительный ответ от сервера (вход в систему), то он считается рабочим. В случае, если вход по каким-либо причинам невозможен, проверка происходит еще 3 раза с периодичностью 1 мин между попытками. Если в итоге система выдает отрицательный результат, то сервер записи получает статус «неработоспособный».

В том случае, если в кластере остается лишь один работоспособный сервер записи, включается новая логическая цепочка: запуск резервного (временного) сервера в облаке AWS. Для автоматической установки нового сервера записи сервер управления имеет набор инструкций (bash-скрипт), который содержит информацию об операционной системе, которую необходимо установить; об версиях устанавливаемого программного обеспечения; об адресе, к которому необходимо подключиться через VPN, и многие другие параметры. С момента получения информации о критическом состоянии кластера до полноценного запуска сервера в AWS проходит не более 3–5 мин, что позволяет в кратчайшие сроки вывести систему в стабильное состояние. Временной промежуток в данном случае был получен эмпирическим спосо-

бом – неоднократным тестированием целостной инфраструктуры.

Если выходят из строя все серверы записи, остается лишь один виртуальный сервер. Он будет и далее единственным, так как в AWS имеется инструмент, позволяющий обеспечить работоспособность сервера в любом случае за счет нескольких зон доступности и Auto Scaling Group [4]. Этот инструмент выполняет непрерывный мониторинг ресурсов пользователя в рамках зоны доступа AWS. Если зона доступа с сервером становится недоступной, инструмент за считанные минуты автоматически перестраивает инфраструктуру в резервной зоне.

В настоящей статье описан подход построения отказоустойчивой системы с вероятностными алгоритмами, который позволяет передавать непрерывный поток данных на серверы записи, используя многопоточную маршрутизацию. Он уникален тем, что система автоматически адаптируется под реальные условия среды с помощью вероятностного алгоритма, предугадывая критические ситуации (например, увеличение количества потерянных пакетов провайдером и дальнейший выход из строя маршрута). Также данный алгоритм позволяет системе автоматически адаптироваться под реальные условия среды передачи данных. Например, снижение скорости передачи данных на одном из каналов в автоматическом режиме позволит скорректировать количество пакетов, передаваемых этим путем, что позволит сохранить нагруженность канала на исходном уровне, позволяя использовать его как для административных целей (управление оборудованием), так и для задач мониторинга.

Основываясь на вышеизложенном, стоит отметить, что настоящая система получает прирост производительности за счет одновременного использования нескольких каналов связи: прирост заключается в возможности использования каждого из имеющихся каналов для административных целей.

Дальнейшей перспективой разработки предложенной схемы коммутации является тестирование, нахождение идеальных настроек алгоритма и окончательное внедрение в рабочую схему. Результаты, полученные в ходе работы, могут быть востребованы при создании нового программного обеспечения производителями оборудования, а также инженерами теле- и радиовещания.

## СПИСОК ЛИТЕРАТУРЫ

1. Мельник В. С. Исследование алгоритмов для решения задачи маршрутизации пакетов в компьютерной сети // XXI Междунар. конф. по мягким вычислениям и измерениям (SCM-2018). СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2018. С. 673–676.

2. Мельник В. С., Горячев А. В. Построение географически-распределенной компьютерной системы и оценка ее надежности // Изв. СПбГЭТУ «ЛЭТИ». 2019. № 5. С. 19–25.

3. How PCC works. URL: [https://wiki.mikrotik.com/wiki/How\\_PCC\\_works\\_%28beginner%29](https://wiki.mikrotik.com/wiki/How_PCC_works_%28beginner%29) (дата обращения 13.05.2019).

4. AWS Auto Scaling. URL: <https://aws.amazon.com/autoscaling/> (дата обращения 13.05.2019).

V. S. Melnik

Saint Petersburg Electrotechnical University

## MULTI-THREAD ROUTING IN THE AUTOMATED SYSTEM OF ACCOUNT OF FINITE NODES OF THE DISTRIBUTED CLUSTER

*This article describes the approach to designing a geographically distributed system that uses multi-threaded routing. This mechanism allows you to increase the fault tolerance of the designed system as a whole through the use of more than one communication channel between the end nodes.*

*In addition to solving the problem of improving reliability, the article outlines a technique that allows balancing resources, allowing you to use all available communication channels.*

*The aim of the research is to develop a comprehensive solution for the design of the infrastructure for transmitting a continuous data stream by several routes, which ensure fault tolerance and share the load among themselves, allowing these routes to be used for administrative purposes, for example, to monitor the system and track critical situations, as well as to manage existing nodes.*

*In the proposed design scheme, it is proposed to ensure the fault tolerance of a geographically distributed system so that each of the message chains acts as an independent unit of the computer system. In the results of the work, an analysis was made of various options for connecting the system taking into account the probability of failure of each of the independent links.*

**Network infrastructure design, fault tolerance, reliability, redundancy, load balancing, routing, distributed systems, cluster systems**

УДК 004.387

И. В. Герасимов, С. А. Кузьмин, Н. М. Сафьянников  
Санкт-Петербургский государственный электротехнический  
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

А. В. Рудинский  
ЗАО «НПЦ «Аквamarin»

## Сложнофункциональные блоки вычислительной перцепции в классе вычислительно-преобразовательных цепей

*Исходя из совокупности обобщенных свойств физической символьной системы (ФСС), рассматривается структурно-аналоговая архитектура средств вычислительной перцепции. При этом привлекаются различные уровни абстрагирования природоподобного объекта. Применение архетипов в классификации вычислительно-преобразовательных цепей (ВП-цепей) на операционных с управляемым параметром элементах обеспечивает возможность систематизации не только имеющихся, но и мыслимых структур. Исходя из онтологической интерпретации структурно-функциональной модели, ее качество определяется качеством понятий и отношений между ними. В условиях дестабилизации воздействий внутренней и внешней среды обеспечение равновесного состояния цели обеспечивается с помощью механизма гомеостазиса. Для математического описания структуры ВП-цепей как потоковых устройств нейросетевой архитектуры используются результаты аппроксимации воспроизводимых зависимостей «вход-выход» в классе рациональных функций (по параметру управления проводимостью операционных элементов). При описании класса объектов и множеств отображений (морфизмов), переводящих одни объекты в другие, привлекается теория категорий. Общая задача синтеза сводится к генерации полного множества неизоморфных циклически связанных неприводимых структур s-моделей ВП-цепей. Приводится процедура синтеза минимальной по рангу s-модели ВП-цепи. Обсуждаются примеры оригинальных разработок, где использованы свойства вычислительной перцепции в природоподобных объектах искусственного происхождения.*

**Вычислительная перцепция, вычислительно-преобразовательные цепи, структурно-аналоговые вычисления, s-моделирование, теория категорий, функтор, структурное число, рельефная виртуализация, сенсор, тремор**

Согласно современным представлениям, перцепция – «суперпозиция трех полевых структур:

поля сознания, поля обработки информации и поля мышления». Привлекаемый формализм име-