## Моделирование вычислительных систем с логическими магистралями

### Введение

Многоканальные магистральные вычислительные системы предназначены для обеспечения эффективной коммуникации между элементами вычислительной системы. Предполагается, что в МРПВС широко используется модель клиент-сервер. Основной задачей любого узла-сервера является предоставление узлам-клиентам информационных и вычислительных ресурсов.

В МРМВС может одновременно решатся целый ряд независимых либо взаимосвязанных задач. Взаимосвязанные задачи, это чаще всего одно задание характеризующиеся высокой вычислительной сложностью, решение которого на базе одного вычислительного элемента является сложным, а в ряде случаев просто невозможным. Процесс разбиения задания на частично независимые фрагменты называть будем разделением. Причиной проведения разделения является не только требование распараллеливания решения задачи о большой трудоемкости. В ряде случаев оно необходимо для того, чтобы вычислительный процесс отражал структуру создания, обработки и использования результатов его проведения. Такое требование, в частности, стоит перед системами использующими модель вычислений клиент-сервер, кооперативную модель, распределенные модели вычислений: кластерную, параллельную [7].

Применение в высокопроизводительной параллельной системе любого из методов распределения обработки неуклонно приводит к возникновению нового рода издержек вычислительного процесса. Кроме издержек связанных с проведением вычислений, появляются издержки связанные с необходимостью передачи сообщений между вычислительными элементами обрабатывающими взаимосвязанные фрагменты задания. Издержки те определяются как коммуникационные издержки. Вторая группа издержек нового класса это издержки синхронизации, связанные с необходимостью обеспечения синхронного выполнения отдельных частей задачи. Дополнительные издержки необходимы также для обеспечения работы операционной системы, которая решает гораздо более сложные проблемы, чем в случае однопроцессорного решения задачи. В дальнейшем интересовать нас будут исключительно коммуникационные издержки. Во-первых, потому что их участие в суммарном значении дополнительных задержек достаточно высокое. Во-вторых, потому что с точки зрения применения предложенных ранее методов улучшения параметров многоканальных магистральных сетей уменьшение их значения является возможным.

Способов анализа параллельных вычислительных систем основанных на применении многоканальных магистральных топологий следует искать среди методов анализа компьютерных сетей либо мультипроцессорных систем. В основном для этой цели применяются два класса методов. Первый из них основан на вероятностном анализе параметров системы, второй на использовании методов теории очередей.

### Допущения процесса моделирования

Одной из важнейших характеристик моделирования, произвольной вычислительной системы, являются вычислительные ресурсы необходимые для его проведения. К наиболее существенным ресурсам следует отнести объем требуемой оперативной памяти ВС на которой производится моделирование и время, необходимое для его проведения. Если вышеназванные параметры, превышают возможности инструментальной вычислительной системы, то решение задачи моделирования невозможно, либо достигается оно за неприемлемое для пользователя время.

Основной целью проводимого вычислительного эксперимента, являются: исследование влияния многоканальности и различных типов кластеризации на работу ВС. Ограниченный набор воздействий и исследуемых характеристик создает предпосылки для упрощения используемой модели, а тем самым и к уменьшению ресурсов необходимых для моделирования. При этом точность определяемых параметров ухудшается только незначительно.

Создание модели, учитывающей все свойства вычислительной системы возможно, однако ее решение требует чрезвычайно больших вычислительных ресурсов. Поэтому, с целью проведения процесса моделирования, следует применить упрощенную (приближенную) модель, сложность решения которой приемлема, а точность достаточно высока. При создании модели МРМВС, предлагается использовать следующие рассуждения:

1. Модель должна описывать систему, работающую в синхронном режиме, которой циклы определяются центрально (тогда, самая модель тоже синхронная). Такое допущение, позволяет использовать вероятностные методы анализа, основанные на биномиальном распределении. Методы данного класса характеризуются незначительной вычислительной сложностью, при высокой точности моделирования. Численное решение уравнений модели, можно осуществить используя общеизвестные итерационные методы. Применение данного ограничения не критично, т.к. большинство магистральных систем передачи имеет асинхронный характер. Если по каким либо причинам, система не может исследоваться как синхронная, в процессе анализа следует применить модели, основанные на методах теории очередей;
2. Вычислительная сложность решения модели уменьшается, когда анализируемая ВС является однородной. Однородность, прежде всего, относится к запросам узлов-клиентов. Однородными, называем запросы, идентичные по вычислительной сложности, выполняемые за одно и то же время, независимо от применяемого для вычислений узла-сервера. Однородность, дополнительно означает постоянство вычислительных мощностей узлов-серверов и не относится к их архитектуре. Если исследуемая ВС имеет партнерский характер, а каждый из ее узлов снабжен многопользовательской операционной системой, необходимо также обеспечить тождественность всех ее узлов;
3. Запросы узлов-клиентов создаются исключительно в начале рабочего цикла. Это допущение, обеспечивает синхронный режим работы модели (допущение 1). Благодаря этому, вероятность  появления в системе новых запросов, одновременно является частотой их возникновения. В реальных системах, запросы появляются на протяжении всего цикла, что для более точного моделирования требует применения асинхронных моделей ВС. Возможность возникновения запроса в произвольном моменте, чревато также тем, что завершение его обработки может не совпадать с концом данного рабочего цикла, что дополнительно затрудняет моделирование;
4. Запросы на проведение вычислительных работ создаются исключительно узлами клиентами и относятся только к узлам-серверам. Характерное для кооперативной модели вычислений, общение между серверами, здесь не появляется;
5. Распределение запросов между узлами-клиентами равномерное. Поэтому, вероятность создания нового запроса идентична для всех клиентов и равна . Благодаря допущению, с точки зрения запросов, система однородная, и каждый узел-клиент вносит в нее идентичную нагрузку. Это является предпосылкой упрощения моделирования. В реальных вычислительных системах (в том числе в МРМВС), распределение запросов между узлами-клиентами нормальное, т.е. имеет неоднородный характер. Учет неоднородности требует применения неоднородных моделей запросов узлов-клиентов и является сложным с вычислительной точки зрения;
6. Вычислительная сложность моделирования уменьшается также, тогда когда распределение запросов между узлами-серверами является однородным. Это означает, что вероятность появления (частота выступления) запроса к заданному узлу, одинакова для всех серверов. В реальных системах, распределение запросов между узлами-серверами тоже нормальное. Однако, для максимизации производительности ВС, следует обеспечить идентичную, по количеству запросов, нагрузку всех узлов-серверов. Это требует применения одного из методов выравнивания нагрузок ВУ;
7. Применение эффективных вероятностных методов моделирования требует независимости вычислительных запросов. Поэтому, запросы, которые по любым причинам созданы, но не приняты на обработку в течение данного цикла, отбрасываются и дальше уже не рассматриваются. В реальных ВС, откинутые вычислительные запросы рассматриваются повторно. Так как, в нормально нагруженных вычислительных системах, количество откинутых запросов невелико, появляющаяся, за счет этого, ошибка невелика и не превышает нескольких процентов.
8. Важнейшей причиной ухудшения производительности МРМВС, являются ограничения в доступе запросов к ресурсам сервера. Они появляются в результате занятости (либо отказа) узла-сервера, а также недостаточного, для передачи появляющиеся запросов, количества логических магистралей. Поэтому, если целью моделирования, является определение влияния на работу системы занятости логических магистралей, то количество  клиентов и количество  серверов должно быть одинаковым, т.е. . Тогда, моделирование сводится к исследованию влияния, на работу ВС, недостаточного количества логических магистралей. В реальных системах, выполнение этого условия мало вероятно. Обычно, количество  узлов-клиентов значительно больше количества  узлов-серверов. Его приближенные выполнение (т.е. ) реально в системах основанных на партнерской модели клиент-сервер;
9. Недоступность ресурсов узла-сервера, может быть также результатом его занятости. Поэтому, если моделирование производится с целью определение влияния на работу МРМВС коллизий, появляющиеся при доступе к узлам-серверам, то количество  логических магистралей должно быт больше либо равно количеству  узлов-серверов (т.е. ). В идеальном (но мало реальном из-за стоимости) случае, количество  ЛМ, должно быть равно количеству  узлов-клиентов. В реальных системах, в которых количество вычислительных узлов неизменно. При этом следует стремиться к состоянию, когда количество  логических магистралей сравнимо со значением , но значительно меньше размера . В партнерских системах клиент-сервер, эти требования могут не выполняться;
10. В реальных вычислительных системах, цикл обслуживания запроса состоит из трех основных этапов: выдачи запроса узлом-клиентом, его приема и обработки узлом-сервером и выдачи результатов обработки узлом-сервером. В большинстве моделей предполагается, что объем запроса значительно меньше размера результатов его обработки. Такое предположение, справедливо и для МРМВС, при условии, что в них используется модель клиент-сервер либо кооперативная модель вычислений. Если в ВС, применение находит модель распределенных вычислений, то размеры потоков трудно предусмотреть. Для упрощения вычислительной сложности моделирования предлагается, что все три этапа обслуживания запросов, это одно целое и они реализуются в течение одного и того же цикла работы системы. Если целью моделирования является исследование влияния магистральных связей на работу системы в целом, то данное ограничение не ухудшает его точности. Однако, когда моделирование должно учитывать характеристики обработки запросов серверами, тогда данного допущения нельзя применить.
11. Если все появляющиеся в системе запросы равноправны (т.е. для них не используются приоритеты), то вычислительная сложность решения модели уменьшается. Это означает, что все запросы однородны, благодаря чему нет необходимости учета различных классов запросов. В реальных системах, запросы можно дополнительно описать приоритетами, для учета которых, в вероятностных методах анализа предлагается использовать условную вероятность. К сожалению, это вызывает увеличение вычислительной сложности модели.

Классификация методов предложенных методов, упрощения вычислительной сложности моделирования многоканальных реконфигурируемых магистральных ВС представлена на рис.4.18.



Рис.4.19.Классификация методов упрощения моделирования МРМВС.

Одновременный учет всех вышеперечисленных ограничений (допущений), гарантирует создание простой, несложной в решении модели МРМВС, низкая точность которой ограничивает диапазон ее применений. В дальнейшем, будут применяться ограниченные наборы допущений, состав которых определяется целью моделирования и применяемым методом исследований.

### Средства анализа многоканальных систем с логическим магистралями

В литературе описано множество методов предназначенных для анализа вычислительных систем, в том числе и параллельных. Однако, все они не учитывают особенностей многоканальных систем. Это приводит к недостаточной точности результатов моделирования либо (и) высокой сложности их получения. Ниже охарактеризованы выбранные методы анализа различных разновидностей МРМВС, использование которых, считается предпочтительным. Для каждого из предлагаемых методов определяется меры эффективности и интенсивности использования ВС. Определим понятия эффективности и интенсивности.

**Определение 4.1.** Эффективностью использования ВС будем называть степень соизмерения результатов функционирования вычислительной системы с затратами. В частности, эффективность определяет уровень использования производственных мощностей системы. Данный тип эффективности называть будем также технической эффективностью.

**Определение 4.2.** Интенсивностью использования ВС будем называть среднее количество заданий поступающих в систему за единицу времени.

Характеристикой любой ВС, проявляющей особое влияние на используемые методы ее анализа, является временная организация модели вычислительной системы. Различаем синхронные и асинхронные модели. Заметим, что синхронность модели, довольно свободно связана с синхронностью ВС. Это означает, что допустима ситуация, когда асинхронная вычислительная система описывается синхронной моделью.

Для анализа синхронных моделей, т.е. таких, в которых запросы появляются исключительно в начале рабочего цикла и управляются центральным таймером системы, предлагается использовать один из описанных ниже методов моделирования. Первый из них, основан на предположении, что в реальной системе, связь между узлом-клиентом и узлом-сервером реализована на основе коммутации цепей. Здесь, производительность и эффективность МРМВС определяется на основе количества альтернативных путей, по которым запрос может быть передан от узла-клиента к узлу-серверу, а также вероятностных характеристик самого запроса. Параметром, описывающим интенсивность употребления системы, является средняя частота  появления запросов, которая из-за синхронного характера модели, заменяется вероятностью  создания узлом-клиентом вычислительного запроса. Параметром, отражающим эффективность использования моделируемой ВС, является ее пропускная способность , определяемая количеством вычислительных запросов обслуживаемых в течение заданного промежутка времени. Мерой эффективности может быть также вероятность  принятия запроса, определяющая вероятность обработки запроса клиента соответствующим сервером. Значение , это соотношение ожидаемой пропускной способности к количеству запросов, ожидаемых в течение рабочего цикла ВС. Предложенный метод обладает удовлетворительной точностью, а необходимые ресурсы чрезвычайно малы. Кроме этого, позволяет он эффективно учитывать характерные свойства многоканальных магистральных систем.

Второй метод, предназначен для систем с коммутацией пакетов. Модель, основана на вероятностных параметрах очередей описывающих ВС, типы используемых очередей, существенным образом зависят от моделируемой архитектуры и задач стоящих перед процессом моделирования. Показателем интенсивности использования ВС, является тоже средняя частота запросов (вероятность  создания запроса). Эффективность использования системы описывается здесь коэффициентом  использования сервера. Метод обладает хорошей точностью, а ресурсы необходимые для его реализации тоже невелики. Несмотря на видные преимущества, он менее пригоден для анализа МРМВС, что объясняется используемым типом коммутации.

Выбор способа коммутации, это одно из наиболее существенных решений принимаемых на этапе проектирования ВС. В рассматриваемых многоканальных магистральных вычислительных системах, допустимо использование обоих типов коммутации, т.е. коммутации цепей либо пакетов. При коммутации цепей, узел-клиент занимает логическую магистраль на все время обработки запроса, т.е. его выдачи клиентом, приема и обработки, а также выдачи результатов обработки сервером. В классических ВС, коммутация цепей понижает эффективность использования системы (в частности, из-за продолжительной процедуры составления цепи) при обработке сложных запросов и повышает, когда обработке подлежит большое количество простых запросов. Если обрабатываются запросы различных типов, то в классических системах, целесообразно применять смешанные методы коммутации. Перечисленные выше ограничения не распространяются на многоканальные системы с логическими магистралями. Здесь, запросы передаются в вещательном режиме, по сверхвысокоскоростных логических магистралях. При чем, в случае занятости одной из них, запрос может передаваться по другой логической магистрали (если это предусматривает алгоритм трассировки). Предлагаемый способ коммутации, за счет использования вещательного режима передачи, объединяет в себе достоинства коммутации цепей и пакетов. Использование в МРМВС коммутации пакетов оказывается затруднительным. Это связано, прежде всего, с высокой стоимостью и значительными временными задержками коммутаторов пакетов.

В асинхронных моделях, запросы не синхронизированы системным таймером и могут появляться и обслуживаться в произвольном моменте рабочего цикла. Их можно построить для систем, как с коммутацией цепей, так и коммутацией пакетов. При этом используемая модель зависит от применяемого типа коммутации. Для анализа систем с коммутацией цепей, рекомендуется использовать модель, основанную на исследовании сети очередей с неограниченными буферами. Интенсивность их использования, описывает время  реакции сервера. Для многоканальных, многопользовательских систем, использующих множество серверов (именно такими являются МРМВС), время между завершением и созданием очередного запроса, описывается экспоненциальным распределением. Показателями эффективности функционирования такой системы являются: пропускная способность  системы либо коэффициент  использования серверов. Метод характеризуется хорошей точностью, достигаемой за счет высокой вычислительной сложности. По этим причинам он не рекомендуется к широкому использованию.

Построение моделей асинхронных систем с коммутацией пакетов предлагается выполнить на основе метода эквивалентных потоков. Здесь, узлы-серверы и система логических магистралей объединяются и замещаются единственным потоком, которого истоком является сервисный центр модели. Интенсивность использования, оценивается временем  реакции сервера, а ее эффективность, с помощью пропускной способности  либо коэффициента  использования серверов. Для широкого диапазона нагрузок (коэффициентов использования), метод характеризуется достаточной точностью, а вычислительные ресурсы необходимые для решения модели, находятся в допустимых пределах.

Кроме методов основанных на анализе синхронных либо асинхронных моделей, находят применение методы, в которых модель учитывает свойства обоих типов моделей (синхронных и асинхронных). Они базируют на применении дискретных вероятностей, а их достоинством является неизменная трактовка, всех возможных методов управления доступом к логическим магистралям, что характерно для синхронных, а невозможно при использовании асинхронных моделей. Построение моделей данного класса, основано на применении методов гарантирующих независимый анализ очередей отдельных элементов системы. Параметром эффективности использования системы, является коэффициент  использования серверов системы. Методы этого класса не будут дальше рассматриваться.

### Анализ систем с однородной моделью запросов

Анализу подвергается вычислительная система состоящая из  серверов и  клиентов объединенных с использованием  логических магистралей. С целью упрощения процесса моделирования, используются допущения 1-7 и 10-11. Упрощение вычислительной сложности моделирования достигается, прежде всего, благодаря допущениям 2 и 7. Однородность вычислительных запросов (допущение 2) значительно упрощает сложность вычислений, при этом однако, точность моделирования незначительно ухудшается. Предположение о независимости запросов (допущение 7 – приближение Стрекера) неприемлемо даже в реальных системах. При этом, его использование не вызывает значительного ухудшения точности моделирования, а вычислительная сложность решения модели уменьшается.

С математической точки зрения предлагаемая модель основана на многократном, последовательном проведении опыта Бернулли, который имеет исключительно два результата: успех либо поражение. Так как все проводимые опыты являются полностью независимыми и одинаково вероятными, вероятность успеха при их последовательном повторении и является мерой пропускной способности канала передачи. Таким образом, с целью определения пропускной способности логической магистрали определять будем среднее количество запросов к узлам-серверам, со стороны узлов-клиентов. Согласно допущений 2 и 3 узлы-клиенты выдают запросы одновременно в начале цикла, их сложность идентична, все они решаются в течение одного рабочего цикла.

Пусть каждый из  клиентов с вероятностью  создает, в течение одного рабочего цикла, один запрос к произвольному узлу-серверу. Тогда, ожидаемое количество  запросов узлов-клиентов, в течение рабочего цикла определяется как . Его значение может превышать допустимое количество сообщений передаваемых по используемым каналам. В результате, возникают коллизии двух взаимосвязанных типов. Во-первых, коллизия появляется тогда, когда более чем один узел-клиент, в тот же момент времени обращается с вычислительным запросом к одному и тому же серверу. Это возможна, т.к. связь вычислительных узлов (клиентов и серверов) многоканальная. Во-вторых, коллизия появляется в случае, когда количество узлов-клиентов, нуждающееся одновременной передаче запроса, больше количества магистралей. Это реально, т.к. в рассматриваемых системах, количество  магистралей, значительно меньше количества клиентов. Кроме этого, в МРМВС, коллизии возникают независимо от количества клиентов, серверов либо связывающих их магистралей и являются последствием одновременного создания запросов. Из-за наличия коллизий, пропускная способность  всегда меньше максимального количества  запросов, создаваемых в системе в течение одного рабочего цикла. Согласно принятых допущений, вероятность появления запроса в течение цикла к заданному серверу, равна  и неизменна для любого узла-клиента либо узла-сервера.

Рассмотрим вычислительные запросы генерируемые -тым узлом-клиентом и адресованные -ым узлом-сервером. Если вероятность создания узлом-клиентом в течение цикла, любого запроса равна , то вероятность формирования любым их них не менее одного запроса адресованного именно к -тому серверу определяется формулой:

,

где:  – вероятность того, что любой из  узлов-клиентов, в течение цикла, не создает запроса к -тому узлу-серверу. Если в рассматриваемой архитектуре количество  магистралей больше количества  серверов (т.е. ), то ограничения пропускной способности , будут исключительным результатом коллизий появляющиеся при доступе к узлам серверам. Тогда, вероятность  создания запроса к любому из  узлов-серверов, а тем самым и пропускная способность  многоканальной системы будет равна:

,

где:  –  пропускная способность МРМВС, ограниченная недостаточным количеством серверов. Так как, значение пропускной способности определяется вероятностью доступа всех  узлов-клиентов, ко всем  серверам, и в реальных системах , значение  зависит, прежде всего от количества узлов-серверов системы.

Вторым существенным ограничением пропускной способности МРМВС, является недостаточная коммуникационная емкость самих логических магистралей (в частности их недостаточное количество). Поэтому, рассмотрим функционирование магистральной вычислительной системы, в которой одновременно появляется точно  вычислительных запросов, адресованных к  различным узлам-серверам. Если количество запросов , не больше количества  логических магистралей (т.е. ), то все вычислительные запросы обслуживаются без препятствий и пропускная способность не лимитируется. В противном случае, блокируется  запросов. Если все вычислительные запросы независимые (допущение 7), то вероятность  появления точно  произвольных запросов, а тем самым и пропускная способность  вычислительной системы, определяется на основе биномиального распределения и равна:



И так, пропускная  способность ВС с недостаточным количеством логических магистралей, соответствующая вероятности появления в системе точно  запросов, является произведением: количества комбинаций  запросов для  узлов-серверов, вероятности создания  запросов к узлам-серверам и вероятности того, что  запросов, просто не появиться.

Пропускная  способность вычислительной системы, с учетом двух рассмотренных выше типов коллизий, определяет формула:

.

Предложенная выше модель, предназначена для анализа вычислительных систем с полными многоканальными магистралями (см. рис.4.10). Однако, после соответствующих модификаций можно нее для определения параметров других, описанных раньше, архитектур магистральных сетей. Рассмотрим многоканальную магистральную архитектуру с однократным подключением клиентов (см. рис.4.11).

В вычислительной системе данного типа, пропускная способность  является суммой пропускных способностей отдельных магистральных каналов. Поэтому, определение , начинается с определения характеристик каждой отдельно взятой логической магистрали. Пусть  означает пропускную способность вычислительной системы построенной исключительно с использованием -той логической магистрали. Тогда, пропускная способность всей многоканальной вычислительной системы:

.

Определим пропускную способность , вычислительной системы использующей, исключительно одну (-тую) логическую магистраль. Пусть к используемому, -тому магистральному каналу подключено  узлов-клиентов, каждый из которых, связан исключительно с данной логической магистралью. Используя введенные раньше обозначения, можно записать, что вероятность появления в течение цикла к заданному узлу серверу равна . Тогда, вероятность создания не менее одного вычислительного запроса к -ому узлу-серверу определяется на основе выражения и для узлов-клиентов равна:

.

Если принять, что количество узлов-серверов равно , а каждый из них подключен ко всем логическим магистралям, то выражение определяет пропускную способность  системы состоящей исключительно из -той логической магистрали. Тогда, формула описывает пропускную способность полной (т.е. использующей все  логических магистралей) вычислительной системы. Анализ модели показывает, что улучшение пропускной  способности системы, достигается исключительно целенаправленным перераспределением узлов-клиентов между магистралями. Предложенная модель пригодна также ля анализа кластерных многоканальных (см. рис.4.12) и симметричных многоканальных кластерных магистралей (рис.4.13).

Рассмотрим МРМВС с объединенными магистралями (см. рис.4.14), состоящую из  логических магистралей, разбитых на  групп. С точки зрения узлов-клиентов, любая группа является полной магистральной системой, для узлов-серверов это система, в которой увеличено количество магистралей объединяющих независимые группы. Пропускная способность данной системы растет, именно благодаря увеличению количества используемых магистралей.

Пусть распределение элементов (логических магистралей и узлов-клиентов) равномерное, а узлы серверы подключаются ко всем логическим магистралям (что обеспечивает эффективный доступ к произвольному узлу-серверу, любого узла клиента). Тогда, любая из  групп состоит из  клиентов и  магистралей, количество серверов каждой их групп равно . Вероятность  создания не менее одного запроса к -ому узлу-серверу, любым из  узлов-клиентов -той группы, можно определить модифицируя формулу к виду:

.

Так как узлы-серверы одновременно подключаются ко всем вычислительным группам, узлы-клиенты -той группы не являются единственными образующими запросы к -ому серверу. Поэтому выражение преобразуем к виду учитывающему появление запросов от всех  групп. Вероятность создания  группами клиентов, любого вычислительного запроса к -ому узлу-серверу, определяется выражением:

.

Если в вычислительной систем выполняется неравенство , то вероятность  создания -той группой клиентов, запроса адресованного к любому серверу можно определить формулой:

.

Вероятность , это по сути пропускная способность, системы состоящей из целевого множества узлов-серверов и -той группы узлов-клиентов.

Применение предложенной организации особо эффективно в случае, когда вычислительные запросы локализованы, т.е. когда узлы-клиенты произвольной группы, обращаются преимущественно к предписанной им группе, состоящей из  серверов. Заметим, что данное допущение не связано с архитектурой связей, а исключительно с распределением запросов. Тогда, если допущение 7 выполнено, то вероятность  создания -той группой узлов клиентов точно  вычислительных запросов представлена выражением:

,

а пропускная способность -той группы, МРМВС с объединенными магистралями равна:

.

Пропускная способность  всей системы равна сумме пропускных способностей всех  групп, т.е.

.

Если в анализируемой вычислительной системе локализация ресурсов не используется, то вероятность появления в целой ВС точно  запросов на доступ ко всем  узлам-серверам определяется формулой:

,

а пропускную способность можно определить выражением:

.

Описанные выше модели МРМВС основывались на предположении, что в вычислительной системе выступают два типа коллизий, понижающих пропускную способность ВС. Основным недостатком такого подхода является отсутствие тесной связи между вычислительными узлами (клиентами и серверами) а логическими магистралями. Заметим, что пропускная способность ВС с учетом ограничений вносимых ограниченным количеством логических магистралей в рассмотренных моделях выражается вероятностью создания  произвольных запросов к узлам-серверам.

Предложенные выше модели, для описания количества комбинаций вычислительных запросов к узлам-серверам, используют бином Ньютона. Поэтому, учет взаимосвязи между количеством доступных узлов-серверов, количеством узлов клиентов генерирующих вычислительный запрос и количеством вызываемых ними серверов является затруднительным. Этот недостаток в системах, в которых отсутствует локализация запросов, оказывается существенным. Поэтому, вместо бинома Ньютона предлагается использовать числа Стирлинга второго рода.

Число Стирлинга второго рода записывается как  и определяет количество способов разделения  элементного множества на  непустых подмножеств [XX]. Его значение можно определить с помощью рекуррентной формулы:

.

Предлагаемый совершенствованный метод, также основан на предположении, что пропускная способность системы ограничивается коллизиями при доступе к узлам-серверам и логическим магистралям. В частности, пропускная способность определенная с учетом коллизий при доступе к серверам, уменьшается о величину, которая не может быть использована из-за ограничений системы логических магистралей.

Пропускная способность  учитывающая исключительно коллизии при доступе к серверам, определяется с использованием выражения . Размер уменьшения полной пропускной способности, определяется как произведение количества недоступных (занятых) серверов и вероятности появления вычислительных запросов, которые не могут быть обслужены из-за недостаточного количества магистралей. Если в системе можно использовать биномиальное распределение, то вероятность выдачи узлами-клиентами точно  запросов, определяется модифицируя выражение . Однако, из-за недостаточного количества логических магистралей (), обработка всех запросов не возможна, а первый член произведения (количество недоступных серверов), является суммой биномиальных распределений, для которых  выполняет условие , т.е.:.

Для определения второго сомножителя (вероятности появления необслуживаемых запросов) используем число Стирлинга второго рода. Пусть вероятность  создания узлами-клиентами  запросов, адресованных к  из  узлов-серверов, обозначена символом  и определяется с использованием выражения:

.

В дальнейших исследованиях, учитываются исключительно те запросы, которые не могут быть обслужены из-за недостаточного количества логических магистралей (т.е. ). Поэтому выражение преобразуется к виду:

.

Тогда, второй член произведения (вероятность появления не обслуживаемых запросов) следует представить в виде суммы вероятностей описанных формулой , для всех узлов-серверов превышающих количество логических магистралей, что можно записать как: . И так, пропускная способность многоканальной магистральной системы, определяется формулой:



Предложенная модель более точно учитывает взаимосвязь между узлами-клиентами, узлами-серверами и логическими магистралями, благодаря чему, обеспечивает более точное моделирование.

### Неоднородная модель запросов

В предыдущих моделях МРМВС, называемых однородными, предполагалось, что любой узел-клиент, создает с идентичной вероятностью запросы к произвольному серверу. Использование в реальных системах однородной модели запросов не отражает процессов в них происходящих. В любой системе вычислительные запросы распределены неравномерно, всегда существует группа узлов-серверов, к которой узлы-клиенты обращаются в первую очередь. Введем два существенных определения.

**Определение 4.3.** Предпочтительным узлом-сервером, заданного узла-клиента будем называть сервер, вероятность обращения к которому со стороны клиента самая высокая.

**Определение 4.4.** Логическим кластером будем называть множество вычислительных узлов (клиентов и серверов), для которых вероятность создания вычислительного запроса любым клиентом кластера, для всех его серверов идентична.

Таким образом логический кластер является виртуальной структурой объединяющей клиентов и соответствующие им серверы. К логическому кластеру принадлежать будут все узлы-клиенты и выбранные узлы-серверы, вероятность обращения к которым одинакова. Количество логических кластеров равно числу различных вероятностей, с которыми клиенты запрашивают выбранные серверы. Для упрощения процесса моделирования дополнительно предположим, что распределение запросов между кластерами является неизменным для каждого клиента.

Очевидно, что для описания неоднородной модели запросов недостаточно использования вероятности  создания запроса узлом клиентом. Для этого требуется дополнительно ввести вероятность  обращения вычислительного запроса к -тому логическому кластеру. И так, вероятность создания запроса к первому логическому кластеру будем обозначать символом , а запросы данного класса будем называть запросами первого типа. По аналогии, вероятность создания запроса второго типа, адресуемых к узлам-серверам второго логического кластера будет равна . Пусть количество типов (значений вероятностей обращения) вычислительных запросов будет равно . Тогда количество логических кластеров запросов равно тоже . При этом вероятность создания запроса к -тому логическому кластеру обозначается как .

Если в системе произведена локализация вычислительных запросов, то узлы клиенты будут обращаться, прежде всего, к своим предпочтительным серверам. Для вероятностей  обращения к серверам -того логического кластера выполняется неравенство:

,

т.е. обращение к серверам логического кластера первого типа наиболее вероятно. Кроме этого, сумма вероятностей  для любого узла клиента равна единице, т.е. справедливо равенство

.

Согласно вышесказанного, вероятность создания узлом клиентом, запроса первого типа равна , второго  и т.д. Обобщая, запишем, что вероятность создания клиентом запроса -того типа равна .

Рассмотрим пропускную способность МРМВС с недостаточным количеством узлов-серверов. Пусть количество серверов -того логического кластера равно . Тогда вероятность создания узлом-клиентом вычислительного запроса -того типа к заданному серверу равна . В состав логического кластера -того уровня входят все узлы-клиенты ВС и  из узлов серверов. Поэтому, вероятность , создания любым из  клиентов не менее одного вычислительного запроса -того типа к заданному -ому серверу -того логического кластера равняется:

.

Логический кластер -того уровня состоит из  узлов-серверов, поэтому вероятность  создания вычислительного запроса -того типа к любому из них, определяет формула:

.

Если количество типов запросов равно , то вероятность  создания вычислительного запроса произвольного типа к -ому серверу равна:

.

Учитывая выражение , формулу можно преобразовать к виду:



Если запросы предъявляемые клиентами независимые (допущение 7), то вероятность  появления в течение цикла работы ВС, точно  произвольных запросов определяется с использованием модифицированной формулы , а пропускная способность с учетом обоих типов коллизий по измененной соответствующим образом формуле . Предложенная модель является видоизмененной версией модели для однородных запросов и предназначена для анализа полной многоканальной магистральной системы. Однако, может она успешно использоваться для анализа всех рассмотренных раньше магистральных систем. В качестве примера рассмотрим использование метода для анализа ВС с симметричной многоканальной кластерной магистралей (см. рис.4.13).

Пусть коэффициент перекрытия магистралей , т.е. кластеры узлов-клиентов по магистралям не перекрываются. Это означает, что МРМВС можем рассматривать как топологию с объединенными магистралями, в которой количество  групп равно количеству  ее логических кластеров. Предположим, что количество логических магистралей, обслуживающих -тый логический кластер равно , а количество его клиентов . Тогда, вероятность создания не менее одного запроса -того типа к -ому узлу-серверу, клиентом -того логического кластера равняется:

,

где:  - количество серверов -того кластера запросов. Предполагая, что количество типов запросов равно , вероятность  создания вычислительного запроса произвольного типа к -ому серверу, клиентом -того логического кластера можно определить преобразуя к виду:

.

Если запросы предъявляемые клиентами независимые (допущение 7), то вероятность  появления в системе точно  вычислительных запросов, адресованных к -ому серверу, создаваемых клиентами -того логического кластера можно определить модифицируя формулу , а пропускную способность логического кластера на основе выражения . Полная пропускная способность является суммой пропускных способностей всех логических кластеров, т.е.

,

где:  - пропускная способность -того кластера определена на основе модифицированной формулы .

Предложенный метод сохраняет все преимущества методов данного класса, т.е. достаточную точность и очень низкую вычислительную сложность решения модели. Основным ее недостатком является ограниченный учет влияния характеристик логических кластеров на работу узлов-серверов ВС. Отмеченный недостаток можно устранить применяя рассмотренную дальше иерархическую модель запросов. Неоднородную модель запросов можно использовать тоже для анализа других архитектур МРМВС.

Модель многоканальной реконфигурируемой вычислительной системы использующую числа Стирлинга второго рода (см. §4.5.4) можно также приспособить к работе с неоднородной моделью запросов. Для этого, достаточно заменить соответствующие члены в формуле . Тогда пропускная способность  будет определена формулой

,

где:  определяется с использованием выражения , а  модифицированной формулы .

### Иерархическая модель запросов

Несмотря на свои преимущества моделирование функционирования МРМВС с использованием неоднородной модели запросов обладает рядом существенных недостатков. Главный из них это факт, что запросы различных типов никак не связаны между собой. Ввиду этого, применение неоднородной модели для анализа иерархических магистральных систем приводит к значительным погрешностям и нецелесообразно. В иерархических системах, на различных уровнях используются различные количества логических магистралей, узлов-клиентов и узлов-серверов. Кроме этого, иерархические МРМВС строятся так, что вероятность создания вычислительного запроса тем больше, чем сервер находится ближе кластера вызывающего узла-клиента. Поэтому, для создания модели многоканальных иерархических магистралей предлагается использовать методы основанные на иерархической модели запросов.

Иерархическая модель запросов предназначена для определения влияния вероятностных характеристик запросов на работу иерархической системы в целом.

Пусть в МРМВС, состоящей из  иерархических уровней существует  типов вычислительных запросов. Все типы запросов непосредственно связаны с определенным уровнем иерархии системы. И так, запрос первого типа реализован в одном из  кластеров первого иерархического уровня. Запрос первого типа является обращением клиента к своему предпочтительному серверу. Запрос второго типа непосредственно связан со вторым иерархическим уровнем системы. Однако не включает он элементов одного из кластеров предыдущего уровня. Исключение имеет количественный характер и объясняется тем, что при определении вероятностных характеристик запроса нельзя включать элементов которые были учтены вероятностных характеристик запроса низшего уровня. Образно говоря при определении вероятностных параметров запросов второго типа следует не учитывать элементов одного кластера первого уровня, т.к. если искомый сервер являлся бы предпочтительным сервером обслуживание запроса прошло бы на предыдущем уровне иерархии, в рамках запроса первого типа.

Определим количество элементов (клиентов либо серверов) которые охватывает запрос заданного типа. Для этого введем понятие кластера запросов. Кластером запросов определять будем часть кластера иерархического уровня которая содержит исключительно те элементы к которым может быть адресован запрос. И так, кластер запросов -того уровня является частью иерархического кластера -того уровня из которого исключено один кластер -ого уровня. Рассматриваемое исключение не охватывает запросов первого типа. Поэтому количество  клиентов кластера запросов первого типа равно количеству  клиентов иерархического кластера первого уровня. Соответственно количество  серверов кластера запросов первого типа равно количеству  серверов иерархического кластера первого уровня.

Количество клиентов кластера запросов второго типа равно

.

где: - количество клиентов кластера запросов второго уровня;  - количество клиентов иерархического кластера первого уровня. С использованием можно определить также количество  серверов кластера запросов второго уровня. Однако вместо параметров относящихся к клиентам (, ) следует использовать параметры относящийся к серверам (, ). Количество клиентов кластера запросов третьего типа определяется выражением

,

а количество серверов определяем идентично как и для кластера второго уровня. По аналогии запишем, что кластер запросов -того типа состоит из  клиентов, а кластер запросов -ого типа соответственно из , где  - количество иерархи­ческих кластеров -того уровня. Обобщая можем записать, что количество клиентов кластера запросов *i*-того типа равно

.

С использованием с использованием выше перечисленных изменений можно определить также количество серверов кластера запросов.

Как и для неоднородной модели запросов вероятность создания запроса -того типа определяет параметр . Вероятность создания запроса первого типа равна , второго , соответственно .

В реальных высокопроизводительных параллельных системах вероятности запросов клиента к различным серверам различные. Иерархические магистраль­ные структуры позволяют реализовать распределение ресурсов таким образом, что вероятность появления запроса заданного типа тем выше чем меньше номер типа запроса. И так, в системе наиболее вероятны запросы первого типа, наименее запросы -ого типа. Для иерархической модели запросов справедливо неравенство

.

С точки зрения вероятностного моделирования выполнение уравнения обязательно.

С использованием предложенной иерархической модель запросов рассмо­трим анализ высокопроизводительных вычислительных систем использующих многоканальные топологии связи. Исследование системы полной многоканаль­ной магистралью опирается на исследованию этой топологической архитекту­ры для неоднородной модели запросов. И так, вероятность создания данным клиентом запроса -того типа равна , вероятность появления запроса -того к заданному северу равна , а вероятность  создания не менее одного запроса -того типа к заданному -ому серверу определяется выражением

.

Вероятность создания запроса произвольного типа к -ому серверу определяется выражением

,

Если принять, что в анализируемой системе  пропускная способность полной многоканальной системы  равна

.

Как для однородной и неоднородной модели запросов, также для иерархи­ческой модели при выполнении условия независимости запросов (допущение 7) вероятность  появления точно  запросов произвольного типа на доступ серверам системы определяется с использованием биномиального разложения. И так  можно определить по формуле

,

а пропускную способность на основании выражения

.

Учитывая результаты проведенного раньше анализа заметим, что идеи построения иерархических кластеров и кластеров запросов сходные. Поэтому применение иерархической модели запросов для анализа иерархических магистральных систем является целесообразным.

Пусть магистральная система состоит из  иерархических уровней. Тогда вероятность  создания не менее одного запроса -того типа к заданному -ому серверу определяется выражением

,

где  находится в диапазоне . Так как, количество серверов кластера -того уровня запросов равна , вероятность создания запроса -того типа к любому из серверов кластера равна

,

а при условии, что распределение вероятностей обращения к каждому из серверов однородное,  определяется формулой

.

Вероятность создания запроса произвольного типа к -ому серверу определяется выражением

.

В анализируемой системе, как и во всех других многоканальных магис­тральных системах возникают два типа коллизий. Во первых, одновременно к данному серверу обращается больше чем один клиент. Система среди множества запросов выбрать один подлежащий обработке. Во вторых, количество клиентов которые хотят связаться с сервером больше количества доступных магистралей. Тогда система должна определить которые среди запросов будут обслужены с использованием доступных магистралей. Таким образом количество запросов появляющихся на выходе клиентов не будет равна количеству запросов появляющихся на входе серверов. Пусть вероятность  относится к выходным интерфейсом клиентов, а величина  определяет вероятность появления запросов -того типа на входных интерфейсах серверов -того кластера запросов. Тогда можно предположить, что величины  и  связаны зависимостью

,

где:  - коэффициент пропорциональности учитывающий свойства метода разрешения перечисленных выше коллизий.