

УДК 004.272.43.003.13

Е.А. Башков¹, В.П. Иващенко², Г.Г. Швачич², М.А. Ткач²
¹Донецкий национальный технический университет,
²Национальная металлургическая академия Украины
bashkov@pmi.dgtu.donetsk.ua

Исследование влияния сетевого интерфейса на эффективность модульной многопроцессорной кластерной системы

Статья посвящена проблеме исследования оценок эффективности многопроцессорной вычислительной системы. Основное внимание уделяется влиянию сетевого интерфейса на оценки эффективности кластерной системы. Выведены аналитические соотношения основных оценок эффективности вычислений через параметры многопроцессорной системы.

Персональный вычислительный кластер, управляемые коммутаторы, реконфигурируемая сеть, вычислительные узлы, ускорение, эффективность

Введение

Применение параллельных вычислительных систем вызвано не только принципиальным ограничением максимально возможного быстродействия обычных последовательных ЭВМ, но объясняется практически постоянным существованием вычислительных задач, для решения которых недостаточно возможностей существующих средств вычислительной техники.

В данной работе рассматриваются так называемые "блейд" серверные решения многопроцессорных систем, при которых несколько однотипных материнских модулей устанавливаются в одном корпусе [1]. Практика показывает, что блейд-системы более компактны и удобны в обслуживании, а их реализация не намного дороже по сравнению с многопроцессорными компьютерными комплексами. Но благодаря растущему спросу и предложению "блейд" конфигураций на нашем рынке была сконструирована такого рода кластерная вычислительная система. Основные особенности конструирования ее архитектуры изложены в [2]. В конфигурации кластера было избрано шесть лезвий и модульный принцип его реализации. Это обеспечивает в случае необходимости его расширение за счет установки дополнительных модулей. Каждый узел работает под управлением собственной копии операционной системы. Состав и мощность узлов такого кластера может меняться, что позволяет создавать неоднородные системы. Коммутирующая сеть соединяет процессоры друг с другом. Система содержит отдельную реконфигурируемую сеть для обмена данными между вычислительными узлами, дополнительные управляемые коммутаторы, которые работают параллельно, промежуточные буферы памяти коммутаторов, а также

предусматривает сетевую загрузку узлов и механизм резервирования ключевых компонентов.

Практически одновременно с появлением первых многопроцессорных систем возникла необходимость в оценке их эффективности, производительности, быстродействия и в последующем сравнении подобных вычислительных систем, учитывая эти параметры. Однако эффективность параллелизации вычислений существенно зависит от многих факторов, один из важнейших – это особенности пересылки данных между соседними узлами многопроцессорной системы, которая обычно является самой медленной частью алгоритма и может свести на нет эффект от увеличения числа используемых процессоров. Указанные вопросы являются определяющими для процедуры моделирования широкого класса задач при помощи модульных многопроцессорных кластерных систем и получили свое развитие в [3].

Важным фактором эффективности функционирования модульных многопроцессорных систем являются вопросы выбора, конструирования и организации их сетевых интерфейсов. Однако в настоящее время не получили должного развития или вообще отсутствуют работы, посвященные исследованию влияния сетевого интерфейса на эффективность модульных многопроцессорных вычислительных систем. Из этих позиций и будет оцениваться эффективность многопроцессорной кластерной системы, которая исследуется в данной работе.

Постановка проблемы исследований, ее актуальность

Одна из основных проблем применения модульных многопроцессорных кластерных систем может быть сформулирована таким образом: имеем разностную сетку размерности M ,

время вычисления задачи, которая решается с использованием однопроцессорной системы, определяется величиной t . Этот параметр является критичным. Необходимо существенно уменьшить время вычислений, сохраняя значение M . В этой связи, вопросам эффективности, быстродействия и производительности уделяется основное внимание при конструировании кластерных систем. Такой подход ориентирован, например, на разработку новых технологических процессов (когда время вычислений является критической величиной) [4,5]. Кроме того, аналогичные задачи часто приходится решать в медицине, военной технике и др. При этом, эффективность распараллеливания вычислений зависит от многих факторов, однако одним из определяющих является организация сетевого интерфейса. В настоящее время такая проблема не получила должного развития для модульных многопроцессорных кластерных систем. В этой связи, рассматриваемые в данной статье исследования, являются актуальными и, несомненно, вызовут интерес у соответствующих специалистов.

Анализ последних исследований и публикаций

Итак, рассматривается задача уменьшения времени расчетов путем увеличения числа узлов кластерной системы. При этом будем считать, что область вычислений равномерно распределяется между узлами кластерной системы. Для удобства исследований будем предполагать, что форма области, на которой проводятся вычисления, является квадратной.

Достаточно полный и всесторонний анализ оценок эффективности модульной многопроцессорной кластерной системы приведен в [3, 6].

С целью исследования путей ускорения вычислений отдельно исследовалось влияние способа передачи данных между узлами кластерной системы. В этой связи исследовалось влияние дуплексного и полудуплексного режима передачи данных в вычислительной сети. Заметим, что дуплексный режим – наиболее скоростной в работе вычислительных систем. Он позволяет эффективно использовать вычислительные возможности кластерных комплексов в сочетании с высокой скоростью передачи данных каналами связи. Учитывая описанные выше обстоятельства, в указанных работах подчеркивается важность исследования характеристик эффективности многопроцессорной системы при реализации дуплексного режима работы. Исходные данные для этого режима исследований представлены в табл. 1.

Здесь приняты следующие обозначения: V_p – протокольная пропускная способность сети

кластера, Гбит/с; T_{it} – время счета одной итерации относительно области вычислений, с; R – объем оперативной памяти узла кластера, Гбит; значение m может равняться единице для одностороннего режима граничного обмена данными, или двум для двустороннего; d – полудуплексный ($d = 1$), или дуплексный ($d = 2$) режим работы вычислительной сети кластерной системы.

Полученные результаты моделирования сведены в табл. 2.

Таблица 1. Исходные данные для расчета характеристик эффективности многопроцессорной системы при реализации дуплексного режима работы кластера

V_p	1 Гбит/с
T_{it}	100 с
R	8 Гбит
M	2
D	2

Таблица 2. Результаты расчета основных характеристик эффективности многопроцессорной системы при реализации дуплексного режима работы кластера

колич. узлов, N	T_n	T_{ex}	T	USK	EF
1	100,00	0,00	100,00	1,00	1,00
2	50,00	2,83	52,83	1,89	0,95
3	33,33	5,66	38,99	2,56	0,85
4	25,00	8,49	33,49	2,99	0,75
5	20,00	11,31	31,31	3,19	0,64
6	16,67	14,14	30,81	3,25	0,54
7	14,29	16,97	31,26	3,20	0,46
8	12,50	19,80	32,30	3,10	0,39
9	11,11	22,63	33,74	2,96	0,33
10	10,00	25,46	35,46	2,82	0,28
11	9,09	28,28	37,38	2,68	0,24
12	8,33	31,11	39,45	2,54	0,21
13	7,69	33,94	41,63	2,40	0,18
14	7,14	36,77	43,91	2,28	0,16
15	6,67	39,60	46,26	2,16	0,14

Результаты моделирования представлены также в виде графических зависимостей (рис. 1, 2).

Проведенный анализ полученных результатов моделирования показал следующее. Как видим, на рис.1 время счета одной итерации при увеличении числа узлов многопроцессорной системы уменьшается по гиперболической зависимости (кривая 1). Наряду с этим время

граничного обмена при увеличении числа узлов многопроцессорной системы увеличивается по линейному закону (линия 2). Общую картину изменения времени счета одной итерации в многопроцессорной системе иллюстрирует зависимость, отображенная кривой 3. Анализ кривой показывает, что время расчета на первом этапе уменьшается при увеличении количества узлов кластера. Подобный результат, казалось, и был предусмотрен. Однако уменьшение такого времени происходит до определенного предела. Если, например, количество узлов будет, превышает шести (рис. 1), то общее время счета начинает возрастать. Происходит это на фоне увеличения объема данных, которые пересылаются между узлами кластера.

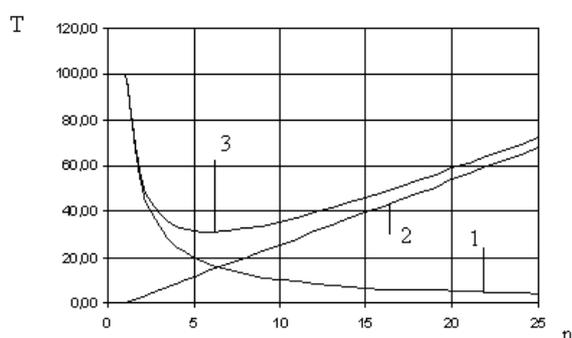


Рисунок 1 – Кривые зависимости времени расчета одной итерации от количества узлов многопроцессорной системы при дуплексном режиме работы

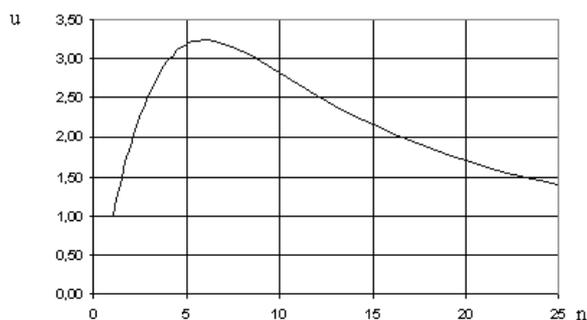


Рисунок 2 – Кривые зависимости ускорения вычислений от количества узлов многопроцессорной системы при дуплексном режиме работы

Итак, проведенный обзор показывает, что, с одной стороны, вопросам исследования эффективности многопроцессорных кластерных систем уделяется достаточно серьезное внимание. С другой стороны, такие исследования охватывают самые разнообразные режимы функционирования кластерных многопроцессорных систем. Однако можно отметить, что на сегодняшний день проблема

влияния сетевого интерфейса кластерной системы на ее эффективность раскрыта не полностью. Более того, также не получило развитие направление исследований, связанное с совершенствованием архитектуры сетевого интерфейса с целью повышения скорости обмена данными между вычислительными узлами многопроцессорной кластерной системы. Проведенные в данной работе исследования направлены на устранение указанного недостатка при анализе эффективности многопроцессорных кластерных систем и в дальнейшем развивают подход к исследованию таких систем, освещенный авторами в [1– 3, 6].

Цель и задачи исследований

Основная цель исследований, представленных в данной работе, заключается в дальнейшем развитии подхода, связанного с анализом оценок эффективности многопроцессорной кластерной модульной вычислительной системы. При этом основное внимание уделяется особенностям влияния сетевого интерфейса такой системы на оценки ее эффективности.

Предполагается решение следующих задач:

- выявить и установить основные режимы работы сетевого интерфейса в многопроцессорных кластерных системах;
- провести анализ основных режимов работы сетевого интерфейса в многопроцессорных кластерных системах и выявить их влияние на оценки эффективности распараллеливания;
- выявить пути повышения эффективности многопроцессорной кластерной системы за счет организации архитектуры ее сетевого интерфейса;
- вывести аналитические соотношения для определения оптимального числа узлов кластерной системы для различных режимов ее функционирования;
- для удобства определения оценок эффективности кластерной вычислительной многопроцессорной системы вывести основные аналитические соотношения через параметры исследуемой системы.

Изложение основного материала исследований

Рассмотрим особенности формирования архитектуры сетевого интерфейса кластерной системы и основные режимы его работы, а затем уже, на втором этапе, проведем анализ взаимодействия процессоров кластерной системы с его интерфейсом.

Итак, для оценки процессов, протекающих в кластерной системе при

организации соответствующих потоков информации, необходимо сравнивать пропускную способность сети кластера и пропускную способность коммутатора. Эта процедура необходима для оптимального подбора компонентов сетевого интерфейса многопроцессорной системы. В этой связи для удобства исследований введем в рассмотрение такой параметр, как общая пропускная способность сети кластерной системы по спецификации производителя (V_s):

$$V_s = V_p \cdot N. \quad (1)$$

Здесь N – число узлов кластера, а V_p – протокольная пропускная способность сети кластера, Гбит/с.

Отметим, что современные коммутаторы имеют такую характеристику, как "пропускная способность шины". Например, восьми портовый коммутатор *SuperStak II Switch 9000* производителя *3COM LG* имеет пропускную способность 17,2 Гбит/с. Это означает, что каждый из восьми портов может одновременно с другими в дуплексном режиме передавать и принимать данные с суммарной скоростью 2 Гбит/с. Точнее говоря, пропускная способность шины задана в пакетах в секунду и заявленные 17,2 Гбит/с достигаются при пересылке больших пакетов. Будем считать, что рассматриваемые граничные области обладают именно такими характеристиками.

При таком подходе уже можно сравнивать общую пропускную способность сети кластера (V_s) и пропускную способность коммутатора (V_b). Очевидно, что пропускная способность используемого коммутатора (V_b) будет соответствовать 16 Гбит/с. Для дальнейшего анализа сетевого интерфейса кластерной системы введем в рассмотрение коэффициент k_s и будем его трактовать, как коэффициент пропускной способности сети кластера, который определим следующим образом:

$$k_s = \frac{V_s}{V_b}. \quad (2)$$

С учетом формулы (1), получаем:

$$k_s = \frac{V_p \cdot N}{V_b}. \quad (3)$$

В работе [3] проведенные исследования справедливы для так называемого "идеального" кластера, когда $k_s=1$, т.е. вся необходимая информация для пересылок в кластерной системе через коммутатор распределяется без задержек

между необходимыми ее узлами. Такой подход для исследования процессов, протекающих в кластерных системах, оказался полезным для того, чтобы выявить основные режимы работы кластерной системы и их соответственным образом оценить. Однако, на практике "идеального" кластера создать практически невозможно. В этой связи, ранее выявленные особенности функционирования кластерной системы необходимо уточнить с учетом новых открывшихся обстоятельств. Для развития такого подхода введем в рассмотрение понятие коэффициента пропускной способности коммутатора (k_b):

$$k_b = \frac{V_b}{V_p \cdot N}. \quad (4)$$

Исходные данные для изучения области изменения коэффициентов сетевого интерфейса многопроцессорной системы приведены в табл. 3.

Таблица 3. Исходные данные для расчета сетевых характеристик кластерной системы

V_p	1 Гбит/с
V_b	16 Гбит/с
k	1

Полученные результаты моделирования сведены в табл. 4.

Результаты моделирования представлены также в виде графических зависимостей (рис. 3). Для удобства анализа исследуемых коэффициентов, соответствующие графические зависимости представлены в одной системе координат.

Проведем предварительный анализ полученных результатов. Очевидно, что с увеличением числа узлов кластерной системы будет возрастать пропускная способность сети кластера (V_s). Тогда изменение коэффициента пропускной способности сети кластера (k_s , формула 3) будет осуществляться по линейному закону (рис. 3, линия 1). С другой стороны, увеличение объема данных, пересылаемых между узлами кластера, приведет к тому, что коммутатор будет перегружаться и его коэффициент пропускной способности (k_b , формула 4) будет уменьшаться по нелинейному закону (рис. 3, линия 2).

Итак, проведенный предварительный анализ результатов моделирования основных сетевых коэффициентов кластерной системы создал предпосылки для обобщения полученных результатов.

Таблиця 4. Результати расчета основных сетевых коэффициентов кластерной системы

Колич. узлов, N	k_s	k_b	c_k
2,00	0,13	8,00	8,00
3,00	0,19	5,33	5,33
4,00	0,25	4,00	4,00
5,00	0,31	3,20	3,20
6,00	0,38	2,67	2,67
7,00	0,44	2,29	2,29
8,00	0,50	2,00	2,00
9,00	0,56	1,78	1,78
10,00	0,63	1,60	1,60
11,00	0,69	1,45	1,45
12,00	0,75	1,33	1,33
13,00	0,81	1,23	1,23
14,00	0,88	1,14	1,14
15,00	0,94	1,07	1,07
16,00	1,00	1,00	1,00
17,00	1,06	0,94	0,94
18,00	1,13	0,89	0,89
19,00	1,19	0,84	0,84
20,00	1,25	0,80	0,80
21,00	1,31	0,76	0,76
22,00	1,38	0,73	0,73
23,00	1,44	0,70	0,70
24,00	1,50	0,67	0,67
25,00	1,56	0,64	0,64

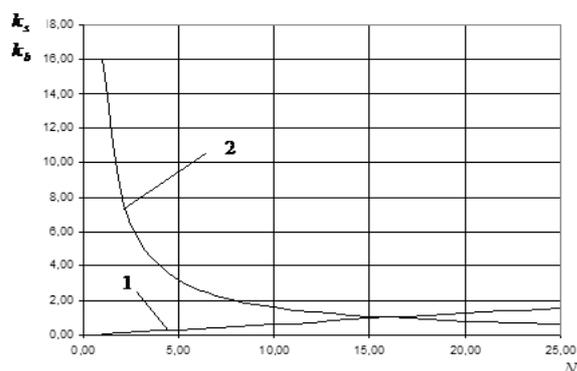


Рисунок 3 – Залежності основних сетевих коефіцієнтів кластерної системи від кількості вузлів

Для повної картини проводимих досліджень, введемо деякі визначення, а потім з урахуванням сформульованих визначень,

проведемо більш детальний аналіз основних сетевих характеристик кластерної системи.

Визначення 1. Точкою *сетевого рівноважності* будемо називати точку, в якій значення коефіцієнта пропускної спроможності мережі кластерної системи і коефіцієнта пропускної спроможності коммутатора будуть рівні між собою.

Визначення 2. *Рівновесним числом вузлів кластерної системи* будемо називати то кількість вузлів, яке відповідає точці *сетевого рівноважності*.

Визначення 3. *Ідеальною багатопроцесорною кластерною системою* будемо називати систему, в якій справедливо рівність виду $k_s = k_b$.

Визначення 4. *Реальною багатопроцесорною кластерною системою* будемо називати систему, в якій справедливо нерівність виду $k_s \neq k_b$.

Визначення 5. *Режимом дефіциту мережевого інтерфейсу кластерної системи* будемо називати варіант функціонування мережі кластера, при якому виконується нерівність виду $k_s < k_b$.

Визначення 6. *Режимом профіциту мережевого інтерфейсу кластерної системи* будемо називати випадок функціонування мережі кластера, при якому виконується нерівність виду $k_s > k_b$.

З урахуванням наведених визначень можна уточнити особливості функціонування мережевого інтерфейсу кластерної системи і виконати детальний аналіз основних розрахункових співвідношень, якими наведені в табл. 3.

В межах аналізу роботи мережевого інтерфейсу кластерної системи на першому етапі відзначимо деякі необхідні особливості функціонування коммутатора. При цьому відзначимо, що продуктивність коммутатора суттєво залежить від типів комутації. Використовуваний коммутатор підтримує чотири типи комутації:

- сквозна комутація (*cut-through*);
- комутація з буферизацією (*store-and-forward switching*);
- безфрагментна комутація (*fragment-free switching*);
- адаптивна комутація (*intelligent*).

При сквозній комутації в буфер вхідного порту поступають лише декілька перших байтів пакета, що необхідно для читання адреси призначення. Після встановлення адреси призначення, паралельно з прийомом решти байтів кадру, відбувається комутація необхідного маршруту, і пакет передається до вихідного порту, якщо він не

используется другими устройствами кластера. В противном случае весь пакет поступает в буфер входного порта. Сквозная коммутация обеспечивает самую высокую скорость коммутации, что дает значительный выигрыш в производительности.

При коммутации с буферизацией пакет поступает в буфер входного порта, где по контрольной сумме проверяется на наличие ошибок. Если ошибки не обнаружены, пакет передается на выходной порт. Этот способ коммутации гарантирует фильтрацию от ошибочных пакетов, однако за счет процедуры буферизации снижается пропускная способность коммутатора по сравнению со сквозной коммутацией.

При бесфрагментной коммутации в буфер входного порта поступает не весь пакет, а только первые 64 байта. Для пакета минимального размера это соответствует полной буферизации, а для пакетов, размер которых больше 64 байт это соответствует сквозной коммутации. Таким образом, при бесфрагментной буферизации проверка подлежат только кадры минимального размера.

При адаптивной коммутации коммутатор сам выбирает для каждого порта оптимальный режим работы.

Предварительно заметим, для получения высоких оценок эффективности и ускорения вычислений кластерной системы необходимо, чтобы она функционировала в одном из указанных типов коммутации в режиме дефицита сетевого интерфейса. Такое утверждение вполне очевидно, т.к. суммарная скорость передаваемых и принимаемых данных всеми узлами многопроцессорной кластерной системы не должна превышать пропускной способности коммутатора.

Основная особенность режима профицита сетевого интерфейса состоит в том, что коммутатор может сталкиваться с перегрузками, когда сумма входящих трафиков превышает сумму выходящих. В таком случае меняются условия коммутации данных. Здесь коммутатор переходит в режим коммутации с буферизацией, что приводит к потере производительности коммутатора. Однако критическое увеличение объема данных в буфере порта приводит к его переполнению и, как следствие, к потере информации. При указанных обстоятельствах коммутатор не сможет обеспечивать максимально стабильное и надежное формирование коммутируемых потоков данных в вычислительной системе. Основные характеристики такой кластерной системы (в т.ч. и быстродействия) будут существенно ухудшаться, а основная задача формирования многопроцессорности вычислений теряет свой смысл.

Вообще заметим, что результаты моделирования основных сетевых характеристик кластерной системы показывают, что режим профицита сетевого интерфейса характерен для малопроизводительных многопроцессорных систем, сетевой интерфейс которых может быть построен на примитивных коммутаторах или при его неудачной организации.

Наконец, проведенные результаты моделирования основных сетевых характеристик кластерной системы убедительно опровергает утверждение "чем больше узлов в кластерной системе, тем быстрее она работает".

Перейдем к более детальному анализу результатов моделирования основных сетевых характеристик кластерной системы с учетом отмеченных обстоятельств. Очевидно, что в рамках рассматриваемых параметров сетевого интерфейса "идеальной" кластерная система будет в случае, когда ее число узлов соответствует $N = 16$ (табл. 4). Однако это противоречит техническим возможностям коммутатора (количество его портов равно восьми). В этой связи, на первый взгляд, может создаваться впечатление такого характера: возможности коммутатора значительно превосходят протокольные характеристики сети. Для более детального анализа такого утверждения вновь рассмотрим некоторые особенности работы коммутатора. Так, если N узлов многопроцессорной системы пытаются установить соединение с одним узлом по протоколу *GigabitEthernet*, то коммутационная шина коммутатора может выделить каждому узлу лишь полосу пропускания (c_k), которая будет определяться соотношением вида:

$$c_k = \frac{V_b}{N}. \quad (5)$$

В табл. 4 приведен расчет полосы пропускания коммутатора (c_k) для такого режима работы сетевого интерфейса. Заметим, что для $N = 8$ значение полосы пропускания на каждый выходной порт кластерной системы будет соответствовать 2 Гбит/с, а это полностью согласуется как с дуплексным режимом обмена данных в многопроцессорной системе, так и с возможностями самой коммутационной шины. При таких обстоятельствах сетевой интерфейс многопроцессорной системы будет функционировать в его режиме дефицита и в условиях максимально допустимой загрузки каналов коммутации коммутатора. Таким образом, технические возможности заявленного коммутатора будут полностью согласовываться с возможностями коммутационной сети.

Анализ выявленных режимов функционирования сетевого интерфейса многопроцессорной системы позволяет сформулировать следующую проблему: каким

образом можно расширить область дефицита сетевого интерфейса многопроцессорных систем? Другими словами эта проблема может быть сформулирована так: как за счет конструктивных особенностей архитектуры вычислительных сетей многопроцессорных кластерных систем можно повысить ее эффективность и быстродействие?

Основная особенность исследуемой многопроцессорной кластерной системы заключается в том, что обмен данных между вычислительными узлами вынесен в отдельную сеть, которая работает на канальном (втором) уровне с использованием технологии *channel bonding* [1]. Такой подход направлен на увеличение скорости обмена данных между узлами кластера, и снижение загрузки канала, который соединяет узлы кластера.

Кроме того, введение дополнительных управляемых коммутаторов, которые работают параллельно, позволило через терминал или WEB-интерфейс изменять конфигурацию сети, повышать ее пропускную способность. Такая архитектура сети обеспечила высокоскоростной доступ к памяти узлов.

Итак, предлагаемая сетевая архитектура многопроцессорной кластерной системы позволяет, во-первых, повышать быстродействие вычислений во время решения сильносвязанных задач и, во-вторых, обеспечивать высокоскоростной доступ к памяти узлов кластера, снижая загрузку канала. В этой связи дальнейшие усилия будут направлены на получение численных оценок сетевого интерфейса кластерной системы за счет реализации технологии *channel bonding* и введением дополнительного управляемого коммутатора.

Для этой цели с учетом применения технологии *channel bonding* необходимо уточнить соотношения (3) и (4), которые характеризуют основные сетевые характеристики кластерной системы. Коэффициент пропускной способности сети кластера будем определять следующим образом:

$$k_s = \frac{V_p \cdot N \cdot k}{V_b \cdot k_m}. \quad (6)$$

Здесь k – количество симметричных вычислительных подсетей, которые работают одновременно за счет реализации технологии *channel bonding*, k_m – количество коммутационных матриц в сети обмена данных.

Коэффициент пропускной способности коммутатора (k_b) уточним аналогично:

$$k_b = \frac{V_b \cdot k_m}{V_p \cdot N \cdot k}. \quad (7)$$

Исходные данные для изучения рассматриваемого режима работы сетевого интерфейса многопроцессорной системы перечислены в табл. 5.

Таблица 5. Исходные данные для расчета сетевых характеристик кластерной системы с использованием технологии *channel bonding*

V_p	1 Гбит/с
V_b	16 Гбит/с
k	2
k_m	2

Полученные результаты моделирования сведены в табл. 6.

Таблица 6. Результаты расчета основных сетевых коэффициентов кластера с использованием технологии *channel bonding*

Колич. узлов, N	k_s	k_b	c_k
2,00	0,13	8,00	16,00
3,00	0,19	5,33	10,67
4,00	0,25	4,00	8,00
5,00	0,31	3,20	6,40
6,00	0,38	2,67	5,33
7,00	0,44	2,29	4,57
8,00	0,50	2,00	4,00
9,00	0,56	1,78	3,56
10,00	0,63	1,60	3,20
11,00	0,69	1,45	2,91
12,00	0,75	1,33	2,67
13,00	0,81	1,23	2,46
14,00	0,88	1,14	2,29
15,00	0,94	1,07	2,13
16,00	1,00	1,00	2,00
17,00	1,06	0,94	1,88
18,00	1,13	0,89	1,78
19,00	1,19	0,84	1,68
20,00	1,25	0,80	1,60
21,00	1,31	0,76	1,52
22,00	1,38	0,73	1,45
23,00	1,44	0,70	1,39
24,00	1,50	0,67	1,33
25,00	1,56	0,64	1,28

Очевидно, что представленный режим работы, при равных прочих условиях, за счет изменения архитектуры сетевого интерфейса многопроцессорной системы будет расширять только полосу пропускания коммутационной шины, оставляя равновесное число узлов кластера

неизменным. Последнее обстоятельство означает, что сформированный режим работы сетевого интерфейса кластерной системы будет предоставлять более широкие возможности для выполнения необходимых расчетов, существенно улучшая характеристики эффективности и быстродействия.

В дальнейшем исследования будут направлены на выявление аналитических зависимостей и числовых характеристик эффективности и ускорения вычислений кластерной системы за счет расширения возможностей сетевого интерфейса. Для этой цели рассмотрим особенности взаимодействия сетевого интерфейса кластерной системы с ее узлами.

Введем характеристику – коэффициент пропускной способности кластерной системы (k_k):

$$k_k = \begin{cases} 1, & \text{для режима дефицита сетевого интерфейса и} \\ & \text{для идеального кластера,} \\ k_b, & \text{для режима профицита сетевого интерфейса} \end{cases} \quad (8)$$

Вполне очевидно, что для режима дефицита сетевого интерфейса коэффициент пропускной способности кластерной системы будет определяться характеристиками сетевого интерфейса. При этом в силу особенностей режима дефицита сетевого интерфейса такой коэффициент принимают равным единице, т.к. коммутационная матрица будет работать в режиме сквозной коммутации (*cut-through*) и информация будет передаться без использования процедуры буферизации передаваемых пакетов с наибольшей скоростью. В режиме профицита сетевого интерфейса такой параметр будет определяться характеристиками коммутатора, когда сумма входящих трафиков превышает сумму выходящих. Здесь коммутатор переходит в режим коммутации с буферизацией, что приводит к потере его производительности, это обстоятельство и отражено спадающей кривой (рис. 4).

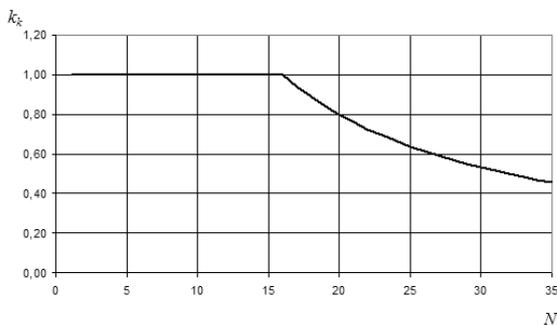


Рисунок 4 – Зависимость коэффициента пропускной способности кластерной системы от числа узлов

Принимая во внимание, что скорость порта узла кластерной системы равна протокольной пропускной скорости сети, общая оценка пропускной способности сети кластера будет определяться следующим соотношением:

$$V = k \cdot d \cdot V_p \cdot k_k \quad (9)$$

Очевидно, что с учетом выражения (8) формула будет распадаться на две: одна описывает общую оценку пропускной способности сети кластера для режимы дефицита сетевого интерфейса (V_1), а другая – режима его профицита (V_2). Для режима дефицита сетевого интерфейса получим:

$$V_1 = k \cdot d \cdot V_p, \quad (10)$$

для режима профицита сетевого интерфейса такая скорость будет выражаться формулой вида:

$$V_2 = \frac{V_b \cdot k_m \cdot d}{N} \quad (11)$$

Анализ соотношения (10) показывает, что для режима дефицита сетевого интерфейса общая оценка пропускной способности сети кластера будет зависеть от скорости порта узла многопроцессорной сети, количества подсетей и режима передачи данных в вычислительной сети (дуплекс или полудуплекс). В тоже время такая скорость не будет зависеть от числа узлов многопроцессорной системы. Этот на первый взгляд парадоксальный вывод можно объяснить тем, что система работает в режиме дефицита сетевого интерфейса, а это означает, что до равновесного числа узлов кластерной системы скорость коммутации данных в сети будет определяться, в основном, скоростью узла порта многопроцессорной системы.

С другой стороны, в режиме профицита сетевого интерфейса оценка пропускной способности сети кластера будет зависеть от пропускной способности используемого коммутатора, количества коммутационных матриц в сети обмена данных, режима передачи данных в вычислительной сети (дуплекс или полудуплекс) и количества вычислительных узлов кластерной системы. В тоже время такая скорость не будет зависеть от количества подсетей сетевого интерфейса. Такой результат очевиден, поскольку коммутационная шина коммутатора может выделить каждому узлу многопроцессорной системы лишь полосу пропускания, которая определяется пропускной способности коммутаторов и числом коммутационных матриц.

В таком случае соотношение для определения времени граничного обмена данных между узлами кластерной системы будет распадаться на два:

$$T_{ex1} = \frac{E}{V_1} \quad (12)$$

для работы кластерной системе в режиме дефицита сетевого интерфейса и

$$T_{ex2} = \frac{E}{V_2} \quad (13)$$

для работы кластерной системы в режиме профицита сетевого интерфейса. В выражениях (12, 13) E – объем данных в области граничного обмена (Гбит).

Выразим соотношения (12) и (13) через параметры кластерной системы:

$$T_{ex1} = \frac{m \cdot (N-1) \cdot \sqrt{R}}{k \cdot d \cdot V_p} \quad (14)$$

$$T_{ex2} = \frac{m \cdot (N-1) \cdot \sqrt{R} \cdot N}{k_m \cdot d \cdot V_b} \quad (15)$$

Оценим количество узлов кластерной системы, при котором задача будет решаться наиболее эффективно. При этом заметим, что время счета одной итерации вычислительного процесса складывается из двух слагаемых – это времени непосредственного счета на процессорах

$T_{calc} = \frac{T_{it}}{N}$ и времени обмена данных между

вычислительными узлами кластера T_{ex} , т.е.

$$T_{it} = T_{calc} + T_{ex} \quad (16)$$

При этом в [3] показано, что скорость вычислений будет расти примерно до момента, когда

$$T_{calc} \approx T_{ex} \quad (17)$$

Таким образом, исходя из соотношения (17) предоставляется возможность оценить количество узлов кластерной вычислительной системы, при котором задача будет решаться наиболее эффективно. Учитывая соотношение (17), получим аналитические выражения для определения оптимального числа узлов кластерной системы:

$$\frac{T_{it}}{N} \approx \frac{m \cdot (N-1) \cdot \sqrt{R}}{k \cdot d \cdot V_p} \quad (18)$$

для работы кластера в режиме дефицита сетевого интерфейса и

$$\frac{T_{it}}{N} \approx \frac{m \cdot (N-1) \cdot \sqrt{R} \cdot N}{k_m \cdot d \cdot V_b} \quad (19)$$

для работы кластера в режиме профицита сетевого интерфейса. На основании выражений (18) и (19) можно получить два уравнения относительно N для определения оптимального числа узлов кластерной системы, при котором

общее время вычислений, требуемое для решения задачи, будет минимальным.

Уравнение (18) преобразуется к квадратичному виду

$$N^2 - N - \frac{T_{it} \cdot k \cdot d \cdot V_p}{m \cdot \sqrt{R}} = 0 \quad (20)$$

Решением такого уравнения будет два корня, при этом один из них положительный, а другой – отрицательный. Исходя из поставленных физических условий задачи, принимается положительный корень, значение которого равно девяти ($N = 9$). Заметим, что такое решение удовлетворяет неравенству из определения 5, которое устанавливает условия функционирования кластерной системы в режиме дефицита сетевого интерфейса.

Уравнение (19) можно свести к кубическому виду

$$N^3 - N^2 - \frac{T_{it} \cdot k_m \cdot d \cdot V_b}{m \cdot \sqrt{R}} = 0 \quad (21)$$

и оно будет иметь два мнимых корня и один действительный. Действительный корень соответствует $N = 11$. Однако анализ такого корня показывает, что он не удовлетворяет условию функционирования кластерной системы в режиме профицита сетевого интерфейса (определение 6).

Отмеченные обстоятельства показывают, что в рамках рассматриваемой задачи оптимальное число узлов кластерной, при котором будет достигаться максимальная эффективность распараллеливания, будет соответствовать $N = 9$. При таком размере кластера задача в 4,6 раза будет решаться быстрее, чем на одном компьютере.

В качестве проверки полученных результатов была проведена процедура расчета основных характеристик эффективности многопроцессорной кластерной системы. Полученные результаты моделирования сведены в табл. 7, а график ускорения вычислений при указанных условиях эксперимента приведен на рис. 5.

Итак, теоретические выкладки полностью согласуются с результатами численного моделирования основных характеристик эффективности кластерной многопроцессорной системы.

Сравнительный анализ результатов расчета без реорганизации архитектуры сетевого интерфейса (табл. 2) и после введения симметричных вычислительных подсетей, которые работают одновременно в результате реализации технологии *channel bonding* (табл. 7) показал, что за счет повышения скорости обмена данных между узлами вычислительной системы

удалось снизить загрузку каналов, которые соединяют эти узлы. Такой подход позволит существенно повысить эффективность кластера.

Таблица 7. Результаты расчета основных характеристик многопроцессорной кластерной системы

колич. узлов, N	T_n	T_{ex}	T	USK	EF
1	100,00	0,00	100,00	1,00	1,00
2	50,00	1,41	51,41	1,94	0,97
3	33,33	2,83	36,16	2,77	0,92
4	25,00	4,24	29,24	3,42	0,85
5	20,00	5,66	25,66	3,90	0,78
6	16,67	7,07	23,74	4,21	0,70
7	14,29	8,49	22,77	4,39	0,63
8	12,50	9,90	22,40	4,46	0,56
9	11,11	11,31	22,42	4,46	0,50
10	10,00	12,73	22,73	4,40	0,44
11	9,09	14,14	23,23	4,30	0,39
12	8,33	15,56	23,89	4,19	0,35
13	7,69	16,97	24,66	4,05	0,31
14	7,14	18,38	25,53	3,92	0,28
15	6,67	19,80	26,47	3,78	0,25
16	6,25	21,21	27,46	3,64	0,23

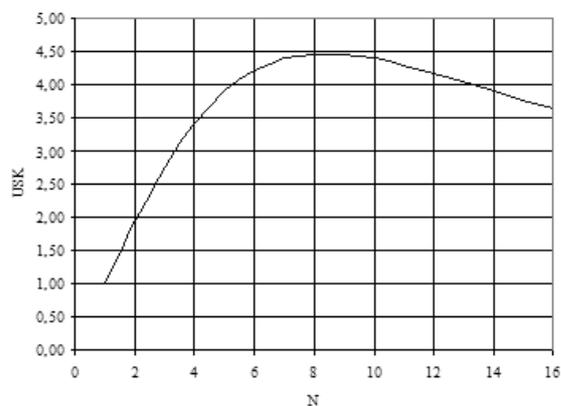


Рисунок 5 – Кривая зависимости ускорения вычислений от количества узлов многопроцессорной системы при двустороннем граничном обмене данными

Выводы и перспективы последующих исследований

В работе показаны пути повышения эффективности многопроцессорной кластерной системы за счет реорганизации архитектуры ее сетевого интерфейса. Предложенный подход позволил не только повысить эффективность распараллеливания, но и существенно уменьшить время вычислений. Таких результатов удалось достичь за счет уменьшения времени граничного обмена данными между вычислительными узлами кластерной системы.

При этом:

1. Выявлены два основных режима работы сетевого интерфейса в модульной многопроцессорной кластерной системе. Показаны условия формирования равновесного числа вычислительных узлов многопроцессорной кластерной системы, когда режим дефицита сетевого интерфейса переходит в режим его профицита.

2. Проведен анализ выявленных режимов работы сетевого интерфейса в модульной многопроцессорной кластерной системе. Показано, что для получения высоких оценок эффективности и ускорения вычислений в кластерной системе необходимо, чтобы она функционировала в режиме дефицита сетевого интерфейса. Выявлено, что основная особенность режима профицита сетевого интерфейса состоит в том, что коммутатор будет сталкиваться с перегрузками, когда сумма входящих трафиков превышает сумму выходящих. При этом основные характеристики эффективности такой кластерной системы будут существенно ухудшаться.

3. Определены условия работы коммутационной матрицы коммутатора в режиме сквозной коммутации (*cut-through*) с тем, чтобы информация передавалась без использования процедуры буферизации. Такой подход обеспечивает передачу пакетов с наибольшей скоростью, что приводит к улучшению оценок эффективности кластерной системы в целом.

4. Установлены оценки эффективности кластерной системы при организации многоканальных режимов функционирования сетевого интерфейса многопроцессорной кластерной системы.

5. Выведены аналитические соотношения для определения оптимального числа узлов многопроцессорной кластерной системы, когда соответствующая задача будет решаться за минимальное возможное время.

6. Выведены основные аналитические соотношения для определения числовых характеристик эффективности кластерной системы через ее основные параметры.

Список літератури

1. Модуль високоефективної багато процесорної системи підвищеної готовності: пат. 57663 Україна, МПК G06F 15/16 (2011.01) / Іващенко В.П., Башков Є.О., Швачич Г.Г., Ткач М.О.; власники: Національна металургійна академія України, Донецький національний технічний університет. – № u 2010 09341; заявл. 26.07.2010; опубл. 10.03.2011, Бюл. № 5.
4. Швачич Г.Г. Математическое моделирование скоростных режимов термической обработки длинномерных изделий / Г.Г. Швачич, В.П. Колпак, М.А. Соболенко // Теория и практика металлургии. Общегосударственный научно-технический журнал. – 2007. – № 4 - 5 (59 - 60). – С. 61 - 67.
5. Швачич Г.Г. Про проблему математичного моделювання термічної обробки довгомірного сталевого виробу / Г.Г. Швачич, М.О. Ткач // VII International Conference ["Strategy of Quality in Industry and Education"], 3 - 10 June, 2011, Varna; Bulgaria. - Proceedings. – V. 2. – P. 561 - 567.
6. Сбитнев Ю.І. Дослідження оцінки ефективності багато процесорної кластерної системи / Ю.І. Сбитнев, Г.Г. Швачич, М.О. Ткач // VI International Conference ["Strategy of Quality in Industry and Education"], 1 - 8 June 2010, Varna; Bulgaria - Proceedings. – V. 2. – P. 288 – 296.

Надійшла до редколегії 15.10.2011

**Є.О. БАШКОВ, В.П. ІВАЩЕНКО,
Г.Г. ШВАЧИЧ, М.О. ТКАЧ**

Донецький національний технічний університет
Національна металургійна академія України

**E.A. BASHKOV, V.P. IVASHENKO,
G.G. SHVACHYCH, M. A. TKACH**

Donetsk national technical university
National metallurgical academy of Ukraine

Дослідження впливу мережевого інтерфейсу на ефективність модульної багато процесорної кластерної системи

Research of the influence of network interface on the efficiency of module multiprocessor cluster system

Статтю присвячено проблемі дослідження оцінок ефективності багато процесорної обчислювальної системи. Основна увага приділяється впливу мережевого інтерфейсу на оцінки ефективності кластерної системи. Виведено аналітичні співвідношення основних оцінок ефективності обчислень через параметри багато процесорної системи.

The article is devoted to the problem of estimation efficiency of the multiprocessor computer system research. Basic attention is focused on the questions of the influence of network interface on the estimations of the cluster's system efficiency. Analytical correlations of the basic estimations efficiency are shown via the parameters of the multiprocessor system.

Персональний обчислювальний кластер, керований комутатори, реконфігурована мережа, обчислювальні вузли, прискорення, ефективність

Personal calculable cluster, guided switchboards, reconfigurable network, calculable knots, acceleration, efficiency