электронный научно-технический журнал
ИНЖЕНЕРНЫЙ ВЕСТНИК
Издатель ФГБОУ ВПО "МГТУ им. Н.Э. Баумана". Эл No. ФС77-51036. ISSN 2307-0595

Отказоустойчивые топологии для циклических компьютерных систем

10, октябрь 2014

Андреев А. М., Можаров Г. П.

УДК: 681.322.01

Россия, МГТУ им. Н.Э. Баумана
arkandreev@gmail.com

Введение.

Циклическая конфигурация сети процессоров компьютерных систем (КС) отличается простотой организации и управления. Сообщения между элементарными процессорами (ЭП) передаются по связям в одном направлении, пока не поступят на требуемый процессор КС, при этом отпадает необходимость в принятии решения о направлении пересылки сообщения.

Для сравнительного изучения различных способов организации циклической сети процессоров КС интересны оценки максимальной и средней задержки сообщения, отказоустойчивости, оперативного времени процессоров, а также максимальной пропускной способности. Эти величины взаимозависимы и связаны с топологией КС.

Что касается отказоустойчивости, то циклическая сеть неустойчива к ошибкам, т. к. любой отказ ЭП нарушает связь между процессорами КС. Далее, при увеличении количества ЭП в КС сообщениям приходится проходить больше промежуточных ЭП. Так, диаметр однонаправленного цикла из N ЭП составляет $N-1$, а средняя длина пути сообщения – $N/2$. С другой стороны, при увеличении размера сети КС требуется большая общая пропускная способность, с которой элементы связи могут не справиться.

Для повышения отказоустойчивости можно, в частности, соединить все ЭП с центральным ЭП, способным рассылать сообщения по периферийным ЭП, но при этом нарушается однородность сети и повышается её зависимость от отказоустойчивости центрального ЭП. Другой подход к проблеме предусматривает введение дополнительных связей между ЭП. Помимо повышения отказоустойчивости это уменьшает время передачи сообщений и разгружает каналы связи, повышая пропускную способность КС.

Топология циклической сети

Рассмотрим топологии двойного цикла, для которых исследуются проблемы отказоустойчивости функционирования и к отказам КС. В рассматриваемых КС N узлов (ЭП) соединены в основной однонаправленный цикл и, кроме того, каждый узел имеет дополнительную связь для улучшения функционирования и повышения отказоустойчивости. Пример такой сети – двунаправленный полный цикл, выдерживающий одиночные сбои. Аппаратный интерфейс цикла может быть выполнен на базе микропроцессоров и обеспечивать простой алгоритм обхода отказавшего ЭП или связи. Диаметр такой сети составляет $N/2$, а среднее время передачи сообщения – $N/4$.

Другим примером подобной топологии является цикл, в котором узел x соединяется с узлами $x+1$ и $x-s(\bmod N)$, т. е. имеет переход на один узел вперёд и на s узлов назад (рис. 1). Такая сеть имеет большую пропускную способность и отказоустойчивость и меньшее время передачи сообщений, чем предыдущая. Для большинства значений N оптимальная (или близкая к оптимальной) в смысле функционирования и отказоустойчивости сеть достигается при $s=\lfloor\sqrt{N}\rfloor$. В оптимальной сети между каждой парой узлов существует несколько путей, поэтому КС может работать при выходе из строя нескольких элементов.

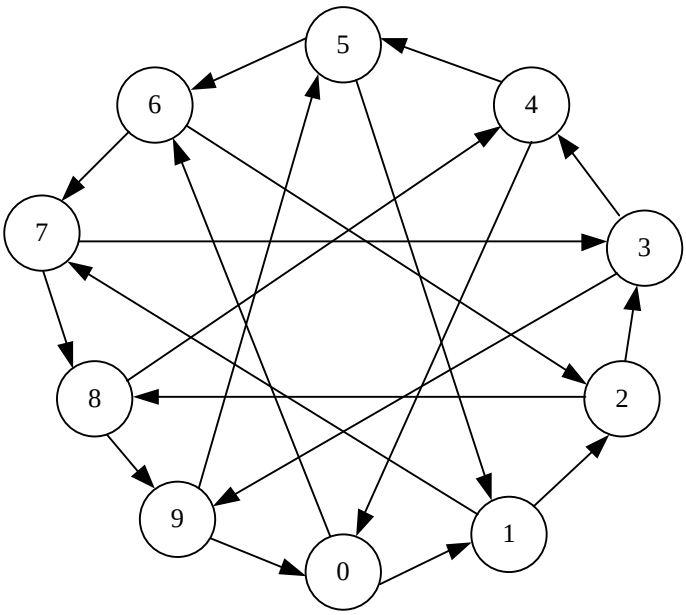

Рис. 1. Десяти узловая циклическая сеть, имеющая переходы на один узел вперёд и на четыре узла назад

Организация циклического интерфейса показана на рис. 2. В интерфейс поступают сообщения от его хозяина (1) и двух соседей. Для большей надежности интерфейс разделён на два одинаковых модуля, содержащих блоки управления (2) и драйверы (3), с выходами на общие передатчик (4) и приёмник (5). Через драйверы осуществляется передача сообщений узлу $i+1$ (6) и $i-s$ (7), а также приём сообщений от узлов $i-1$ (8) и $i+s$ (9).

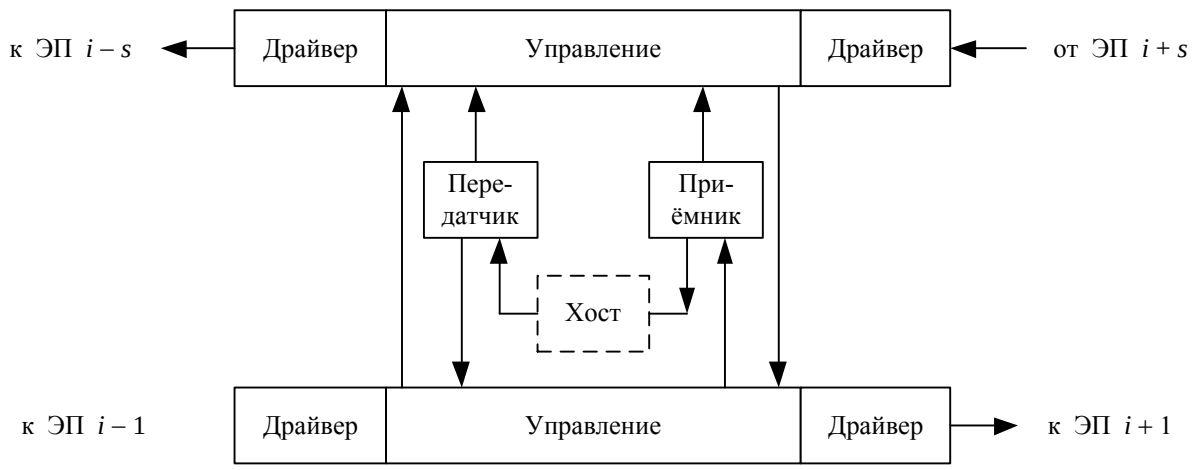

Рис. 2. Функциональная организация циклического интерфейса

Сообщение, помимо данных и служебной информации, содержит адреса источника и процессора назначения. Если адрес назначения совпадает с адресом интерфейса, то сообщение изымается из сети и передаётся хозяину, иначе посылается дальше по одной из исходящих связей. Может быть предложена схема, в которой процессор назначения лишь копирует посланное ей сообщение; аннулировать его из сети может только источник. Этим обеспечивается автоматическое уведомление о получении сообщения. Для устранения конфликтных ситуаций можно потребовать, чтобы процессор не посылал очередного сообщения до получения уведомления о предыдущем. Уведомление позволяет также убедиться, что в результате сбоев один или несколько процессоров не оказались изолированными. (В дальнейшем предполагается, что сообщение, дойдя до адресата, изымается из сети).

Диаметр сети и производительность

Для каждого узла существует диаметрально противоположный узел, отстоящий на расстоянии d диаметра сети. Параметр s выбирается таким образом, чтобы минимизировать

$$d = \max d_{ij} \text{ для всех } i,\ i = 0,\ 1,\ ...,\ N-1,$$

где d_{ij} – длина кратчайшего пути между узлами i и j.

Для получения кратчайшего пути между узлами i и j выполняются переходы назад до тех пор, пока это выгодно. Для диаметрально противоположных узлов переходы назад выгодны, пока d больше, чем $(N - bs)$, где b – количество шагов назад. Дополнительно к этому для диаметрально противоположных узлов требуется $s-1$ переход вперёд. Таким образом, соотношение между d и s имеет вид: $d = N - (d - s + 1)s + (s - 1)$, откуда

$d = \left\lfloor \frac{N}{s+1} \right\rfloor + (s-1)$. Для минимизации d при заданном N подходит, в частности, значение $s = \left\lfloor \sqrt{N} \right\rfloor$.

Средняя длина пути $\bar{\rho}$, который проходит сообщение, может быть определена из уравнения $\bar{\rho} = \frac{1}{N}\sum_{i=0}^{N-1}\left[\frac{1}{N-1}\sum_{j=0}^{N-1} d_{ij}\right]$, где $d_{ij} = 0$. Если КС симметрична, то выражение в квадратных скобках одинаково для всех i, следовательно, средняя длина пути прохождения сообщения составляет $\bar{\rho} = \frac{1}{N-1}\sum_{j=0}^{N-1} d_{ij}$ при всех i, а максимальное количество пересылок назад – $\rho_{b_{\max}} = \left\lfloor \frac{N}{s+1} \right\rfloor$.

Если N – полный квадрат, то $\rho_{b_{\max}} = s-1$. Отсюда можно посчитать сумму расстояний от заданного узла до других, составляющую

$$\rho_{\Sigma} = \frac{\left(N - \rho_{b_{\max}} s\right)\left(N - \rho_{b_{\max}} s - 1\right)}{2} + \frac{\rho_{b_{\max}} s}{2}\left(\rho_{b_{\max}} + s\right).$$

Среднее число пересылок $\Pi_{\text{ср}}$ будет определяться выражением $\Pi_{\text{ср}} = \rho_{\Sigma} / (N-1)$. Подставляя вместо $\rho_{b_{\max}}$ и ρ_{Σ} их выражения, можно получить

$$\Pi_{\text{ср}} = \frac{1}{2(N-1)}\left[\left(N - \left\lfloor \frac{N}{s+1} \right\rfloor s\right)\cdot\left(N - \left\lfloor \frac{N}{s+1} \right\rfloor s - 1\right) + \left\lfloor \frac{N}{s+1} \right\rfloor s\left(\left\lfloor \frac{N}{s+1} \right\rfloor + s\right)\right].$$

Чтобы минимизировать среднее число пересыпок, требуется минимизировать выражение во вторых квадратных скобках. Это достигается при $s+1 = N/(s+1)$, откуда $s = \left\lfloor \sqrt{N} \right\rfloor$.

В таблице для различных значений N приведены максимальные (левые колонки) и средние (правые колонки) количества пересылок сообщения для двунаправленного цикла $(s=1)$, для сети при $s=2$ и для КС при $s = \left\lfloor \sqrt{N} \right\rfloor$.

Таблица Максимальное и среднее расстояния пересылок (скачков) для трёх циклических архитектур КС

N	**Двойное кольцо**		**Гирляндное кольцо**		**Оптимальное кольцо**	
	макс.	средн.	макс.	средн.	макс.	средн.
10	5	2,80	4	2,33	4	2,33
15	7	4,00	6	3,21	5	3,00
30	15	7,76	11	5,69	9	4,65
50	25	12,75	17	9,00	12	6,14

Для устранения конфликтных ситуаций, связанных с одновременным поступлением в интерфейс нескольких сообщений от разных источников, каждый интерфейс должен содержать буферный регистр поступающих сообщений. Среднее время задержки сообщения в этом регистре примерно пропорционально расстоянию между источником и приёмником. При отсутствии конфликтов среднее время задержки $\overline{T}$ характеризуется выражением $\overline{T} = \rho_{\Sigma} + \overline{\rho} \times T_{буф}$, где ρ_{Σ} – задержка передачи, $T_{буф}$ – задержка в буфере на запоминание и проверку шапки сообщения.

Если сообщение буферизуется в каждом промежуточном узле, то средняя задержка $\overline{T'}$ составляет $\overline{T'} = \overline{\rho} \times \left(\rho_{\Sigma} + T_{буф}\right)$.

Дать точную оценку задержки в промежуточных узлах трудно, однако видно, что время задержки пропорционально количеству промежуточных пересылок, т. к. среднее количество конфликтующих сообщений пропорционально количеству процессоров, через которые они проходят.

Для изучения задержек в сетях двойного цикла был разработан имитатор дискретных событий. Возникающие сообщения составляет пуассоновский поток, а длины сообщений, времена задержки и обработки заголовков сообщений распределены экспоненциально. Для каждого звена (связи между соседними узлами) создаётся своя очередь с достаточно большим буфером во избежание переполнения. Сравнительное моделирование для простого двунаправленного цикла и оптимального двойного цикла показывает, что средние времена задержек в сети оптимального цикла значительно уменьшены в сравнении с двунаправленным для больших нагрузок и слабо отличаются при малых.

Для анализа скорости пересылки сообщений сделаны следующие предположения: сообщения образуют пуассоновский поток со средней скоростью γ; все звенья равнозначны; система работает на пределе пропускной способности.

Предполагается также, что для передачи сообщений используются кратчайшие маршруты между узлами. Если таких маршрутов несколько, то сообщения передаются поровну по каждому из маршрутов. Среднее число переходов $\Pi_{ср}$ вычисляется как среднее число кратчайших маршрутов от одного узла ко всем другим. Общий объем пересылок характеризуется соотношением $\rho_{f_{max}} + \rho_{b_{max}} = \Pi_{ср} \times \gamma / N$, где $\rho_{b_{max}}$ – вклад в пересылку сообщений переходов назад, а $\rho_{f_{max}}$ – вперёд. Посредством перераспределения сообщений по нескольким кратчайшим маршрутам равной длины возможен баланс передачи сообщений в обоих направлениях, при этом $\rho_{f_{max}} = \rho_{b_{max}} = \frac{1}{2} \cdot \Pi_{ср} \times \gamma / N$. Если приравнять объем пересылок к пропускной способности системы, можно получить: $\left(\rho_{f_{max}} + \rho_{b_{max}}\right) N = 2N\mu$, где μ – пропускная способность звена. Посредством преобразований и подстановки в результирующую формулу выражения для $\Pi_{ср}$, имеем:

$$\gamma_{ST} = \frac{2N(N-1)\mu}{\frac{1}{2}\left[\left(N - \left\lfloor \frac{N}{s+1} \right\rfloor s\left(N - \left\lfloor \frac{N}{s+1} \right\rfloor s - 1\right) + \left\lfloor \frac{N}{s+1} \right\rfloor s\left(\left\lfloor \frac{N}{s+1} \right\rfloor + s\right)\right)\right]},$$

где γ_{ST} максимально, если средняя длина пути сообщения минимальна. Максимум γ_{ST} достигается при $s = \left\lfloor \sqrt{N} \right\rfloor$. Для двунаправленного цикла эта величина составляет $\gamma_{ST} = \frac{2N(N-1)}{\left\lfloor \frac{N^2}{4} \right\rfloor}$, а для двойного цикла при $s = 2$

$$\gamma_{ST} = \frac{4N(N-1)\mu}{\left[\left(N - \left\lfloor \frac{N}{3} \right\rfloor 2\left(N - \left\lfloor \frac{N}{3} \right\rfloor 2 - 1\right) + \left\lfloor \frac{N}{3} \right\rfloor 2\left(\left\lfloor \frac{N}{3} \right\rfloor + 2\right)\right)\right]}/$$

Так как γ_{ST} обратно пропорционально $\Pi_{ср}$ и в рассматриваемой топологии $\Pi_{ср}$ почти минимально, то производительность для данной топологии оптимальна.

В особом случае, если N – полный квадрат, то $d = 2\left(\sqrt{N} - 1\right)$, средняя длина пути сообщения будет равна $\bar{\rho} = \frac{N}{\left(\sqrt{N} + 1\right)}$, откуда $\gamma_{ST} = \frac{2(N-1)}{\left(\sqrt{N} - 1\right)}\mu$.

При выходе элемента сети из строя максимальная и средняя длина пути сообщения увеличивается, однако и в этом случае параметры оптимального цикла лучше параметров других рассматриваемых структур.

В оптимальной сетевой архитектуре используется алгоритм распределённой рассылки, в соответствии с которым сообщение передаётся по назначению, обходя неисправные узлы и связи, на основе лишь локальной информации при выборе маршрута. Локальные соседи – это узлы, связанные друг с другом короткими прямыми связями, а отдалённые соседи – это узлы, связанные длинными обратными связями. Кратчайший путь между источником и приёмником может состоять как из длинных, так и коротких переходов. Локальный путь состоит только из коротких, а путь высокого уровня – только из длинных переходов.

На рис. 3 представлена сеть оптимального цикла. В каждом узле хранится таблица локальных адресов всех узлов, расположенных от данного на расстоянии, меньшем a в обоих направлениях по локальным путям, а также адрес отдалённого соседа.

При отсутствии неисправных узлов узел, получив сообщение, читает адрес приёмника D из заголовка и пытается найти его среди адресов узлов впереди по основному циклу, содержащихся в его таблице. Если выяснилось, что D – собственный адрес узла, то сообщение пришло по назначению. Если D имеется в просмотренной части таблицы локальных адресов, то сообщение направляется по этому адресу через короткие связи. В остальных случаях сообщение передаётся отдалённому соседу.

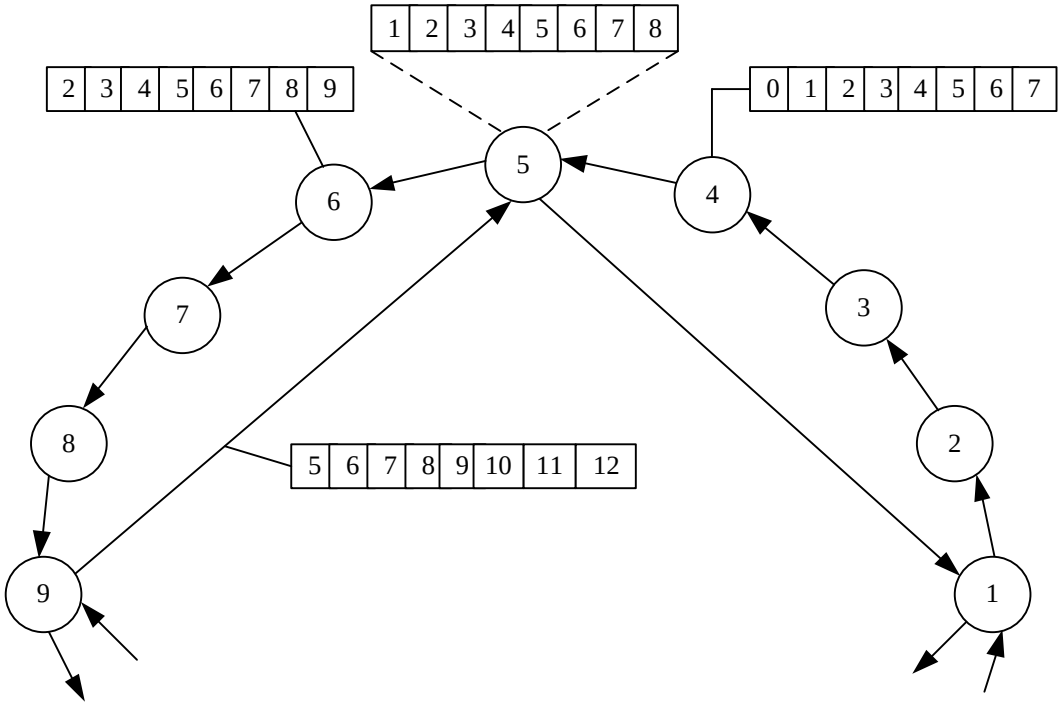

Рис. 3. Маршрутизация в циклической сети, имеющей переходы на один узел вперёд и на четыре узла назад

Таблица локальных адресов составляется и модифицируется при помощи распределённого алгоритма. Периодически, но асинхронно каждый узел передаёт своим, локальному и отдалённому, соседям свою таблицу локальных адресов. На основе таблицы, полученной от соседнего узла, узел модифицирует свою собственную таблицу.

При выходе из строя узла или связи (без ограничения общности – узла) в алгоритме рассылки рассматриваются случаи, когда неисправность возникла на пути высокого уровня или на локальном пути, В первом случае передающий узел, обнаружив неисправность отдалённого соседа (в соответствующей позиции таблицы адресов проставлен признак ∞), пересылает сообщение локальному соседу, который, не найдя адресата в своей таблице, перешлёт сообщение по длинной связи, и неисправный узел будет обойдён. Если же неисправность обнаружена па локальном пути (рис. 4), например, пусть неисправность возникла в узле 2 и обнаружена узлом 5 при просмотре таблицы узла 4, то узел 5 попытается найти адрес сообщения во всей своей таблице (а не только среди адресов, куда узел 5 может передать сообщение по локальным связям). Если искомый адрес обнаружен, то сообщение пересылается по локальной связи, иначе сообщение пересылается по длинной связи. Например (см. рис. 4), узел 5, получив сообщение с адресом 3, находит 3 в своей таблице, пересылает его по локальной связи в узел 6, который доставляет его в узел 3 через узел 7, в соответствии с первым случаем. Сети двойного цикла обладают более высокой надёжностью и устойчивостью к неисправностям КС, чем сети простого цикла, вследствие наличия нескольких путей. Сети двойного цикла обеспечивают несколько альтернативных путей между двумя узлами. Например, для сети из 20 узлов наличие неисправного

процессора влечёт за собой увеличение диаметра от 9 до 18 в двунаправленном цикле, от 7 до 8 в циклах при $s=2$ и $s=\left\lfloor \sqrt{N} \right\rfloor$. Влияние неисправных узлов на среднее число пересылок сообщения в сетях двойных циклов относительно мало.

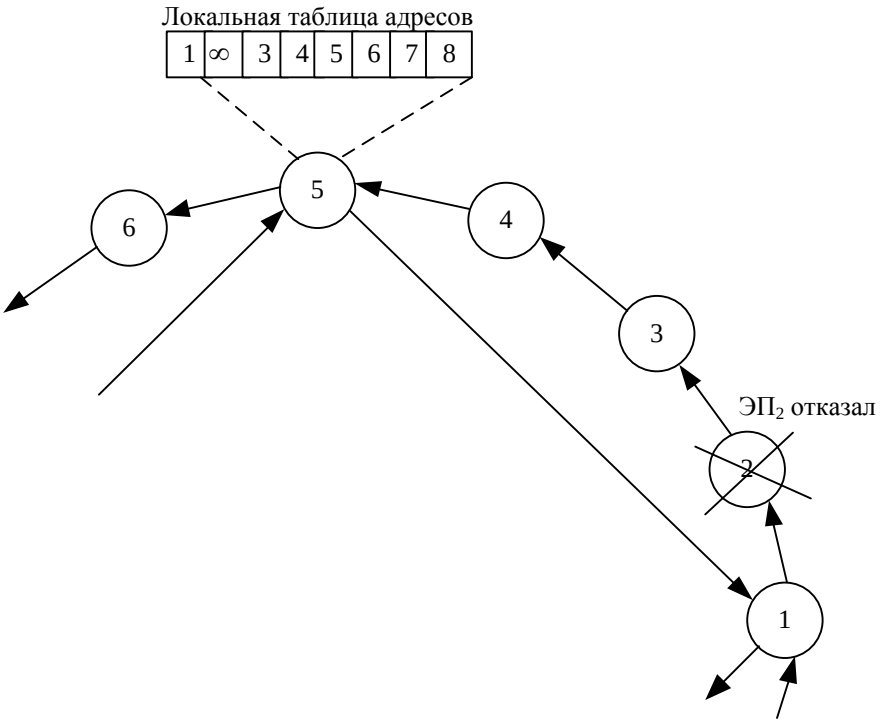

Рис. 4. Индикация относительно отказавшего узла локальной таблице адреса

Надёжность и отказоустойчивость сети

Мерой устойчивости к отказам является надёжность пересылки сообщения между двумя наиболее удалёнными друг от друга (противоположными) узлами. Интуитивно, она пропорциональна количеству различных маршрутов между этими узлами и обратно пропорциональна диаметру сети.

При $s=\left\lfloor \sqrt{N} \right\rfloor$ диаметр минимизируется, а число альтернативных маршрутов максимизируется. Каждый из кратчайших маршрутов состоит из $\rho_{b_{\max}}=\left\lfloor \frac{N}{s+1} \right\rfloor$ длинных пересыпок назад и $s-1$ коротких пересылок вперёд. Число альтернативных маршрутов между противоположными узлами составляет $N_R-1=\binom{\rho_{b_{\max}}+s-1}{s-1}$, откуда $N_R=1+\binom{\rho_{b_{\max}}+s-1}{s-1}$.

При $s=2$ получаем $N_R = 2+\lfloor N/3 \rfloor$, а при $s=\lfloor \sqrt{N} \rfloor$ – $N_R = 1+\binom{\left\lfloor \frac{N}{s+1} \right\rfloor + s-1}{s-1}$.

Максимум N_R достигается при значениях $s=\lfloor \sqrt{N} \rfloor$.

В сети с двумя неисправными узлами особый случай составляют изолированные узлы, когда оба неисправных узла являются соседями такого узла, выключая его из сети. Вероятность появления отрезанного узла (если известно, что два узла неисправны) составляет

$$N_{DN} = \frac{2N}{N(N-1)} = \frac{2}{N-1}$$

и уменьшается при увеличении размера сети. В случае отрезанного узла в сети оптимального двойного цикла выключается $1/N$ всех возможных маршрутов в сети, в то время как в неоптимальных сетях двойного никла при наличии двух неисправных узлов перестаёт действовать до половины всех возможных связей, откуда следует, что вероятность разрыва каждой связи из-за двух неисправных узлов в оптимальной циклической сети на два порядка ниже, чем в неоптимальной.

Выводы

Таким образом топология двойного цикла для локально распределённых сетей обеспечивает оптимальную производительность за счёт максимального количества транзитных участков, среднего числа транзитных участков, уровня пропускной способности и отказоустойчивости. Лучшая производительность достигается в предложенных топологиях: двойной и гирляндный циклы. Топология схемы межсоединений процессоров может найти приложения в многопроцессорных КС. Кроме того достаточно просто выполняется реконфигурирование КС в случае отказов.

Список литературы

1. Андреев А.М., Можаров Г.П., Сюзев В.В. Многопроцессорные вычислительные системы: теоретический анализ, математические модели и применение. - М.: Изд-во МГТУ им. Н. Э. Баумана, 2011. - 334 с.
2. Андреев А.М., Можаров Г.П. Анализ основных параметров компьютерных систем методом спектральной теории графов // Наука и образование: электронное научно-техническое издание, 2011, 10 [электронный ресурс] http://technomag.edu.ru/doc/232774.html (77-30569/232774).
3. Андреев А.М., Березкин Д.В., Можаров Г.П., Свирин Ил.С. Математическое моделирование надежности компьютерных систем и сетей // Вестник

МГТУ им. Баумана. Сер. «Приборостроение». - 2012. Спец. выпуск «Моделирование и идентификация компьютерных систем и сетей». - с. 3-46.

4. Андреев А.М., Можаров Г.П. Инженерная методика оценки компьютерных систем // Инженерный журнал: наука и инновации. 2013, вып. 11. URL: http://engjournal.ru/ catalog/it/network/1007.html, с. 1-8.
5. Андреев А.М., Можаров Г.П. Методика выбора программного обеспечения компьютерных сетей // Инженерный журнал: наука и инновации. 2013, вып. 11. URL: http://engjournal.ru/ catalog/it/network/1055.html, с. 1-10.
6. Андреев А.М., Можаров Г.П. Моделирование надежности компьютерной сети // Инженерный журнал: наука и инновации. 2013, вып. 11. URL: http://engjournal.ru/ catalog/it/network/1007.html, с. 1-6.
7. Черкесов Г.Н. Надёжность аппаратно-программных комплексов. Учебное пособие. - СПб.: Питер, 2005. - 479 с.