МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 517.21

А.С.Сорокин

МОДЕЛИ СИММЕТРИЧНЫХ СИСТЕМ ОПРОСА В ТЕОРИИ ОРГАНИЗАЦИИ ОЧЕРЕДЕЙ

§1. Введение. В этой работе представлено исследование моделирования, демонстрирующее использование языка PEPA для оценки характеристик. Примеры, полученные при исследовании моделирования, будут использоваться для того, чтобы показать методы упрощения модели, которые будут указаны позднее. Это исследование рассматривает и сравнивает различные системы: *мульти-очереди и мульти-серверы*. Такие системы являются расширением традиционной системы опроса и обычно применяются к моделям, в которых множественные ресурсы разделены среди нескольких пользователей возможно с различными требованиями. Примеры включают локальные сети с многочисленными маркерами и сетями взаимосвязи мультишин в распределенных системах. Подобные системы были исследованы в [1-16].

Система опроса состоит из нескольких очередей и единственного сервера, который перемещает очереди в циклическом порядке. Эти очереди всегда были хорошими моделями для многих систем, возникающих в компьютерных сетях и в коммуникациях и, следовательно, были хорошо изучены [15] .

Множество расширений и модификаций традиционной системы опроса было исследовано [15], включая нециклический опрос, очереди приоритета и очереди с обратной связью. Одно расширение, которое особенно подходит для моделирования инновационных локальных сетей, это введение дополнительных серверов, каждый из которых перемещает очереди, обслуживающие там, где это необходимо. Эти системы, иногда известные как системы *мульти-очереди и мульти-сервера,* быстро не поддаются решению в теории организации очередей. Несколько предложенных приближенных методов, основанных на теории организации очередей, а также точных решениях, основанных на GSPNs, рассмотрены в [6-9].

Системы мультиочереди и мультисервера были выбраны как основание для представленного исследования моделирования в данной работе, потому что они просто сформулированы и понятны, хотя выбор критериев качества работы есть задача не тривиальная. Тонкость этих систем состоит в зависимостях, которые существуют между очередями. Скопление в каждой очереди зависит от скопления в других очередях в системе, а также и между серверами.

В остальной части работы представлена подготовка мультиочереди, мультисервера и системы опроса. Иллюстрированы некоторые особенности моделей, представленных на языке PEPA. В разделе 2 описаны основные особенности систем опроса и рассмотрены их решения. Как иллюстрация в разделе 2.2 дана простая система опроса модели PEPA вместе с некоторыми численными результатами.

§2. Системы опроса. *Элементы системы опроса* развивались из схем опроса, используемых для передачи данных между терминалами и центральным компьютером, а также с применением многоабонентской линии (Рис. 1). В свою очередь центральный компьютер должен быть приближен к каждому терминалу, для того чтобы установить есть ли у него какие-нибудь данные для передачи. Если это так, то терминал передаёт данные, и компьютер переходит к опросу следующего терминала; в противном случае компьютер немедленно переходит дальше к следующему терминалу. Впоследствии опрос системы использовался для широкого диапазона применений, характерных и основанных на требованиях множественного доступа к разделенным ресурсам. В приведенном выше примере каждый терминал имеет плановую возможность передать данные на центральный компьютер.

Вообще, система опроса состоит из набора *узлов* или *очередей* и единственного *сервера,* который циркулирует между ними в циклическом порядке. Для каждого обслуживания требования *клиентов* узла накапливаются в *буфере*. Сервер посетит каждый узел, в свою очередь, оказывая обслуживание, если буфер не пуст, иначе перемещается прямо в следующий узел. Время, требуемое серверу, для того чтобы переместиться от одного узла к следующему, известно как время *блуждания* или *переключения*.

Важно делать различие между системами опроса и системой синхронного времени мультиплексированного разделения (STDM).

Сервер потратит некоторое количество времени в каждом узле независимо от того, тре

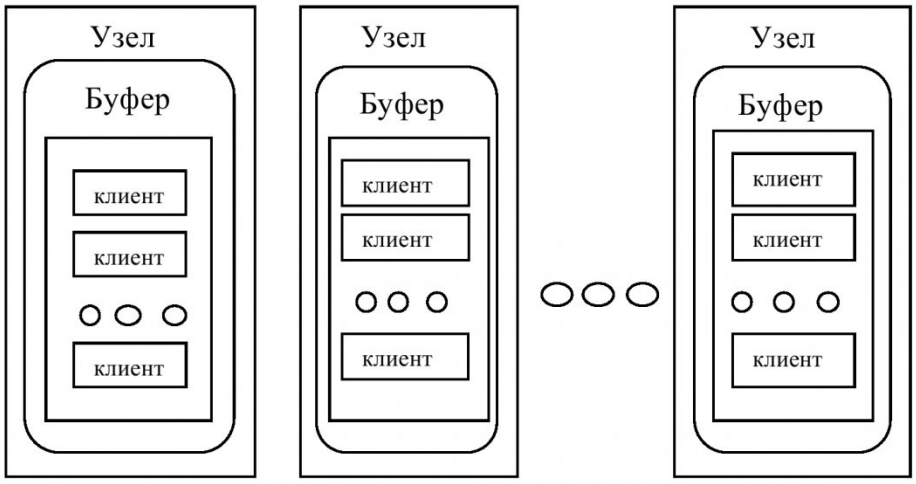

Рис. 1: Фрагмент схематического представления системы опроса.

буется ли обслуживание, или оно закончено, и переходит к следующему узлу.

В результате скопление в каждом узле в системе STDM независимо от скопления в других узлах. Каждое скопление может быть проанализировано отдельно как единственная очередь с *освобождением* сервера. В системах опроса дело обстоит не так, потому что продолжительность посещения сервера узла будет зависеть от особенностей узла. Время до возвращения сервера к узлу будет зависеть от особенностей других узлов в системе.

Особенности системы опроса имеют три категории: особенности клиента, особенности опроса и особенности обслуживания.

Особенности клиента. Поведение клиентов определено оценкой, с которой они прибывают в узел, *оценкой прибытия*, и количеством обслуживания, которого они требуют от сервера, когда они подключен, *требованием обслуживания*. Они имеют стандартные особенности любой модели организации очередей. Время между двумя входами (межвремя прибытия) обычно берется распределенным по экспоненте, хотя другие распределения также были рассмотрены. Предполагалось, что требование обслуживания детерминировано и распределено по экспоненте.

Также рассматривается количество клиентов, которые могли бы ждать обслуживания в любое время. Это определяет *буферную ёмкость*. Два случая, которые рассмотрены в литературе, находятся в бесконечном буфере и выбирают буферные узлы, в которых могут ждать, соответственно, неограниченное число клиентов или только единственный клиент. Однако, была изучена $\hat{E}$ *-ёмкость* буферов, в которых могут ждать $\hat{E}$ клиентов, где $\hat{E}$ - некоторая конечная постоянная. В случае конечных буферов, включая единственный буфер, предполагается, что процесс прибытия приостановлен, когда буфер полон, или что любые последующие клиенты, которые прибыли туда прежде, потеряли место в буфере.

В некоторых моделях предполагается, что клиент занимает место в буфере до полного обслуживания (*ограниченное буферизование*), в то время как в других полагают, что клиент во время обслуживания покидает буфер (*смягченное буферизование*) . Если буфер конечен, то важно различить процесс приостановки от начала прибытия до полного наполнения буфера.

Особенности опроса. Особенность опроса выражается в количестве времени, которое сервер затрачивает на то, чтобы переместиться от узла к другому узлу, на время *блуждания,* и на порядок, по которому сервер решает какой узел посетить следующим. Различные авторы считают, что время блуждания распределено по экспоненте.

Вообще, порядок опроса, как предполагается, является циклическим. Однако, несколько в иной форме он представлен в литературе [15]. Порядок опроса может быть *детерминированным*, *вероятностным* или *зависимым состоянием*.

В детерминированных порядках опроса каждый узел имеет циклический порядок доступа к серверу. Однако если список сформирован, то порядок может измениться.

Например, были изучены системы, в которых:

1) сервер чередует управление циркуляцией между узлами после каждого посещения фиксированного узла;

2) основной узел посещается между каждым посещением других узлов;

3) сервер перемещает узлы согласно некоторому фиксированному порядку в таблице опроса.

В вероятностных порядках опроса следуемый маршрут сервером не предопределен. Когда сервер оставит один узел, он переместится с малой вероятностью. В *случайном* порядке опроса в каждом шаге следующий узел будет узлом i с p_i

вероятностью, где $\sum_{i=1}^{N} p_i = 1$, если N - число узлов. При м*арковском порядке опроса*, направляющая вероятность между узлами p_{ij}, это вероятность того, что когда узел ветви сервера i посетит следующий узел, то это будет узлом j. Время блуждания между узлами может также зависеть от индексов i и j.

В опросе зависимого состояния предъявляются невысокие *требования*, когда сервер перемещается от узла к узлу, его решение, какой узел посетить будет основано на текущем состоянии системы. Например, сервер находится в положении «*жадный сервер»,* он переместится в самый ближайший узел, в котором найдется ждущий клиент, а если он будет пуст, то перемещение отменяется. Это основано на минимальном времени поиска в следующем порядке:

«*кратчайший – поиск – время - начало*»

В *начале переключения* порядка очереди двух систем сервер останется в очереди до тех пор, пока число сообщений, ждущих в другой очереди, не превысит данный порог, или пока длины очередей достигнут указанного размера. Система, как полагают, будет *симметричной,* если все её узлы обладают одинаковыми особенностями клиента (узлы статистически идентичны), и все времена блуждания между узлами в системе одинаковы.

Особенности обслуживания. Во всех случаях сервер достигает узла и находит пустой буфер на следующем узле. Если найден непустой буфер, то немедленно начнется обслуживание первого клиента в буфере. Число ждущих клиентов, которые будут обслужены во время этого посещения узла, зависит от *порядка обслуживания*. Возможности обслуживания, которые были описаны в литературе, являются *исчерпывающими*, *стробированными*, *ограниченными* и *уменьшенными*.

В исчерпывающем обслуживании сервер останется в узле, пока нет никаких клиентов, оставшихся чтобы быть переданными в пустой буфер. В стробированном (запертом) обслуживании сервер останется в узле, пока не поданы все клиенты, которые присутствовали в момент, когда он достиг узла. Любые клиенты, которые прибыли позже, останутся в буфере до следующего посещения сервера. С точки зрения теории организации очередей это наиболее простые случаи.

В $k-$ограниченном обслуживании сервер останется в узле, пока k клиентов не будет обслужено (для некоторой постоянной k), или пока буфер пуст. Например, в случае, когда $k = 1$, т.е. 1- ограниченного обслуживания, которое часто называют просто *ограниченным* обслуживанием, сервер будет обслуживать единственного клиента, прежде чем оставить узел. Для ограниченного числа k, где $k > 1$, порядок должен быть более квалифицированным, чтобы быть исчерпывающим или стробированным.

В уменьшении обслуживания упорядочивается сервер, который останется в узле до пор пока число клиентов, ждущих в буфере, меньше числа, при котором сервер достиг узла. Вариация - рассмотрение обслуживания k-уменьшения, которое должно быть в дальнейшем подходящим, чтобы быть исчерпывающим или стробированным.

2.1 Анализ моделей системы опроса. Критерий качества работы, обычно предъявляемый к системе опроса, является средним *временем ожидания клиента* или распределением. Это время, которое проводит клиент в системе, предшествующей стартовому обслуживанию. Если система будет асимметричной, то эта мера будет отличаться для различных узлов и должна быть вычислена отдельно для каждого узла. Интересны другие меры – *среднее время опроса* (среднее время берет сервер, для того чтобы закончить кругооборот системы), *подразумевается время пребывания клиента* (среднее время, которое клиент тратит для включения системы в обслуживание), *пропускная способность системы* и *средняя длина очереди.*

Большая часть работы, выполненной для систем опроса, привлекает теорию сетей организации очередей и прямую манипуляцию стохастическими процессами. В результате применения многих различных методов возникает особенно много изменений в системах опроса с различными степенями благоприятного исхода. В прошлом десятилетии к точному и приближенному решению моделей было применено много различных комбинированных сложных методов.

Точные явные формы решения, т.е. решения, в которых выражения для критериев качества работы даются в терминах параметров системы, были найдены для симметричного и конечного буфера системы с ограниченным, исчерпывающим или стробированным обслуживанием. Точные решения, основанные на численном решении систем линейных уравнений, были даны для единичных буферных систем (симметричных и асимметричных), и асимметричных в конечных буферных системах с исчерпывающим или стробированным обслуживанием.

Несколько приближенных методов было предложено для систем, которые не приводили к точному решению или для которых точное решение трудоёмко в вычислительном отношении. Они базируются на независимом анализе каждого узла очереди с освобождениями сервера, длина простоя которого находится при анализе взаимодействия между очередями. Это взаимодействие оценено с использованием ожидаемого времени цикла и вероятности того, что сервер считает каждую очередь пустой. Несколько авторов предлагают подобные схемы решения, основанные на этих методах.

Недавно появилось несколько работ, приме-

няющих GSPN при моделировании метода опроса системы [11,16, 17]. В этом подходе модель GSPN системы опроса используется, для того чтобы произвести непрерывный временной марковский процесс. Этот процесс реализован в численном виде для нахождения установившегося состояния решения, из которого получены критерии качества работы.

Ограничения этого подхода были идентифицированы [15-17], все буферы должны быть конечны; все случайные величины, используемые в модели, должны быть распределены по экспоненте; и пространство состояний основной модели растет достаточно быстро. Более точное описание ограничения для конечных буферов при применении отличается от установленного подхода теории организации очередей. Приближенный анализ таких систем был выполнен [17]. В модели GSPN, используя метод стадий, возможно, определить тип распределения в течение времени блуждания, межвремени прибытия и обслуживания. Напротив, задача разрушения пространства состояний серьезная, не допускающая применения методов упрощения, могут быть решены только системы, в которых усреднены сортированные величины. Преимущество подхода GSPN состоит в том, что с асимметричными системами обращаются также как и с симметричными.

2.2 Пример: Модель PEPA для системы опроса. В этом разделе рассмотрена модель PEPA, представляющая собой простую симметричную отдельную буферную систему опроса с ослабленным буферизованием и ограниченным обслуживанием. Модель показана на рис. 3, её компонентами являются сервер и узлы. S_j обозначает сервер, который присутствует в *j-ом* узле системы. Достигая узла, сервер проверит наличие в нем клиента для обслуживания. Если клиент есть, то сервер его обслужит и удалит из буфера в узле прежде, чем переходить к следующему узлу; в противном случае сервер сразу переходит к следующему узлу. У каждого узла j есть два различных состояния в зависимости от того, пуст ли буфер в узле или полон. Эти состояния представлены как два производных компонента узла $Node_{j0}$ и $Node_{j1}$. Вход может произойти только тогда, когда узел пуст; в любом состоянии узел соответственно ответит серверу.

Каждый компонент узла содержит сведения для заполнения буферного пространства в ответ на запрос сервера, указывающего, что буфер пуст, и удаляется сервером, если есть клиент в буфере. Действие сервера включает *блуждание,* которое перемещает его в следующий узел, но возникает ситуация: или узел пуст, или нужно совершить удаление при наличии клиента в узле. Если клиент удален, то сервер, обслуживая клиента, переходит к следующему узлу и реализует следующую деятельность, а именно, деятельность *подачи*.

Рассматриваемая система включает три таких узла, что когда сервер оставляет $Node_3$, то блуждает к $Node_1$. Узлы независимы друг от друга, но каждый из них должен кооперироваться с сервером для деятельности любого *освобождения* или *удаления*.

Будем предполагать, что оценка $empty_j$ деятельности в узле $Node_j$ определяется сервером (оценка в узле не указана). Напротив, деятельность $remove_j$, как предполагается, требует некоторой работы сервера и $Node_j$, а оценка определяется по формуле $r = \min(r_N, r_S)$.

У модели реализованы 72 состояния и 180 переходов. Значения параметров указаны в табл. 1. Был исследован эффект изменения оценок прибытия клиентов в узел от среднего времени ожидания клиента, с тремя различными оценками деятельности *блуждания.* Зависимость времени ожидания клиента W от оценки прибытия клиента λ показана на рис. 3.

Математическая модель зависимости времени ожидания клиента W от оценки λ прибытия клиента, представленная на рис. 3, будет иметь вид:

при $\omega = 10$ W=4.346 $\lambda^{0.9368}\, l^{-0.649\lambda}$;

при $\omega = 15$ W=4.13 $\lambda^{0.8869}\, l^{-0.5677\lambda}$;

при $\omega = 20$ W=3.9774 $\lambda^{0.80198}\, l^{-0.4895\lambda}$/

На рис.4 представлена зависимость времени ожидания клиента W, как от оценки λ прибытия клиента, так и от и оценки блуждания ω.

Математическая модель времени ожидания W клиента зависит как от оценки λ прибытия клиента, так и от и оценки блуждания ω и будет иметь следующий вид:

$$W = (4.70325 - 0.0368175\,\omega)\lambda^{(1.07758 - 0.013488\,\omega)} \times \times \ell^{(0.01594\,\omega - 0.80793)}.$$

Так как система симметрична, то можно использовать любой из узлов, для того чтобы

Таблица 1: Значения параметров для модели опроса на языке PEPA.

in	$remove(N)$	$remove(S)$	$remove(Polling)$	*empty*	*serve*	*walk*
λ	r_N	r_S	$r = \min(r_N, r_S)$	e	μ	w
0.1-0.9	50	100	50	100	1	10,15,20

$$Node_{j0} \overset{def}{=} (in, \lambda).Node_{j1} + (empty_j, \bot).Node_{j0} \qquad 1 \le j \le N$$

$$Node_{j1} \overset{def}{=} (remove_j, r_N).Node_{j0}$$

$$S_j \overset{def}{=} (remove_j r_S).(serve, \mu).(walk, w).S_{j\oplus 1} + (empty_j, e).(walk, w).S_{j\oplus 1}, \quad \text{где } j \oplus 1 = 1 \text{ когда } j = N$$

$$Polling \overset{def}{=} (Node_{10} \| Node_{20} \| Node_{30}) \underset{(empty_j, remove_j)}{\triangleright\!\triangleleft} S_1 \quad \text{где } 1 \le j \le N$$

Рис. 2: Модель симметричной системы опроса со смягченным буферизованием на языке PEPA

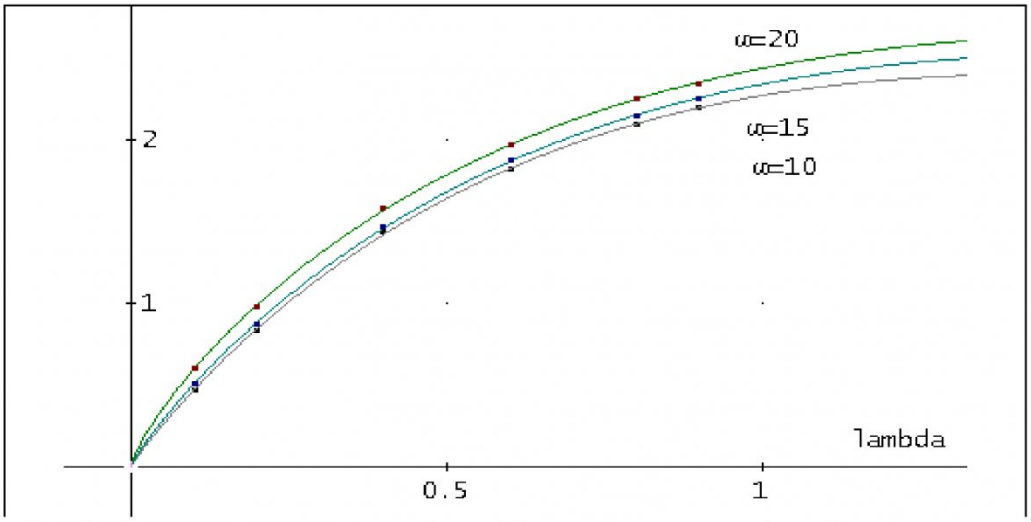

Рис. 3.: Зависимость времени ожидания W клиента от оценки λ прибытия клиента

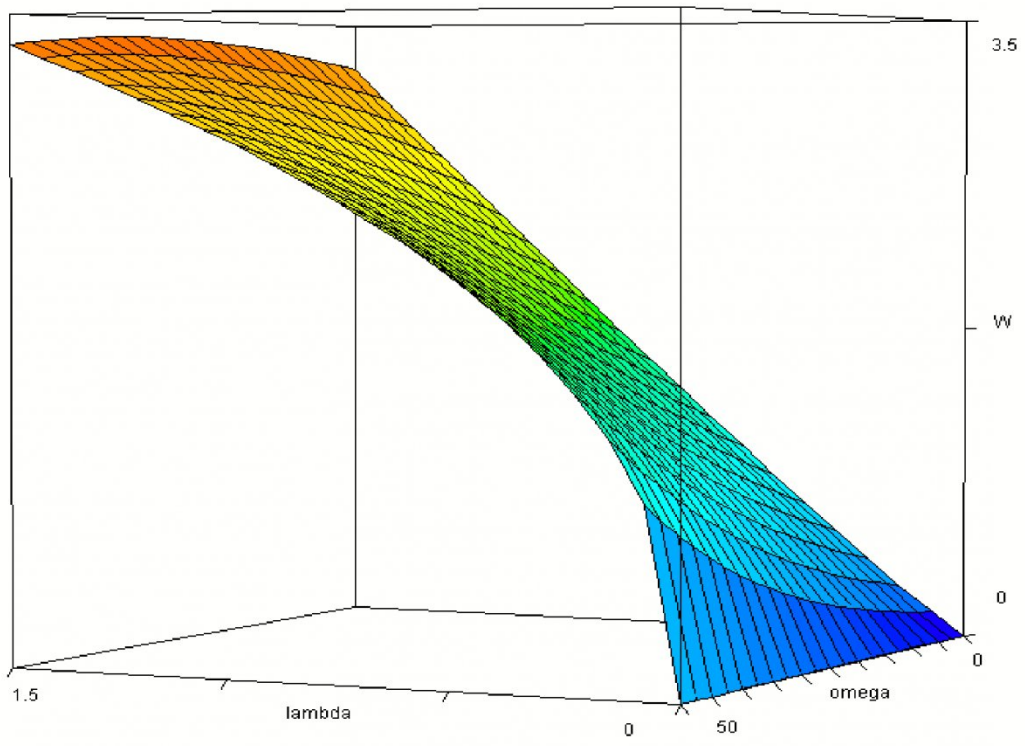

Рис. 4: Зависимость времени ожидания W клиента как от оценки λ прибытия клиента, так и от оценки блуждания ω.

вычислить среднее время ожидания W клиента,поскольку оно одинаково во всех узлах.

Для нахождения W применен закон о малом числе испытаний для $Node_1$, чтобы определить среднее время окончания деятельности $remove_1$. По закону о малом числе испытаний для состояний имеем, что среднее число объектов в системе равно произведению среднего числа оценок, по которому объекты прибывают в систему, на среднее время объекта , в течение которого он является представителем в системе.

Этот закон справедлив для всех систем, для которых существуют эти средние числа.

Среднее число клиентов в буфере N находится из предположения, что полная деятельность соответствует 1. Когда разрешена деятельность in , то она вычисляет R_{in}. Клиент присутствует всякий раз, когда деятельность in не разрешена и тогда $N = 1 - R_{in}$.

Пропускная ёмкость узла это пропускная способность $remove_1$ деятельности X_{remove_1}, и она находится при помощи r деятельности $remove_1$. В действительности это присоединяет r со всеми состояниями, для которых занят буфер в $Node_1$ и когда сервер присутствует в узле. Поскольку обслуживание имеет место вне узла, то в отличие от ограниченных буферизовавших систем время пребывания клиентов в узле равно среднему времени ожидания клиента.

Таким образом, из закона о малом числе испытаний следует, что

$$W = N / X_{remove_1} = (1 - R_{in}) / X_{remove_1} .$$

СПИСОК ЛИТЕРАТУРЫ

1. *Сорокин А.С.* Марковские процессы в теории надежности технологических систем гидродобычи угля // Вестник КузГТУ. 2008. № 1. С. 61-69.

2. *Коэн Дж., Боксма О.* Граничные задачи в теории массового обслуживания. М.: МИР, 1987.

3. *Королюк В.С., Томусяк А.А.* Описание функционирования резервированных систем посредством полумарковских процессов. //Кибернетика, вып.5, 1965.

4. *Баруча-Рид А.Т.* Элементы теории марковских процессов и их применение. Наука, М., 1969. 511С.

5. *Hillston J.* A Compositional Approach to Performance Modelling. Cambridge Univ. Press, 1996.

6. *Сорокин А.С.* Применение полумарковских процессов к определению характеристик надежности технологических схем. // Вестник КузГТУ, 2005. № 1. С. 3 -9.

7. *Сорокин А.С.* Структурное моделирование надежности технологических систем с использованием скелетонов// Вестник КузГТУ, 2008. № 4(68). С. 31-45.

8. *Сорокин А.С.* Математическое моделирование оценки надежности технологических систем// Вестник КузГТУ. 2008. № 5(69). С. 28-37.

9. *Сорокин А.С.* Применение методов теории вероятностей к исследованию некоторых процессов производства.//Труды 4-ой междунар. Конф. Кибернетика и технологии XXI века. Воронеж, 2003. С. 312-323.

10. *Сорокин А.С.* Алгоритм решения систем уравнений Колмогорова (Оценка качества системы).//Вторая Всероссийская научная конф. Проектирование инженерных и научных приложений в среде MATLAB. М., 2004. С. 389 – 397.

11. *Marsan, M.A., Donatelli S., Neri F.* GSPN Models of Markovian Multiserver Multiqueue Systems.// Performance Evaluation, 11, 1990.

12. *Raith T.* Performance Analysis of Multibus Interconnection Networks in Distributed Systems.//In M. Akiyama, editor, Teletraffic Issues in an Advanced Information Society ITC-11. Elsevier, 1985.

13. *Morris R.J.T., Wang Y.T.* Some Results for Multiqueue Systems with Multiple Cyclic Servers. In H. Rudin and W. Bux, editors, //Performance of Computer Communication Systems. Elsevier, 1984.

14 *Yuk T.I. , Palais J.C.* Analysis of Multichannel Token Ring Networks.//In Proceedings of the International Conference on Communication Systems, 1988.

15. *Takagi H.* Queueing Analysis of Polling Models: An Update.//In H. Takagi, editor, Stochastic Analysis of Computer and Communication Systems. IFIP/North Holland, 1990.

16. *Choi H., Trivedi K.S.* Approximate Performance Models of Polling Systems Using Stochastic Petri Nets//In Proceedings of INFOCOM' *92*, 1992.

17. *Ibe O.C. , Trivedi K.S.* Stochastic Petri Net Models of Polling Systems.//IEEE Journal on Selected Areas of Communication, 8(9), 1990.

□ Автор статьи:

Сорокин
Андрей Семенович
- канд. физ.-мат.наук, доцент, ст.н.с.
(филиал КузГТУ , г. Новокузнецк).
Тел.: 8(3843) 772459