

СЕКЦИЯ

ТЕОРИЯ ТЕЛЕТРАФИКА

Руководитель – д.т.н., профессор **СТЕПАНОВ С.Н.**
д.т.н., профессор **ХАРКЕВИЧ А.Д.**

УСТОЙЧИВЫЕ МОДЕЛИ ТРАФИКА МУЛЬТИСЕРВИСНЫХ СЕТЕЙ

Цитович И.И.

ИППИ РАН

В докладе рассматриваются задачи исследований, связанные с описанием абонентского трафика, обслуживаемого мультисервисной сетью связи (МСС). Основная цель исследования - выделить основные компоненты трафика таким образом, чтобы обеспечить разумный компромисс между точностью описания моделей функционирования МСС, пригодных для анализа качества обслуживания, и сложностью описания обслуживаемого трафика. Основное внимание необходимо обратить на выделение наиболее существенных компонент, которые необходимы для построения математической модели обслуживаемого трафика, позволяющей производить оценки качества услуг связи. Под устойчивостью модели понимается непрерывная зависимость выходных параметров от входных. В нашем случае это означает, что при малом изменении параметров, задающих обслуживаемый трафик, мало изменяются вероятности потерь требований в узлах сети.

В зависимости от сложности модели используются различные параметры абонентского трафика. Однако точность расчетов зависит не только от точности модели, которая описывает функционирование сети, но и от точности входных параметров. Это обстоятельство должно учитываться при разработке моделей, поскольку при большой неопределенности в значениях входных параметров будет и неопределенность в значениях выходных параметров моделей. С учетом этого обстоятельства и данных измерений и должны выделяться основные компоненты трафика исходя из принципа существенности этой компоненты на расчет итоговых характеристик модели. Принципы разделения трафика учитывают различия в тарификации трафика, в обслуживании трафика сетью и различия в модели прогнозирования этой компоненты трафика.

Математические модели расчета пропускной способности телекоммуникационных сетей или их фрагментов используют различные данные о трафике в зависимости от степени детализации сети и точности ее описания. Наиболее грубые модели учитывают лишь средние значения характеристик или их доверительные границы. Такие модели используются обычно для вычисления предварительных грубых оценок необходимой пропускной способности сети. Такие же модели целесообразно использовать и в тех случаях, когда точные и надежные значения исходных значений параметров трафика недоступны по каким-то причинам.

При проведении более точных расчетов пропускной способности узлов и каналов сети используются различные модели описания обслуживаемой нагрузки как случайного процесса. Известно, что в зависимости от емкости узла или пучка доля дополнительной емкости, необходимой для обеспечения обслуживания нагрузки с заданным качеством, относительно уменьшается и теоретически может быть сколь угодно малой. Однако усложнение в последнее время природы абонентского трафика приводит к тому, что учет случайных флуктуаций может приводить к перебоям в обслуживании даже для больших узлов, поэтому рассматриваются различные математические модели абонентского трафика, использующие и различные средства его описания.

Классические модели телетрафика, ориентированные на передачу голосовой нагрузки, используют различные марковские или сводящиеся к ним модели входного потока вызовов.[1,2,4] Однако такой подход далеко не всегда дает удовлетворительный результат, поскольку в реальных сетях связи присутствуют эффекты обратной связи и зависимости между вызовами абонента или различных абонентов.

Анализ средств телекоммуникаций на современном этапе показывает стремительное движение по пути интеграции [3]. Интеграция как процесс широко проявляется в таких направлениях развития телекоммуникаций как: цифровизация, унификация, глобализация и стандартизация. Желание пользователя получать одновременно услуги традиционной телефонии и передачи данных определяют вектор развития сетей связи. Направление этого вектора - проектирование и построение цифровых сетей связи с интеграцией служб.

Интеграция служб, имеющих различный по природе трафик (аудио, видео, данные), частично или полностью реализуется рядом телекоммуникационных технологий.

Необходимо отметить, что распространение новых видов услуг меняет характер нагрузки, передаваемой по цифровым линиям [5]. Речевой трафик с коммутацией каналов вытесняется нагрузкой, новых коммуникационных приложений, использующих для своей транспортировки протоколы коммутации пакетов (КП). Уже сейчас в развитых странах мира доля речевого трафика, передаваемого по технологии коммутации каналов, составляет менее половины всего объёма нагрузки. Это не позволяет надеяться, что весь трафик МСС можно описать с использованием лишь одной из математических моделей абонентского трафика.

Одним из важных статистических свойств нагрузки, создаваемой современными приложениями, является наличие зависимости между отдельными её частями, которая не исчезает при изменении шкалы времени в сторону укрупнения, как это происходит с традиционными случайными процессами, используемыми для описания нагрузки в теории телетрафика. Так, например, для пуассоновского процесса соответствующая корреляция всегда равна нулю, а для марковского процесса с конечным числом состояний - экспоненциально убывает к нулю. Интеграция в сетях с КП различного по своей природе трафика, а так же особенности используемых технологий передачи данных для построения сетей связи и повышенная активность пользователей сетей, являются некоторыми из причин проявления самоподобного характера сетевого трафика или, иначе, его фрактальных свойств [6,7].

Анализ современных технологий и сетей передачи информации показывает, что необходимо рассматривать вектор абонентской нагрузки, состоящий из нескольких компонент, каждая из которых должна отражать свойства трафика некоторых групп абонентов с одинаковыми интегральными характеристиками. Поскольку предполагается, что деление на компоненты осуществлено таким образом, что каждая из них описывает трафик принципиально различного происхождения: например одна – голосовой трафик, вторая – трафик данных, третья – видео в режиме реального времени и т.п., то можно считать эти компоненты статистически независимыми.

В связи с этим становится понятным, что обслуживаемый трафик X должен рассматриваться как сумма нескольких компонент:

$$X = \sum_{i=1}^n X_i = \sum_{i=1}^n (\lambda_i t + \sqrt{a_i \lambda_i} Z_t^{(i)}) = (\sum_{i=1}^n \lambda_i) t + \sum_{i=1}^n \sqrt{a_i \lambda_i} Z_t^{(i)}, \quad (1)$$

где X_i - некоторый стандартный процесс, описывающий i -ую компоненту трафика, и процессы $X_i, i=1,2,\dots,n$, статистически независимы. Для описания компонент целесообразно использовать диффузионное приближение [10]. Это означает, что $Z_t^{(i)}$ - фрактальное броуновское движение со значением параметра Херста H_i . Таким образом, предметом рассмотрения становится смесь нескольких процессов с различными значениями параметра Херста.

Ясно, что даже при сделанных предположениях процесс Y имеет сложную природу и трудно исследовать его свойства. Поэтому наиболее важным является выяснить, насколько существенно зависят его свойства от значений параметра Херста его компонент. Понятно, что с ростом параметра Херста такие статистические свойства процесса как предсказуемость, размах значений и т.п. лишь ухудшаются. С другой стороны, точное определение значения параметра Херста затруднительно даже в том случае, когда имеется лишь одна компонента. Поэтому понятно, что определить статистическими средствами все параметры, характеризующие смесь (1), является практически невозможно, особенно если интенсивности отдельных компонент являются относительно малыми. Поскольку основным параметром, характеризующим качество обслуживания МСС, является вероятность потери ячеек, то наиболее важной является задача определить чувствительность этого параметра от значений параметра Херста процессов, входящих в рассматриваемую смесь.

В [8,9] рассмотрена двухкомпонентная смесь. Смысл такого предложения состоит в том, чтобы рассмотреть ситуацию, когда присутствуют компоненты с наибольшим и наименьшим из возможных значений параметра Херста, а остальные компоненты добавлены к той или иной из рассматриваемых. Такое упрощение будет давать процесс, который будет по своим статистическим свойствам хуже или лучше исходного. Там исследована устойчивость модели относительно значения параметра Херста в следующем смысле. Предположим, что в смеси присутствует компонента с малой интенсивностью (т.е. относительно других компонент малым значением величины $C_i = \sqrt{a_i \lambda_i}$), но большим значением параметра Херста. Если малое значение C_i не приводит к сильному изменению вероятности потерь ячеек, то модель можно считать устойчивой относительно наличия малых примесей трафика с большим значением параметра Херста. Если же наличие компонент с малой интенсивностью приводит к существенному увеличению вероятности потерь, то это означает, что модель чувствительна к значению параметра Херста. В последнем случае мы не можем считать,

что модель трафика, основанная на предположении, что весь трафик описывается процессом с фиксированным значением параметра Херста, соответствующим группе абонентов, дающих основной вклад в общую интенсивность обслуживаемого трафика, т.е. однокомпонентная модель, будет адекватной моделью, позволяющей рассчитывать МСС с заданным качеством обслуживания.

Проведенные исследования показали, что устойчивости относительно параметра Херста нет. Это означает, что при проектировании МСС необходимо учитывать все компоненты трафика; при наличии трафика, параметр Херста которого больше проектного, возможны перегрузки в сети, которые будут возникать существенно чаще, чем расчетное значение, полученное при проектировании.

Следовательно, при проведении исследований свойств МСС необходимо выделить основные группы пользователей по их статистическим свойствам, существенно влияющим на значение параметра Херста, и далее в качестве моделей обслуживаемого МСС трафика рассматривать модели вида (1). При этом важно отметить, что имеется достаточно широкий, с практической точки зрения, диапазон изменения интенсивностей компонент трафика, при котором сохраняется устойчивость функционирования сети, что позволяет гарантировать качество обслуживания сообщений при наличии регулирования доступа к сети, поскольку в транзитных узлах анализировать структуру смеси (1) статистическими средствами практически невозможно.

Литература

1. Городецкий А. Я., Заборовский В.С. Информатика. Фрактальные процессы в компьютерных сетях. Учебное пособие. СПб.: Издательство СПбГТУ. 2000.
2. Корнышев Ю. Н., Пшеничников А. П., Харкевич А. Д. Теория телетрафика. М.: Радио и связь. 1996. С. 272.
3. Кучерявый А. Е. Новые технологии телекоммуникации на сетях связи Российской Федерации. Вестник связи, N4. 2000. С. 83 -87.
4. Лагутин В. С., Степанов С. Н. Телетрафик мультисервисных сетей связи. М.: Радио и связь. 2000. С. 320.
5. Митилино С. Фрактальная катастрофа TCP/IP . Компьютерное обозрение. №9. 2001.
6. Назаров А. Н. Модели трафика служб с битовой скоростью передачи информации в широкополосных цифровых сетях интегрального обслуживания. Автоматика и телемеханика, №9. 1998.
7. Нейман В. И. Самоподобные процессы и их применение в теории телетрафика. Труды МАС, № 1(9). 1999. С. 11-15.
8. Шмелев И. В. Влияние фрактальных процессов на сетевой телетрафик в современных распределенных информационных сетях. / Реинжиниринг бизнес-процессов на основе информационных технологий. М.: Моск. Госуд. Ун-т экономики, статистики и информатики. 2004. С.11-12
9. Шмелев И. В. Модель трафика мультисервисной сети на основе смеси самоподобных процессов / Международный форум информатизации МФИ-2004. М.:МТУСИ. 2004. С. 12
10. Norros I. A storage model with self-similar input. Queueing systems. V16. P. 387-396.

ROBUST MODELS OF MULTYSERVICE SYSTEM'S TRAFFIC

Models of an input traffic based on a approximation by the fractal diffusion processes are discussed. It is found that the multiservice system's traffic need be a mixture of self-similar processes with different values of Herst parameter for describing the systems with a guaranteed probability of cells loss.



МЕТОД РАСЧЕТА ВЕРОЯТНОСТЕЙ БЛОКИРОВОК ДЛЯ ЛИНЕЙНОГО ФРАГМЕНТА СЕТИ WDM С МНОГОАДРЕСНЫМИ СООБЩЕНИЯМИ

Башарин Г.П., Грубник А.В.

Российский университет дружбы народов

В последние годы идет активное развитие волоконно-оптических сетей, которые позволяют передавать огромные объемы данных на большие расстояния и с большой скоростью. Благодаря использованию технологии мультиплексирования с разделением длины волны (wavelength division multiplexing – WDM) по одному кабелю могут одновременно передаваться многие сигналы, что в сотни раз увеличивает пропускную способность сети [1].

Особый интерес представляют полностью оптические сети (all optical networks – AON), в которых сигнал передается только по оптической среде и отсутствуют оптико-электронные преобразования, которые существенно снижают производительность сети. Среди сетей AON выделяются сети с маршрутизацией по длине волны (wavelength-routed networks). Такая сеть состоит из узлов – оптических коммутаторов (optical

cross-connect – ОХС), соединенных между собой оптическими волокнами, в которых имеется некоторое количество WDM-каналов или длин волн. Маршрутом передачи сообщения в сети является набор звеньев. Для одноадресных сообщений, которые передаются от единственного источника к единственному получателю, маршрутом является линейный путь, а для многоадресных сообщений, которые передаются от единственного источника к фиксированному набору получателей, маршрутом является дерево. Сообщения передаются по звеньям сети с использованием определенной длины волны. Если в узлах сети отсутствуют волновые конвертеры, то на всех звеньях маршрута передачи сообщения длина волны должна быть одинакова, это т. н. требование непрерывности длины волны (wavelength continuity constraint – WCC). Два сообщения могут одновременно передаваться по одному звену, если они установлены на различных длинах волн, это т. н. требование различия длины волны (distinct wavelength constraint).

Данная работа посвящена исследованию метода расчета вероятностей блокировок сообщений во фрагменте сети с маршрутизацией по длине волны, способной передавать одноадресные и многоадресные сообщения. Сделаем следующие предположения: (i) в узлах сети не установлены волновые конвертеры; (ii) на каждом звене сети имеется W длин волн; (iii) маршруты передачи сообщений фиксированы; (iv) потоки заявок на установление соединений являются пуассоновскими с параметром ρ_r , время обслуживания распределено экспоненциально с параметром μ_r , где r – номер маршрута; (v) при установлении соединения длина волны для передачи сообщения назначается случайным образом из всех доступных; (vi) заявка, которая не может быть обслужена, блокируется и теряется.

1. Фрагмент сети с одноадресными сообщениями

Сначала рассмотрим двухзвенный фрагмент сети, в которой передаются только одноадресные сообщения. Перенумеруем имеющиеся маршруты передачи одноадресных сообщений: маршрут 1 – однозвенный маршрут передачи сообщения от узла 0 к узлу 1, маршрут 2 – однозвенный маршрут передачи сообщения от узла 1 к узлу 2, маршрут 3 – двухзвенный маршрут передачи сообщения от узла 0 к узлу 2 (см. рис. 1).

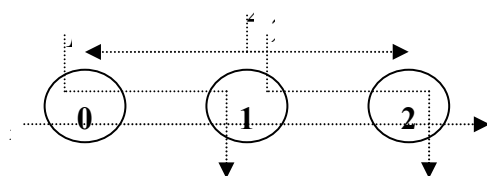


Рис. 1. Двухзвенный фрагмент сети с одноадресными и многоадресными сообщениями

Функционирование такой системы описывается четырехмерным марковским процессом $M_2=(n_1, n_2, n_3, f_3)$, где $n_r, r \in \{1,2,3\}$ – число активных соединений на маршруте r , f_3 – число длин волн, свободных одновременно на всех звеньях маршрута 3. Однако такой процесс не является обратимым, поэтому путем модификации некоторых интенсивностей перехода между состояниями марковского процесса строится приближенный марковский процесс [2, 4], который имеет мультипликативное решение:

$$\pi_{n_1, n_2, n_3, f_3} = \pi_{0, 0, 0, W} \cdot \frac{\rho_1^{n_1} \rho_2^{n_2} \rho_3^{n_3}}{n_1! n_2! n_3!} \cdot \frac{\binom{W - n_1 - n_3}{f_3} \binom{n_1}{W - n_3 - n_2 - f_3}}{\binom{W - n_3}{W - n_3 - n_2}} \quad (1)$$

2. Пуассоновский поток многоадресных сообщений

Теперь рассмотрим сеть, в узлах которой установлены оптические сплиттеры [1] и которая способна передавать как одноадресные, так и многоадресные сообщения (от единственного источника к множеству получателей). Маршрутом многоадресного сообщения в общем случае является дерево и соединение может быть установлено, если существует длина волны, свободная одновременно на всех звеньях сети, составляющих такое дерево [3].

Рассмотрим линейный фрагмент сети с маршрутизацией по длине волны, состоящий из двух звеньев (см. рис. 1). В такой модели имеется три маршрута передачи одноадресных соединений (два однозвенных и один двухзвенный) и один маршрут передачи многоадресных сообщений.

К множеству маршрутов одноадресных сообщений, рассмотренному в п. 1, добавляется маршрут 4 – маршрут передачи многоадресного сообщения от узла 1 к узлам 0 и 2. $R=\{1, 2, 3, 4\}$ – множество маршрутов.

В данном случае узел-источник 1 многоадресного сообщения и узлы-получатели 0 и 2 лежат на одном линейном маршруте, узлы 0 и 2 – крайние узлы этого маршрута. С точки зрения оценки вероятностей блокировок, такое многоадресное сообщение эквивалентно одноадресному сообщению от узла 0 к узлу 2.

Пусть λ_r – интенсивность пуассоновского потока сообщений на маршрут $r \in \mathbf{R}$, время передачи сообщений распределено экспоненциально с параметром μ_r и $\rho_r = \lambda_r / \mu_r$. Тогда такая сеть с многоадресными сообщениями может анализироваться как сеть, в которой имеются только одноадресные сообщения и для которой: $\tilde{\rho}_3 = \rho_3 + \rho_4$, где ρ_3 – предложенная нагрузка на двухзвенный маршрут передачи одноадресных сообщений, ρ_4 – предложенная нагрузка на маршрут передачи многоадресных сообщений.

3. Поток многоадресных сообщений с интенсивностью, зависящей от состояния

На практике не допускается ситуация, когда на одном маршруте устанавливается более одного многоадресного соединения. Чтобы учесть это требование, рассмотрим модель пуассоновского потока многоадресных сообщений с интенсивностью, зависящей от состояния системы.

Пусть теперь $1/\lambda_4$ – среднее время до поступления многоадресного сообщения, при условии, что в системе нет установленного многоадресного соединения, $1/\mu_4$ – среднее время передачи сообщения, $\rho_4 = \lambda_4 / \mu_4$. Таким образом, процесс поступления многоадресных сообщения является пуассоновским с интенсивностью, зависящей от состояния. Для нахождения вероятностей блокировки сообщений в такой сети, используем приближенную оценку.

Предположим, что на каждом звене сети имеется $W=\infty$ длин волн. Тогда состояние сети описывается только числом активных (установленных) соединений на каждом маршруте, включая многоадресные. Пусть $N_r(t)$ – число активных соединений на маршруте $r \in \mathbf{R}$ в момент времени $t \geq 0$. Тогда состояние всей сети в момент времени t описывается марковским процессом $\mathbf{S}(t) = \{N_1(t), N_2(t), N_3(t), N_4(t)\}$. Пусть $p(\mathbf{n}) = P\{\mathbf{S}(t) = \mathbf{n}\}$ в равновесном состоянии, $\mathbf{n} = \{n_1, n_2, n_3, n_4\}$, $n_i \geq 0$. Поскольку число длин волн считаем неограниченным, то марковский процесс $\mathbf{S}(t)$ имеет мультипликативное решение, и $p(\mathbf{n})$ может быть записано в виде:

$$p(n_1, n_2, n_3, n_4) = \prod_{r=1}^4 P_r(n_r), \text{ где } P_r(n_r) - \text{вероятность того, что на маршруте } r \text{ имеется } n_r \text{ активных соедине-}$$

ний. Для одноадресных сообщений выполняется: $P_r(n_r) = \frac{\rho_r^{n_r}}{n_r!} e^{-\rho_r}$, $r \in \{1, 2, 3\}$. Для многоадресных сообщений рассмотрим уравнение равновесия:

$$\begin{cases} \mu_4 P_4(1) = \lambda_4 P_4(0) \\ P_4(n_4) = 0, n_4 > 1 \\ P_4(1) + P_4(0) = 1 \end{cases}$$

Откуда получаем:

$$P_4(0) = \frac{1}{1 + \rho_4}, P_4(1) = \frac{\rho_4}{1 + \rho_4} \text{ и } P_4(n_4) = 0 \text{ при } n_4 > 1 \quad (2)$$

Тогда средняя интенсивность Λ_4 поступления многоадресных сообщений равна:

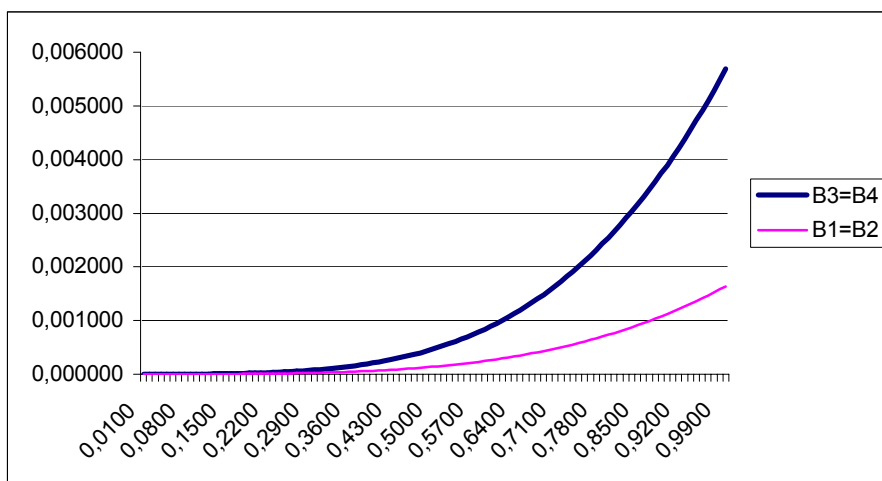
$$\Lambda_4 = \lambda_4 P_4(0) + 0 + 0 + \dots = \lambda_4 P_4(0) = \frac{\lambda_4}{1 + \rho_4} \quad (3)$$

Используя такую интенсивность, можно применить метод, описанный в п. 2 для нахождения вероятностей блокировки многоадресного сообщения.

В случае длинных маршрутов целесообразно применять алгоритмы декомпозиции маршрута на более короткие фрагменты, которые анализируются аналогично [2, 5].

Численный анализ

На рис. 2 представлена зависимость вероятностей блокировки многоадресных и одноадресных сообщений в двухзвенном фрагменте сети от величины нагрузки на одноадресные соединения, которая зависит от параметра a : $\rho_1 = \rho_2 = 0.5a$, $\rho_3 = 0.25a$.



Литература

- [1] Гринфилд Д. Оптические сети // К.: ООО «ТИД «ДС», 2002 г.
- [2] Zhu Y., Rouskas G.N., Perros, H.G. A Path Decomposition Approach for Computing Blocking Probabilities in Wavelength-Routing Networks // IEEE/ACM Transactions on Networking, 8(6): 747-762, December 2000.
- [3] Ramesh S., Rouskas G.N., Perros, H.G. Computing Blocking Probabilities in Multiclass Wavelength-Routing Networks with Multicast Calls// IEEE Journal on Selected Areas in Communications, Vol. 20 No. 1 89-96, January 2002.
- [4] Basharin G. P. Savochkin E. A. Groubnik A. V. A Product-form Solution for a Class of Wavelength Routed All-Optical Networks without Wavelength Conversion // Вестник РУДН Т. 2, №1 2003 г., стр. 13-22.
- [5] Башарин Г.П., Савочкин Е.А. Анализ пропускной способности линейного фрагмента оптической сети с маршрутизацией по длине волны.// Электросвязь 2005 г. (принято в печать).

We study a class of all-optical wavelength-routing network with unicast and multicast calls. We develop an analytical method to compute the call-blocking probabilities in two-hop linear fragment. For multicast calls we consider state-dependent Poisson arrival process.

СИСТЕМА МАР/РН/С С АДРЕСНОЙ СТРАТЕГИЕЙ ПОВТОРНЫХ ВЫЗОВОВ

Мушко В.В.

Белорусский государственный университет

Важным разделом теории массового обслуживания является теория систем с повторными вызовами ("retrials" или "repeated customers"). Такие системы характеризуются следующим поведением. Если вызов поступил в систему, в которой все приборы и места для ожидания (если имеется очередь) заняты, то он покидает систему на некоторое случайное время, иначе говоря, уходит на орбиту ("orbit" или "retrial group"), а затем повторяет попытки попасть на обслуживание.

Системы с повторными вызовами отличаются числом приборов (одноканальные или многоканальные), видом интенсивности повторных вызовов (постоянная, классическая, линейная, квадратичная и т. д.), размерами орбиты (конечная или бесконечная), наличием очереди (конечная или бесконечная), наличием нескольких типов вызовов (например, обладающие более высоким приоритетом, настойчивые или нетерпеливые), стратегией повторов (например, классическая стратегия, стратегия полного доступа), стратегией первичного доступа (например, использование зарезервированных каналов).

В данной работе будет рассмотрена адресная стратегия доступа к приборам.

Рассматривается многоканальная система, имеющая s идентичных приборов. Время обслуживания вызова имеет распределение фазового типа (общепринятый термин для такого процесса обслуживания – РН (Phase-type)-обслуживание (см., например, [1]). Процессом обслуживания управляет цепь Маркова m_t , $t \geq 0$ с непрерывным временем. Состояние процесса m_t , $t \geq 0$ в момент начала обслуживания опре-

деляется стохастическим вектором-строкой $\bar{v} = (v_1, \dots, v_M)$. Переходы процесса $m_t, t \geq 0$, которые не приводят к завершению обслуживания, определяются неприводимой матрицей S размерности $M \times M$. Диагональные элементы матрицы S отрицательны и $-S_{m,m}$ определяет параметр экспоненциально распределенного времени пребывания процесса в состоянии $m, |S_{m,m}| < \infty, m = \overline{1, M}$. Недиagonalные элементы матрицы S определяют интенсивности переходов процесса $m_t, t \geq 0$ в пространстве состояний $\{1, \dots, M\}$. Значения $-\sum_{m'=1}^M S_{m,m'}$ определяют интенсивности перехода процесса $m_t, t \geq 0$ из состояния

m в абсорбирующее состояние. Момент перехода процесса $m_t, t \geq 0$ в абсорбирующее состояние определяет момент завершения обслуживания. Интенсивности переходов, которые приводят к завершению обслуживания, определяются вектором $S_0 = -Se$, где e – вектор-столбец, состоящий из единиц. Предполагается, что все компоненты вектора-столбца S_0 неотрицательны и, по крайней мере, одна из них положительна.

Поток, входящий в систему, является Марковским (\overline{MAP} – Markovian Arrival Process, введен в [2]). Поведением \overline{MAP} управляет цепь Маркова $z_t, t > 0, z_t = \overline{0, W}$, с непрерывным временем. Элементами матриц $D_k, k = \overline{0, 1}$ являются интенсивности перехода этой цепи. Переходы цепи, управляемые матрицей D_0 , не приводят к генерации вызова, в то время как, переходы цепи, управляемые матрицей D_1 , сопровождаются генерацией вызова. Обозначим через \bar{i} вектор-строку стационарных вероятностей цепи $z_t, t > 0$. Он удовлетворяет уравнениям $\bar{i}(D_0 + D_1) = \bar{0}$, и $\bar{i}e = 1$, где $\bar{0}$ – вектор-строка, состоящий из нулей.

В момент прибытия вызов выбирает для обслуживания r -й прибор с вероятностью $q_r, 0 < q_r < 1, r = \overline{1, c}, \sum_{r=1}^c q_r = 1$. Если выбранный прибор свободен, то вызов занимает прибор и после обслуживания покидает систему. Если выбранный прибор занят, то вызов направляется в некоторую виртуальную динамическую область, называемую орбитой, и пытается получить обслуживание позже. Каждый вызов, находящийся на орбите, делает повторные попытки через интервалы времени, имеющие экспоненциально распределенную длину с параметром $\bar{b}, \bar{b} > 0$, независимо от других вызовов. В момент повтора вызов

выбирает r -й прибор с вероятностью $\varphi_r, 0 < \varphi_r < 1, r = \overline{1, c}, \sum_{r=1}^c \varphi_r = 1$. Если прибор свободен, то вызов за-

нимает его и покидает систему после обслуживания. Если прибор занят, то вызов возвращается на орбиту, даже если несколько других приборов свободно в этот момент. Вызовы с орбиты пытаются получить обслуживание до тех пор, пока им не удастся занять прибор, выбранный при соответствующей попытке.

Рассмотрим процесс $\bar{z}_t = \{i_t, m_t^{(1)}, \dots, m_t^{(c)}, z_t\}, t \geq 0$, где $i_t, i_t \geq 0$ – число вызовов на орбите в момент времени $t, m_t^{(r)}$ – состояние r -го прибора в момент времени $t, r = \overline{1, c}$:

$$m_t^{(r)} = \begin{cases} 0, & \text{если } r\text{-й прибор свободен в момент времени } t, \\ m, & \text{если } r\text{-й прибор обслуживает вызов на } m\text{-ой фазе} \\ & \text{в момент времени } t, m = \overline{1, M}, \end{cases}$$

$z_t, t > 0$ – состояние управляющего процесса \overline{MAP} -потока, $z_t = \overline{0, W}$,

Очевидно, что процесс $\bar{z}_t, t \geq 0$ является цепью Маркова с непрерывным временем. Обозначим стационарное распределение этой цепи через

$$p(i, m_1, \dots, m_c, z) = \lim_{t \rightarrow \infty} P\{i_t = i, m_t^{(1)} = m_1, \dots, m_t^{(c)} = m_c, z_t = z\}, \\ i \geq 0, m_r = \overline{0, M}, r = \overline{1, c}, z = \overline{0, W}.$$

Условие существования данного предела приведено ниже и предполагается далее выполненным. Занумеруем состояния цепи Маркова $\mathfrak{X}_t, t \geq 0$ в лексикографическом порядке и сформируем векторы-строки вероятностей \vec{P}_i , соответствующие состоянию i числа вызовов на орбите. Размерности этих векторов равны $(M+1)^c \bar{W}$, $\bar{W} = W+1$. Сформируем также макровектор $\vec{P} = (\vec{P}_0, \vec{P}_1, \dots, \vec{P}_i, \dots)$.

Лемма. Вектор \vec{P} является единственным решением системы: $\vec{P}Q = \vec{0}$, $\vec{P}e = 1$, где инфинитезимальный генератор Q цепи Маркова $\zeta_t, t \geq 0$ имеет следующую форму:

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & 0 & 0 & \dots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & 0 & \dots \\ 0 & Q_{2,1} & Q_{2,2} & Q_{2,3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (1)$$

где блоки $Q_{i,j}$ вычисляются следующим образом: $Q_{i,i-1} = i\bar{b}L_1, i \geq 1$, $Q_{i,i} = A_1 - i\bar{b}B_1, i \geq 0$, $Q_{i,i+1} = S_1, i \geq 0$, где матрицы A_1, B_1, L_1, S_1 , размерности $(M+1)^c \bar{W}$, вычисляются из следующих рекуррентных отношений:

$$\begin{aligned} A_{c+1} &= D_0, B_{c+1} = I_{\bar{W}}, L_{c+1} = O_{\bar{W}}, S_{c+1} = O_{\bar{W}}, \\ A_{c-k+1} &= \begin{pmatrix} A_{c-k} & \bar{b} \otimes I_{(M+1)^{k-1}} \otimes D_1 q_{c-k+1} \\ S_0 \otimes I_{(M+1)^{k-1}\bar{W}} & S \otimes A_{c-k} \end{pmatrix}, k = \overline{1, c}, \\ B_{c-k+1} &= \text{diag}\{B_{c-k}, I_M \otimes (B_{c-k} - \varphi_{c-k+1} I_{(M+1)^{k-1}\bar{W}})\}, k = \overline{1, c}, \\ L_{c-k+1} &= \begin{pmatrix} L_{c-k} & \varphi_{c-k+1} \bar{b} \otimes I_{(M+1)^{k-1}\bar{W}} \\ O_{M(M+1)^{k-1}\bar{W} \times (M+1)^{k-1}\bar{W}} & L_{c-k} \end{pmatrix}, k = \overline{1, c}, \\ S_{c-k+1} &= \text{diag}\{S_{c-k}, I_M \otimes (S_{c-k} + I_{(M+1)^{k-1}} \otimes D_1 q_{c-k+1})\}, k = \overline{1, c}, \end{aligned}$$

где O_l – нулевая матрица размерности l , $O_{l,m}$ – нулевая матрица размерности $l \times m$, I_l – тождественная матрица размерности l , \otimes – символ кронекерова произведения матриц, \oplus – символ кронекеровой суммы матриц (см., например, [1]), $\text{diag}\{a_1, \dots, a_c\}$ – диагональная матрица с диагональными блоками a_1, \dots, a_c .

Очевидно, что цепь Маркова $\mathfrak{X}_t, t \geq 0$ принадлежит к классу асимптотически квазитеплицевых цепей Маркова с непрерывным временем [3]. Поэтому аппарат асимптотически квазитеплицевых цепей Маркова с непрерывным временем может быть применен для установления условий существования стационарного распределения и вычисления вектора \vec{P} стационарных вероятностей.

Обозначим $\tilde{Y}(z) = \tilde{Y}_0 + \tilde{Y}_1 z + \tilde{Y}_2 z^2, |z| \leq 1$, где $\tilde{Y}_k = \lim_{i \rightarrow \infty} P_{i,i+k-1}, k = 0, 1, 2$,

$$P_{i,i-1} = i\bar{b}(F + i\bar{b}\mathfrak{F})^{-1}L_1, i \geq 1, P_{i,i} = I + (F + i\bar{b}\mathfrak{F})^{-1}(A_1 - i\bar{a}B_1), i \geq 0, P_{i,i+1} = (F + i\bar{b}\mathfrak{F})^{-1}S_1, i \geq 0,$$

где матрица F состоит из диагональных элементов матрицы A_1 , матрица \mathfrak{F} получена из единичной матрицы I , заменой диагональных элементов, номера которых совпадают с номерами нулевых диагональных элементов матрицы B_1 , на 0.

Как следует из [3], достаточное условие существования стационарного распределения цепи Маркова $\mathfrak{X}, t \geq 0$ имеет следующую форму: $\bar{X} \bar{Y}'(1) \mathbf{e} < 1$, где вектор \bar{X} удовлетворяет уравнениям: $\bar{X} \bar{Y}'(1) = \bar{X}$, $\bar{X} \mathbf{e} = 1$. Используя это утверждение, можно доказать следующую теорему.

Теорема. Стационарное распределение цепи Маркова $\mathfrak{X}, t \geq 0$ существует, если выполнено следующее условие:

$$\bar{X}[I - B_1 + \bar{F}^{-1}(A_1 + 2S_1)]\mathbf{e} < 1,$$

где вектор \bar{X} удовлетворяет уравнениям:

$$\bar{X}[L_1 - B_1 + \bar{F}^{-1}(A_1 + S_1)] = \bar{0}, \bar{X}\mathbf{e} = 1.$$

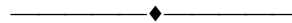
Стационарное распределение \bar{p} может быть найдено с помощью алгоритма, предложенного в [3].

Литература

1. Бочаров П.П., Печинкин А.В. Теория массового обслуживания. – М.: УДН, 1995. – 528 с.
2. Lucantoni D.M., Meier-Hellstern K.S., Neuts M.F. A single-server queue with server vacations and a class of non-renewal arrival processes // Adv. Appl. Prob. – 1990. – Vol. 22. – P. 676–705.
3. Klimenok V.I., Dudin A.N. Continuous-time asymptotically quasi-Toeplitz Markov chains // Queues: flows, systems, networks. Proceedings of the International conference "Mathematical methods of optimization of telecommunication networks", 22 – 24 February, 2005, Minsk. – P. 77–85.

THE MAP/PH/c SYSTEM WITH ADDRESSED RETRIALS

Abstract. Multiserver retrial model in which arrivals occur according to a Markovian arrival process. The service time has phase-type distribution. The intensity of retrials linearly depends on the number of the customers from the orbit. The continuous-time multi-dimensional Markov chain describing the behavior of the system belongs to the class of multi-dimensional asymptotically quasi-Toeplitz Markov chains. Sufficient condition for stationary state distribution is proven.



НАХОЖДЕНИЕ КАЧЕСТВА ОБСЛУЖИВАНИЯ АБОНЕНТОВ АТС ПРИ КОНЕЧНОЙ НАДЕЖНОСТИ СОЕДИНИТЕЛЕЙ И ОГРАНИЧЕНИЯХ НА ОБЪЕМ ЗИП

Цимбал В.А., Косарева Л.Н., Кулешов В.В.

Серпуховской военной институт РВ

Качество обслуживания абонентов АТС оценивается, в основном, вероятностью отказов в обслуживании заявок по причине занятости всех соединителей станции. В теории телетрафика задачи подобного типа рассмотрены подробно, при этом получены первая и вторая формула Эрланга, формула Энгсета и их обобщения [1]. Кроме того, рассмотрены задачи нахождения качества обслуживания абонентов в случае ненадежных обслуживающих приборов (соединителей) и их восстановления при бесконечном числе резервных соединителей [1, 2].

Известно, что на каждой АТС имеется конечный комплект ЗИП на соединители. В сложных экономических условиях, переживаемых страной, пополнение ЗИП – процедура достаточно длительная, и на стадии эксплуатации АТС имеются этапы, когда ЗИП на соединители полностью расходуется. Нахождение качества обслуживания входящей нагрузки на данном этапе – предмет данной задачи.

АТС представляется полнодоступной коммутационной схемой (ПКС), количество соединителей в которой равно ν_0 . На вход ПКС поступает нагрузка в виде простейшего потока заявок с интенсивностью λ . Обслуживание поступившей заявки происходит с интенсивностью μ при экспоненциальном времени обслуживания заявки. Каждый соединитель может выходить из строя с интенсивностью a , причем время его безотказной работы распределено экспоненциально. В АТС реализован способ скользящего резервирования соединителей [3], для этого в ЗИП АТС имеется $\nu_{ЗИП}$ соединителей. Восстановление вышедшего из строя соединителя осуществляется с интенсивностью b , распределение времени восстановления экспоненциальное.

Из способа скользящего резервирования известно, что среднее время исчерпания ЗИП равно [3]

$$\bar{\tau}_{ЗИП} = \frac{\nu_{ЗИП}}{a\nu_0}. \quad (1)$$

Более точно интервал функционирования АТС до исчерпания ЗИП $\tau_{ЗИП}$ можно определить из формулы Пуассона, определяющей вероятность выхода из строя равно $\nu_{ЗИП}$ соединителей за рассматриваемый интервал [1].

$$P_{\nu_{ЗИП}}(\tau_{ЗИП}) = \frac{(\nu_0 a \tau_{ЗИП})^{\nu_{ЗИП}}}{\nu_{ЗИП}!} e^{-\nu_0 a \tau_{ЗИП}}. \quad (2)$$

Так как аналитически разрешить выражение (2) относительно $\tau_{ЗИП}$ невозможно, используем метод итераций [4], при этом основная формула данного метода имеет вид

$$\tau_{k+1} = \frac{1}{\nu_0 a} \nu_{ЗИП} \sqrt{\nu_{ЗИП}! P e^{\nu_0 a \tau_k}}, \quad (3)$$

и условие сходимости метода итераций таково

$$\text{при } k=0 \quad \tau_k \geq \frac{1}{\nu_0 a} \quad (4)$$

Отметим, что интервал времени, найденный методом итераций, существенно меньше интервала, определённого выражением (1).

На интервале $\tau_{ЗИП}$ соединители будут выходить из строя по одному вследствие ординарности данного процесса. Восстановление неисправного соединителя из состава ЗИП будет происходить за время b^{-1} . Следовательно, качество обслуживания поступающих вызовов на соединение, определяемых первой формулой Эрланга, будет иметь вид

$$P_{отк} = 1 - K_{\Gamma} [1 - E_{\nu_0}(y)], \quad \text{где} \quad (5)$$

$$K_{\Gamma} = \frac{b-a}{b+a} - \text{коэффициент готовности ПКС}, \quad (6)$$

$$E_{\nu_0}(y) = \frac{y^{\nu_0} / \nu_0!}{\sum_{i=0}^{\nu_0} y^i / i!} - \text{первая формула Эрланга}, \quad (7)$$

$$y = \frac{\lambda}{\mu} - \text{интенсивность поступающей на ПКС нагрузки}. \quad (8)$$

Однако, на интервале восстановления вышедшего из строя соединителя b^{-1} возможен выход из строя одиночным образом ещё $(\nu_{ЗИП} - 1)$ соединителей. Вероятность выхода из строя одного соединителя из совокупности $\nu_{ЗИП}$ на рассматриваемом интервале вследствие экспоненциального распределения времени данного процесса равна

$$P_1(b^{-1}) = 1 - e^{-\nu_{ЗИП} a b^{-1}} \quad (9)$$

Учитывая, что произведение $\nu_{ЗИП} a b^{-1} \ll 1$, получаем, что $P_1(b^{-1}) = \nu_{ЗИП} a b^{-1}$.

Тогда вероятность обслуживания поступающих заявок равна

$$P_{обсл} = K_{\Gamma} \left[\sum_{i=1}^{\nu_{ЗИП}} P_1^{i-1}(\tau) (1 - E_{\nu_0-i}(y)) \right]. \quad (10)$$

В свою очередь, вероятность отказа в обслуживании заявок будет равна

$$P_{отк} = 1 - K_{\Gamma} \left[\sum_{i=1}^{\nu_{ЗИП}} P_1^{i-1}(\tau) (1 - E_{\nu_0-i}(y)) \right]. \quad (11)$$

По исчерпанию всех соединителей ЗИП в АТС начнется процесс деградации, связанный с выходом из строя рабочих соединителей без их замены. В этом случае вероятность обслуживания, а соответственно и вероятность отказа, можно найти, анализируя переходный режим работы АТС с конечным составом ЗИП на

соединители. При этом для вероятностей состояний процесса обслуживания заявок необходимо составить систему линейных однородных дифференциальных уравнений (СЛОДУ) Колмогорова. Матричная запись такой СЛОДУ имеет вид:

$$\frac{d\bar{P}(t)}{dt} = H \cdot \bar{P}(t). \quad (12)$$

где

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{1n} \\ h_{21} & h_{22} & h_{23} & h_{2n} \\ h_{31} & h_{32} & h_{33} & h_{3n} \\ h_{n1} & h_{n2} & h_{n3} & h_{nn} \end{pmatrix} \quad (13)$$

– основная матрица системы (12).

$$\bar{P}(t) = \begin{pmatrix} P_1(t) \\ P_2(t) \\ \vdots \\ P_n(t) \end{pmatrix} = (P_1(t), P_2(t), P_3(t), \dots, P_n(t))^T \quad (14)$$

– матрица-столбец неизвестных функций вероятностей состояний.

Граф состояний и переходов процесса обслуживания заявок в этом случае будет треугольным [5].

Количество состояний графа, которыми описывается процесс обслуживания заявок в условиях выхода соединителей из строя, равно

$$n = \frac{(v_0 + 1)(v_0 + 2)}{2} \quad (15)$$

Для автоматизированного синтеза СЛОДУ (12) сформированы следующие правила, определяющие процедуры нахождения элементов главной диагонали, под главной диагональю и над главной диагональю, при этом используются такие параметры графа:

d – номер вертикального ряда (показывает число соединителей, занятых обслуживанием заявок);

c – номер горизонтального ряда (показывает число неисправных соединителей);

u – номер уровня (диагонали) (показывает число соединителей, не участвующих в обслуживании заявок по причине их занятости или неисправности).

Правило 1. Элемент главной диагонали

$$h_{ii} = \begin{cases} 0, & \text{при } d = 0 \wedge c = v \\ -(\lambda + (v - c)a), & \text{при } d = 0 \wedge c \neq v \\ -(d\mu + (v - c)a), & \text{при } c + d = v \\ -(d\mu + \lambda + (v - c)a), & \text{при } d \neq 0 \wedge c + d \neq v \end{cases} \quad (16)$$

Правило 2. Элемент под главной диагональю

$$h_{i,i-u-1} = \begin{cases} \lambda, & \text{при } d \neq 0 \\ 0, & \text{при } d = 0 \end{cases} \quad (17)$$

$$h_{i,i-u} = \begin{cases} a(v + 1 - d - c), & \text{при } c \neq 0 \\ 0, & \text{при } c = 0 \end{cases} \quad (18)$$

Правило 3. Элемент над главной диагональю

$$h_{i,i+1} = \begin{cases} a(u - c + 1), & \text{при } c \neq 0 \\ 0, & \text{при } c = 0 \end{cases} \quad (19)$$

$$h_{i,i+u+2} = \begin{cases} \mu(d + 1), & \text{при } u < v \\ 0, & \text{при } u = v \end{cases} \quad (20)$$

Вероятность отказа в обслуживании заявок будет находиться как сумма вероятностей состояний графа, расположенных на главной диагонали, т.е.

$$P_{отк}(t) = \sum_{i=n-\nu}^n P_i(t). \quad (21)$$

Решение СЛОДУ можно осуществить одним из известных численных методов, например, методом Рунге-Кутты. По результатам решения и анализа вероятности отказа (выражение 21) можно найти интервал деградации АТС, после которого обслуживание заявок с допустимой вероятностью невозможно.

Для обеспечения дальнейшего обслуживания отдельных абонентов (например, приоритетных) необходимо вводить ограничения либо на число абонентов станции, либо на информационную нагрузку неприоритетных абонентов.

Литература

1. Шнепс М.А. Системы распределения информации. Методы расчета: Справочное пособие. – М.: Связь, 1979 – 344 с.
2. Методы определения оптимальной надежности элементов сетей связи/ Суторихин Н.Б., Буров П.Н., Захарова С.М.; Под ред. Н.Б. Суторихина. – М.: Связь, 1979 – 104 с.
3. Надежность технических систем: Справочник/ Под ред. И.А. Ушакова. – М.: Радио и связь, 1985. – 608 с.
4. Баврин И.И., Матросов В.А. Общий курс высшей математики: Учеб. для студентов физ.-мат. спец. пед. вузов.-М.: Просвещение, 1995,-464с.
5. Косарева Л.Н. Синтез систем линейных алгебраических уравнений, описывающих поведение систем массового обслуживания с неидеальными восстанавливающимися приборами // XXII Межведомственная научно-техническая конференция. – Серпухов – 2003.– ч.4. – с.137-141.

DETERMINING QUALITY OF SERVICE OF AUTOMATIC TELEPHONE STATION SUBSCRIBERS BY FINAL RELIABILITY OF CONNECTORS AND RESTRICTIONS OF RESERVE ELEMENTS

Tsimbal V., Kosareva L., Kuleshov V.

Serpukhov Military Institute of Rocket Forces

The automatic telephone station in which the service channels fail and are restored through reserve is considered. The reserve of channels is limited. The service quality of the entering applications can be calculated both at the presence of reserve and at its absence. The stationary operation mode and transitive operation mode are analyzed.



ПОСТРОЕНИЕ МАРКОВСКИХ МОДЕЛЕЙ ДОВЕДЕНИЯ МНОГОПАКЕТНЫХ СООБЩЕНИЙ В СЕТИ ПЕРЕДАЧИ ДАННЫХ МЕТОДОМ РАНГОВЫХ СОСТОЯНИЙ

Кирдяшов Ф.Г., Солдатенко Э.Н., Солнцев М.В.

Серпуховской военной институт РВ

В современных сетях передачи данных (СПД) используется большое количество протоколов канального уровня доведения сообщений. Наиболее распространенным является протокол HDLC, который хорошо адаптирован к каналам достаточно плохого качества и ориентирован на передачу одно пакетных сообщений, при этом каждый передаваемый пакет квитируется самостоятельным образом [1]. Процесс доведения сообщений по протоколу HDLC хорошо описывается теорией конечных марковских цепей (КМЦ), и достаточно детально изложен в [2]. Отметим, что вероятностно-временные и временные характеристики процесса доведения сообщений описываются уравнением Колмогорова-Чепмена (УКЧ) и требуют знания матриц переходных вероятностей (МПВ), соответствующей КМЦ.

Однако, во многих СПД с каналами хорошего качества используется протокол передачи много пакетных сообщений, в котором квитирование осуществляется после передачи нескольких пакетов, при этом на недоведенные пакеты выдается соответствующий запрос на повторение [3].

Мы рассматриваем двузвенную СПД, в которой реализован указанный протокол доведения много пакетных сообщений с квитированием, например, после передачи двух пакетов. Анализ такого процесса доведения сообщения показывает, что он также является конечной марковской цепью с одним поглощающим состоянием.

Граф включает следующие состояния:

S_0 – квитанция за все сообщения получена, конец сеанса связи;

S_1 – два пакета получены, передача квитанции за пакет;

S_2 – принят один пакет при передаче одного пакета, передача квитанции за пакет;

S_3 – передача сообщения из двух пакетов;

S_4 – передача одного непринятого пакета;

S_5 – получение одного пакета при передаче двух пакетов, запрос непринятого пакета.

Матрица переходных вероятностей для данной КМЦ имеет вид

$$P_{[k,k]} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ p_{10} & 0 & 0 & p_{13} & 0 & 0 \\ p_{20} & 0 & 0 & 0 & p_{24} & 0 \\ 0 & p_{31} & 0 & p_{33} & 0 & p_{35} \\ 0 & 0 & p_{421} & 0 & p_{54} & 0 \\ 0 & 0 & 0 & p_{53} & p_{54} & 0 \end{bmatrix}$$

Из рассмотрения физики процесса доведения многопакетных сообщений и анализа графа ясно. Что при увеличении числа пакетов в сообщении существенно увеличиваются как размеры графа, так и размеры МПВ КМЦ. Данный факт требует машинного синтеза МПВ для последующего решения УКЧ.

Для решения этой задачи разработан метод ранговых состояний. Он заключается в том, что множество состояний КМЦ разбивается на состояния с различными рангами и в соответствии с ними каждому состоянию присваивается его порядковый номер (осуществляется нумерация состояний графа).

Граф данной КМЦ является двудольным. При этом в одно подмножество двудольного графа входят состояния, принадлежащие отправителю сообщения, а во второе подмножество входят состояния, принадлежащие получателю сообщения. Метод ранговых состояний разбивает множество состояний КМЦ на четыре ранговые группы. При присвоении состоянию того или иного ранга определяется принадлежность состояния к своей ранговой группе, а затем определяется ранг внутри самой группы. Рассмотрим указанные ранги групп и состояний.

В группу первого ранга входит одно состояние, которое соответствует полному доведению сообщения и получению квитанции об этом. Это состояние есть состояние окончания сеанса связи, оно в КМЦ является поглощающим.

В группу второго ранга входят состояния, для которых получено ровно столько пакетов, сколько было передано. И чем больше пакетов было получено, тем большим рангом обладает состояние внутри группы.

В группу третьего ранга входят состояния передачи различного числа пакетов. При этом, чем больше пакетов передается, тем большим рангом обладает состояние внутри группы.

Группу четвертого ранга составляют так называемые неблагоприятные состояния, так как в этих состояниях получатель получает только какую-то часть от всех передаваемых пакетов. Для определения ранга состояния внутри этой группы рассчитывают процентное соотношение числа принятых пакетов к числу передаваемых пакетов. И чем больше это соотношение, тем большим рангом обладает состояние внутри группы. При этом может возникнуть ситуация, когда появится несколько равных ранговых соотношений. Большой ранг в этом случае отдается тому состоянию, в которое передавалось меньшее число пакетов. Таким образом, при определении ранга состояния в четвертой группе получатель сообщения как бы отбирает для себя более благоприятные состояния.

Разделение множества состояний модели на ранговые группы позволило вывести формулу общего числа испытаний КМЦ. Это число зависит от того, на сколько пакетов разбивается сообщение, и равно сумме состояния отправителя и получателя сообщения.

Число состояний отправителя определяется как сумма состояний первой и третьей ранговых групп. При этом первая группа содержит одно состояние, а количество состояний третьей группы равно числу пакетов – l , на которое разбивается сообщение.

Из этого следует, что число состояний отправителя равно

$$m = l + 1 \tag{1}$$

Количество состояний получателя сообщения найдем через сумму арифметической прогрессии (так как на приемной стороне могут возникать неблагоприятные состояния). Оно равно

$$n = \frac{l(l+1)}{2} \tag{2}$$

Сложив числа состояний, принадлежащих получателю и отправителю сообщения, получим выражение для размерности МПВ

$$k = \frac{l^2 + 3l + 2}{2}. \quad (3)$$

Разбиение множества состояний на состояния с различными рангами позволило не только определить число состояний КМЦ, но и сформировать правила синтеза элементов МПВ.

В зависимости от того, что множество состояний разбивается на четыре ранговых группы матрица переходных вероятностей заполняется в четыре этапа. При этом заполнении происходит по строкам последовательно от одного этапа к другому.

Этапы (правила) заполнения МПВ таковы:

1. Группа первого ранга содержит одно состояние, поэтому в МПВ заполняется на данном этапе одна строка. В данную строку записывается вероятность завершения сеанса связи, которая имеет координаты $i = 0$ и $j = 0$, где i - строки матрицы, j - столбцы матрицы, $i = j$.

2. На этом этапе происходит заполнение строк, принадлежащих состояниям группы второго ранга. Количество строк равно количеству пакетов, на которые разбиты сообщения. Строки данного этапа будут содержать вероятности получения квитанции за пакеты, за сообщение и вероятности неполучения за пакеты и за сообщения. При этом вероятности получения квитанций будут иметь координаты $i = 1, \dots, l$ и $j = 0$. Вероятности неполучения квитанции будут иметь координаты $i = 1, \dots, l$ и $j = i + 1$.

3. На этом этапе происходит заполнение строк, принадлежащих состояниям группы третьего ранга. Количество заполняемых строк будет равно количеству пакетов, на которые разбито сообщение. Заполнение МПВ начинается с заполнения главной диагонали той части строк, состояния которых принадлежат к группе третьего ранга. Заполнение главной диагонали в этой части матрицы обусловлено тем, что вершины графа, которым соответствуют состояния отправителя сообщения, имеют петли. На главной диагонали будут находиться вероятности недоведения различного числа пакетов с одного повтора. Ниже главной диагонали располагаются вероятности получения различного числа пакетов. При этом, чем больше номер строки, тем меньше пакетов получено. Эти вероятности будут находиться в столбцах, сдвинутых на l столбцов от главной диагонали. Выше главной диагонали стоят вероятности получения только части от всех передаваемых пакетов. Вероятности будут располагаться так, что чем меньше пакетов было получено, тем на большее число столбцов они сдвинуты выше от главной диагонали. Количество данных вероятностей будет уменьшаться от $l - 1$ до 1 ровно на единицу, если продвигаться от строки с наименьшим порядковым номером к строке с наибольшим порядковым номером в этой части матрицы. Вероятности в отдельных строках данной части матрицы отстоят от главной диагонали на различное число столбцов. При этом самые «близкие» вероятности к главной диагонали отодвинуты от нее на l столбцов, а последующие вероятности отстоят от предыдущих им вероятностей на число столбцов, которое уменьшается ровно на единицу, начиная с величины 1.

4. На этом этапе происходит заполнение строк, принадлежащих состояниям группы четвертого ранга. В этих строках находятся вероятности получения и неполучения запроса на передачу различного числа пакетов. Данный этап разделяется на несколько небольших подэтапов, окончание каждого из которых дает начало следующему подэтапу. Количество подэтапов будет на единицу меньше числа пакетов, на которые разбито сообщение. Заполнение МПВ происходит следующим образом. Под вероятностью, находящуюся в последней строке третьего этапа, которая располагается на главной диагонали, записывается вероятность получения запроса. Запись этих вероятностей происходит в столбец до окончания подэтапа. На этой же строке, являющейся первой строкой подэтапа, но в столбце с номером, равным номеру столбца, в который записана вероятность, находящаяся на главной диагонали первой строки третьего этапа, записывается вероятность неполучения запроса на передачу неприятых пакетов. При этом вероятности неполучения запроса по диагонали «приближаются» к столбцу, в котором находятся вероятности получения запроса на передачу неприятых пакетов. Признаком окончания подэтапа является «сближение» вероятностей получения и неполучения запроса. Признаком следующего подэтапа является то, что столбец, содержащий вероятности получения запроса, смещается на один столбец влево. Вероятность неполучения запроса располагается в том же столбце, что и в первом подэтапе. Далее процесс внутри каждого подэтапа повторяется, и после того, как будут выполнены все подэтапы, заполнение МПВ заканчивается.

Конкретные значения ненулевых элементов МПВ находятся по биномиальному закону, исходя из физики процесса доведения пакетов сообщения.

Литература

1. Протоколы информационно-вычислительных сетей. Справочник./ С.А.Аничкин и др. – М.: Радио и связь, 1990. – 504 с.
2. Цимбал В.А., Косарева Л.Н. Корректная математическая модель доведения сообщений по протоколу HDLC в вариантах «точка-многоточка»//LVII научная сессия, посвященная Дню Радио. Труды, т. 2, - М., 2002. – с. 209-211.
3. Передача дискретных сообщений. Учебник для высших учебных заведений./Под ред. В.П.Шувалова. – М.: Радио и связь, 1990. – 462 с.

SYNTHESIS OF MARKOVIAN MODELS OF MULTIPARK MESSAGE BRINING IN TWO RUNG DATA COMMUNICATION NETWORK

Kirdjashov F., Soldatenko E., Solntsev M.

The Serpukhov military institute of rocket armies

In article developed the method of priority status which allows to automatize the process of synthesis of markovian models which are necessary to research the system of database transportation.



МОДЕЛЬ ДВУХУРОВНЕВОЙ СЕТИ СОВОЙ ПОДВИЖНОЙ СВЯЗИ С РАЗНОТИПНЫМИ ВЫЗОВАМИ

Ефимушкин В.А., Чашин А.А.

ФГУП ЦНИИС

Переход к сетям 3G ставит задачу выделения частотных ресурсов в условиях их ограниченности, что приводит к необходимости оптимизации распределения каналов при предоставлении услуг. В работе были исследованы вероятностно-временные характеристики двухуровневого кластера, состоящего из одной макросоты и нескольких микросот (рис.1) с двумя типами заявок.

Предполагается, что голосовые (г-) вызовы обслуживаются в макросотах, а мультимедиа (м-) вызовы – в микросотах. При отсутствии в макросоте достаточной для обслуживания ширины полосы пропускания (ШПП), г-вызов направляется в микросоту. После освобождения в макросоте достаточной для обслуживания г-вызова ШПП, он возвращается на обслуживание в макросоту. Мультимедиа вызов обслуживается аналогичным образом с обратным порядком перехода (рис.1). На рис.2 схематично изображен алгоритм обслуживания поступающих вызовов, а также переходы вызовов между макро- и микросотами, связанные с переполнением емкости сот (недостатком ШПП).

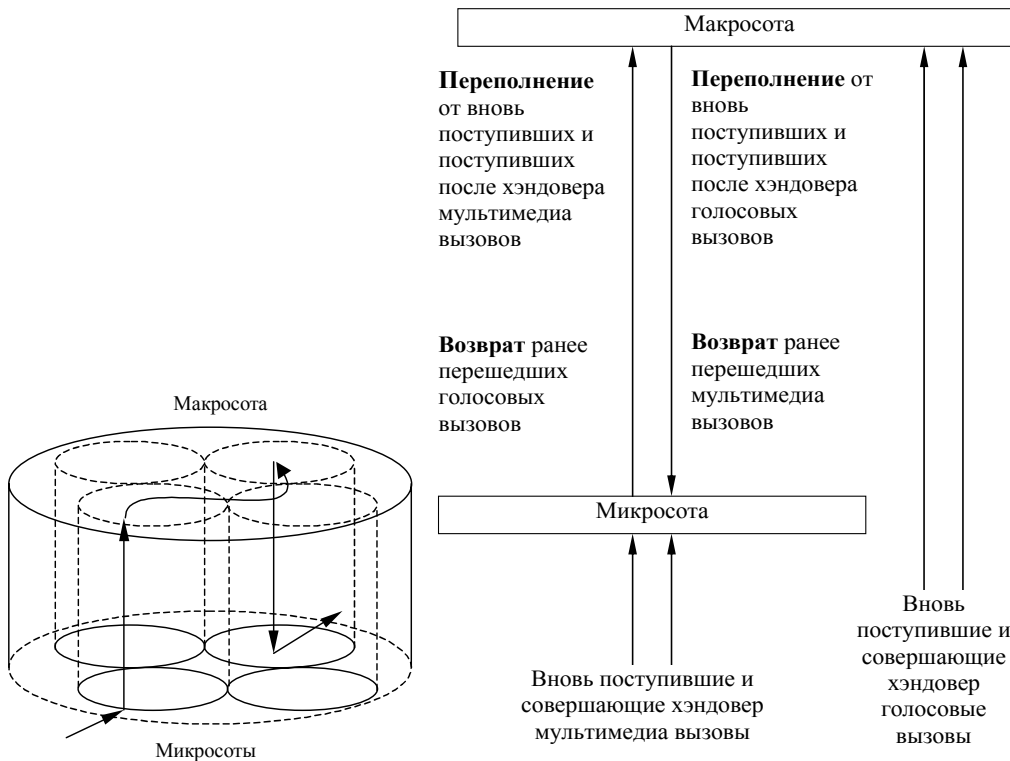


Рис.1. Обслуживание вызова по траектории его движения

Рис.2. Схема и порядок обслуживания вызовов различных типов в кластере

На рис.3 представлена модель функционирования двухуровневого кластера сети сотовой подвижной связи в виде открытой экспоненциальной сети массового обслуживания. Узлы СеМО с номерами от 1 до N состоят из C_m приборов, соответствующих числу каналов микросоты (одинаковому для всех микросот), а узел с номером $N+1$ - из C_r приборов, соответствующих числу каналов макросоты.

С вероятностью α_i , $i = \overline{1, N}$, поступившая заявка идет на узел i (при этом она становится м-заявкой) и с вероятностью α_{N+1} - на узел $N+1$ (при этом она становится г-заявкой).

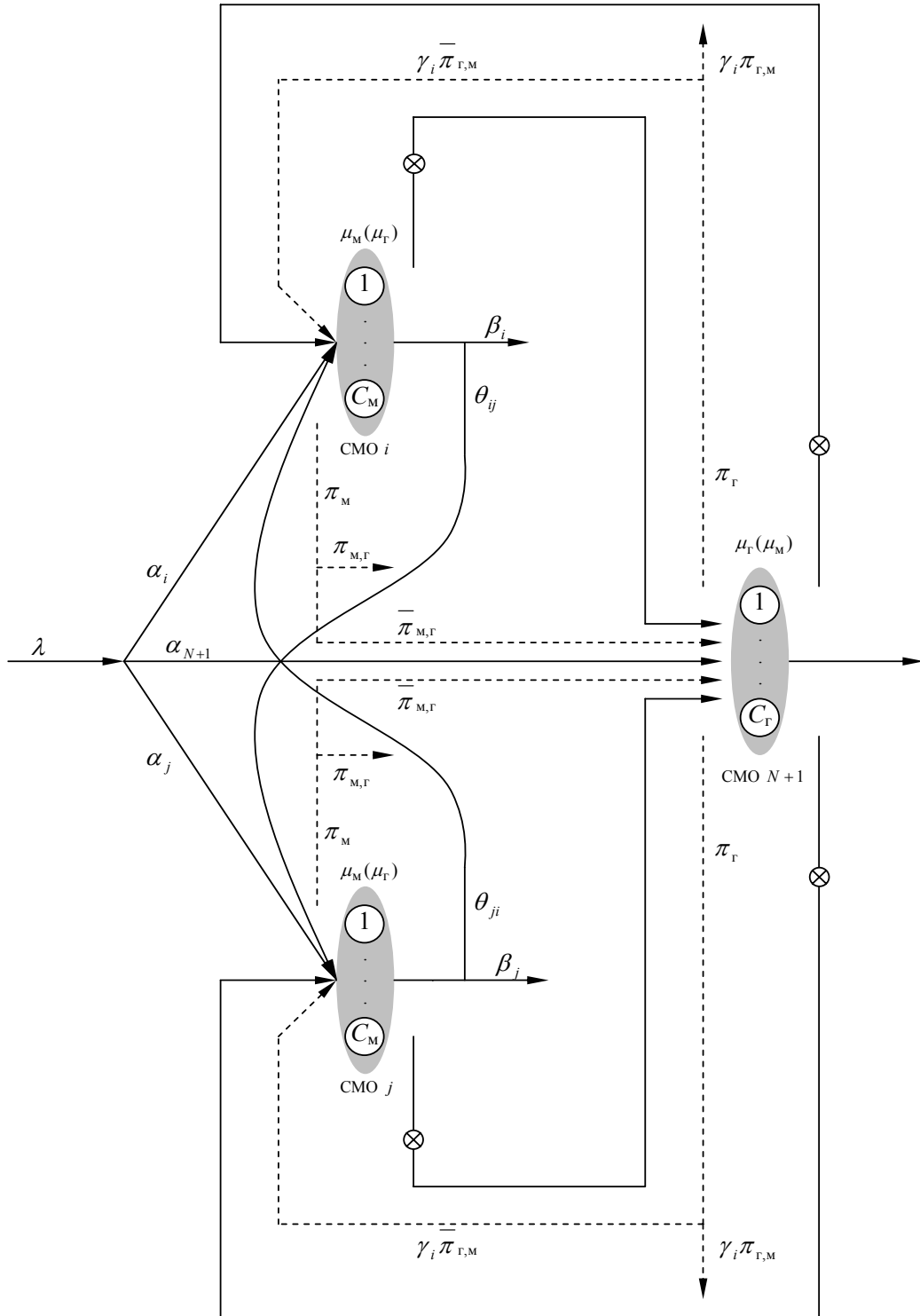
Время обслуживания м- и г-заявок имеет экспоненциальное распределение с параметрами μ_m и μ_r соответственно. Обозначим θ_{ij} - вероятность перехода между узлами СеМО, при этом вероятность покинуть систему, находясь в узле i , равна β_i , $\beta_i + \theta_{i\bullet} = 1$, $i, j = \overline{1, N}$.

При поступлении г-заявки на узел $N+1$, заявка принимается на обслуживание, если для ее обслуживания в узле имеется достаточное количество приборов. В противном случае заявка не теряется, а переходит с вероятностью $\gamma_i = \frac{\alpha_i}{\alpha_{\bullet}}$ на обслуживание на узел i , при этом тип заявки не изменяется.

Заявка, перешедшая на узел i , обслуживается на нем до тех пор, пока, либо не закончится время обслуживания, и она не покинет СеМО, либо не будет высвобождено достаточное количество приборов для ее обслуживания на узле $N+1$, либо она не окончит обслуживание на узле i и с вероятностью θ_{ij} не перейдет на узел j , $i \neq j$.

Проиллюстрируем это на физической модели. Голосовой вызов, не принятый на обслуживание в макросоту вследствие недостатка ШПП, будет обслуживаться на уровне микросот, пока в макросоте не освободится достаточная для обслуживания ШПП; до этого момента вызов будет перемещаться между микросотами. В случае, если в микросоте i недостаточно ШПП для обслуживания, вызов теряется.

При поступлении м-заявки на узел i , $i = \overline{1, N}$, она принимается на обслуживание, если имеется достаточное количество приборов. В противном случае заявка переходит на узел $N+1$, соответствующий уровню иерархии макросот, если в нем имеется достаточное для обслуживания м-заявки количество приборов, при этом тип заявки не меняется; в противном случае, заявка теряется.



⊗- переход возможен при условии достаточного для обслуживания количества свободных приборов

Рис. 3 Модель функционирования сети

Для построенной СеМО вводится марковский процесс в стационарном режиме, для которого находятся система уравнений равновесия и ее решение. Найденное решение позволяет получить ряд вероятностно-временных характеристик, в том числе: интегральную вероятность блокировки м- и г-заявки, среднее число занятых приборов в сети, количество заявок на несоответствующих типу узлах, вероятность нахождения на

узлах заявок соответствующего типа, среднее число заявок на узле, среднее число заявок в сети, вероятность простоя узла, вероятность простоя сети, и провести соответствующее численное исследование.

Литература

1. Lee S.-H., Lim J.-S. Performance analysis of channel allocation schemes for supporting multimedia traffic in hierarchical cellular system // IEICE Trans. Commun., 2003/04, v.E86-B, No.4. – Pp. 1274-1285
2. Jabbari B., Fuhrmann W.F. Teletraffic Modeling and analysis of flexible hierarchical cellular networks with speed-sensitive handoff strategy // IEEE J. SAC, 1997/10, v.15, No.8. – Pp. 1539-1548
3. Boggia G., Camarda P., Di Fonzo N. Teletraffic Analysis of hierarchical cellular communication networks // IEEE Trans. Vehicular Tech., 2003, Vol.52, No.4. – Pp. 931- 946
4. Basharin G.P. Merkulov V.E. Performance Analysis of a Two-Layer Cellular Mobile Communication System Fragment // Вестник РУДН. Прикладная математика и информатика, 2002, Т.1, № 1. – С.14-24.
5. Башарин Г.П., Меркулов В.Е. Анализ пропускной способности в иерархических сетях сотовой связи // Электросвязь, 2003, №4. - С.45-47.

MODEL OF TWO LEVEL CELLULAR NETWORK WITH MULTIPLE TYPE CALLS

Abstract. We investigate a 3G two level network cluster with macro and micro cells. There are two types of calls: multimedia and voice. The queueing network model is developed. The states of the queueing network are described by a complex vector and the exact Markov process is defined over it. We derive combined equations in steady state. With the system equations solution we derive different performance characteristics.



СТРУКТУРА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ АЛГОРИТМОВ ОБРАБОТКИ ПАКЕТОВ IP-ТЕЛЕФОНИИ НА КЛАССИФИКАТОРЕ С НЕЙРОННОЙ СЕТЬЮ

Архангельский А.А., Рябоволик Б.Б.

СПбГУТ, Россия, Санкт-Петербург

Структура программного обеспечения, выполняющего классификацию пакетов IP-телефонии на основе адаптивной нейронной сети, включает три основных функциональных блока: интерфейс пользователя, блок предварительной обработки статистики, нейронную сеть и блок представления и обработки данных, полученных с выходов нейронной сети. Интерфейс пользователя дает возможность устанавливать различные параметры и режимы работы нейронной сети и обеспечивает представление выходной информации в виде графиков и таблиц. Блок предварительной обработки статистики производит считывание статистики с буферных накопителей и преобразует ее в вид пригодный для обработки в нейронной сети. Нейронная сеть выполняет обработку информации и передает результат на блок представления и обработки полученных данных, который преобразует полученные данные с выхода нейронной сети в управляющие сигналы.

Программное обеспечение разработано на языке программирования Delphi и состоит из следующих программных модулей:

- NEURONET.dll
- SetNet.pas
- StatUnit.pas
- TchUnit.pas
- WorkUnit.pas
- GrafUnit.pas

Все функции программного продукта реализованы в виде классов, нейронная сеть реализована в виде динамической библиотеки (NEURONET.dll). Функции основных программных модулей приведены в табл.1.

Система работает в двух основных режимах: немедленной передачи пакетов IP-телефонии сразу после классификации в одном буферном накопителе; передачи всех пакетов IP-телефонии после классификации во всех буферных накопителях. Первый режим используется в тех случаях, когда количество каналов, по которым приходят пакеты IP-телефонии ограничено и известно, второй режим используется при отсутствии информации о таких направлениях.

Данный программный продукт можно использовать, в качестве классификатора, выполняющего обработку пакетов IP телефонии на граничном маршрутизаторе.

Таблица 1.

Модуль программного продукта	Функции	Основные классы реализующие данные функции
NEURONET.dll	Создание нейронной сети	TEvent
SetNet.pas	Установка нейронной сети в начальное положение	TSet_Net
StatUnit.pas	Обработка статистики	TStat_proc, TStat_calc
TchUnit.pas	Обучение нейронной сети	TTeach
WorkUnit.pas	Классификация пакетов на основе обученной нейронной сети	TProc_start, TAction, TFunc
GrafUnit.pas	Представление выходной информации в графическом виде	TGraf

ОПТИМИЗАЦИЯ ЭФФЕКТИВНОСТИ СЕТЕЙ СОТОВОЙ СВЯЗИ С ПОМОЩЬЮ УСОВЕРШЕНСТВОВАННОГО МЕТОДА ЦЕНТРАЛИЗОВАННОГО УПРАВЛЕНИЯ МОЩНОСТЬЮ ПЕРЕДАТЧИКОВ

Попова Л., Герстакер В., Кох В.

Баварский Университет им. Фридриха-Александра,
Эрланген - Нюрнберг, кафедра мобильной связи

Работа посвящена проблеме поиска оптимального метода распределения выходной мощности передатчиков в сетях сотовой связи в целях увеличения теоретического предела емкости вышеназванных сетей. Поиск осуществляется при следующих ограничениях: а) отношение полезного сигнала, принятого мобильным терминалом (МТ) от обслуживающей его базовой станции (БС), к уровню интерференции, поступающей от других базовых станций, предписано заранее и является предельным с точки зрения рекомендуемого стандартом [1] минимума; это отношение принято за основной критерий, характеризующий качество соединения, б) расход мощности на выходах БС должен быть минимальным.

Для решения этой проблемы разработан алгоритм, в основе которого лежит метод централизованной адаптации выходной мощности базовых станций с использованием нового критерия распределения МТ в сети.

Представленные в работе обнадеживающие результаты были получены моделированием на основе Radio Network Simulator RUNE (Rudimentary Network Emulator, Ericsson Research), интегрированного в MATLAB.

Введение

Проблема оценки емкости сетей мобильной связи существует с момента введения в эксплуатацию сетей подвижной связи 1-го поколения. В последние годы стремительного роста мобильных сетей исследователи видят решение этой проблемы в нахождении оптимального алгоритма по регулированию выходной мощности передатчиков (Power Control Process).

Задачей Power Control Process является адаптация мощности передатчиков БС и МТ к условиям радиополя таким образом, чтобы, с одной стороны, за счет достаточно высокой мощности обеспечить надежную связь и, с другой стороны, за счет минимально возможной мощности обеспечить щадящий режим расходования ограниченной емкости батареи МТ и минимизацию уровня интерференции в сети, увеличивая таким образом количество пользователей с допустимым качеством соединения, а следовательно суммарную емкость сети.

При этом надо четко понимать, что уменьшение уровня мощности передатчика для определенного соединения, делает именно этот трафиковый канал (Traffic Channel) еще более уязвимым к интерференции. Поэтому нахождение оптимального алгоритма для регулирования уровня мощности передатчиков не является тривиальной задачей.

Разработка и реализация алгоритма

В данной работе предложен алгоритм централизованного осуществления Power Control Process на основе метода приближенных решений, базирующегося на текущих измерениях соотношения полезного сигнала к уровню интерференции (C/I) для каждого МТ одновременно. Алгоритм обладает рядом существенных преимуществ по сравнению с традиционными методами. Уровень C/I отражает количество информации-

ных блоков, принятых с ошибкой BLER (Block Error Rate), для анализа контрольных каналов связи (Common Control Channel) и количество удаленных фреймов FER (Frame Erasure Rate), характеризующих состояние трафиковых каналов. Именно BLER служит одним из важных оценочных параметров качества соединения (QoS).

Предъявляемые в данной задаче требования можно записать в виде системы неравенств:

$$(C/I)_i = \frac{G_{ii}P_i}{\sum_{\substack{j \\ j \neq i}} G_{ij}P_j + N_i} \geq (C/I)_{\min}, \quad i = 1, \dots, n,$$

где n – номер МТ, P_i – выходная мощность i -ой БС и N_i – уровень термального шума на приемном конце МТ. В данной работе было рассмотрено одностороннее соединение «вниз» (БС – МТ). Каждое активное соединения (МТ-БС) в сети характеризуется коэффициентом G_{ij} , показывающим затухание полезного сигнала от i -ой БС в направлении обслуживаемого ею i -го МТ при $i=j$. Поскольку излучаемый i -ой БС радиосигнал является полезным сигналом только для его адресата, т.е. i -го МТ, то для прочих элементов сети, использующих этот же трафиковый канал, он одновременно является помехой и порождает источник нежелательной интерференции в сети.

Таким образом, мощность получаемого i -ым МТ полезного сигнала будет равна $G_{ii}P_i$ и уровень нежелательной интерференции $G_{ij}P_j, i \neq j$.

Основная проблема, возникающая при поиске максимального предела емкости сети, заключается в правильном выборе оптимального критерия распределения в сети МТ. Определив его, мы получаем в качестве искомого результата вектор выходной мощности с положительными компонентами $P \geq 0$, который максимизирует емкость системы (количество пользователей в сети) при заданных ограничениях. А именно, отношение полезного сигнала к уровню соконалной интерференции для каждого активного соединения (МТ-БС) в сети должно быть равно заранее определенному минимуму, C/I_{\min} .

Заметим, что для удобства аналитического анализа в работе была рассмотрена модель, в которой в каждой соте в каждый момент времени имело место только одно активное соединение. Таким образом, нас интересовала только соконалная интерференция. Заданная выше система неравенств может быть записана в следующем виде: $ZP=N, P$ – вектор выходной мощности $P=[P_1 \dots P_N]^T$, N – вектор нормированного шума, определенный как $N = \begin{bmatrix} N_1 & N_2 & \dots & N_n \\ G_{11} & G_{22} & \dots & G_{nn} \end{bmatrix}^T$, и Z – нормированная неотрицательная квадратная $n \times n$ матрица случайных переменных. Элементами матрицы являются:

$$Z_{ij} = \begin{cases} 1 \\ (C/I)_{\min} \\ -\frac{G_{ij}}{G_{ii}}, \end{cases} \quad \begin{cases} \text{если } i = j, \\ \\ \text{если } i \neq j. \end{cases}$$

В качестве индикатора емкости сети мы использовали вероятность интерференции (т.е., вероятность того, что соединение будет прервано из-за плохого качества соединительной линии). Подробное описание модели дано в [2].

Предложенный в работе рекурсивный алгоритм состоит из нескольких шагов:

1. Из теории положительно определенных матриц, элементами которых являются случайные переменные, известно, что матрица Z имеет полный ранг с вероятностью равной единице [3]. Поэтому, обращая матрицу Z , определяем вектор мощности P , т.о., $P=Z^{-1} N$.
2. Алгоритм проверяет, все ли элементы вектора мощности являются положительными числами. Если это так, он останавливается, в противном случае программа удаляет k -ое активное соединение (индекс k пояснен ниже); как следствие мы получаем матрицу Z' уменьшенного размера $(n-1) \times (n-1)$.
3. Матрица Z заменяется матрицей Z' , $Z=Z'$ и программа возвращается к первому шагу.
4. Работа алгоритма останавливается, когда все элементы, содержащиеся в векторе P , станут положительными.

Индекс k соответствует максимальной сумме $s_i = r_i + c_i$, состоящей из суммы строк r_i , и суммы столбцов c_i , $i = 1, \dots, n$ матрицы \mathbf{Z} и является критерием для удаления того или иного активного соединения, $k = \arg \max_i s_i$.

Сумма элементов i -ой строки, исключая элементы главной диагонали, отражает суммарную интерференцию, которую испытывает i -ый МТ; сумма элементов i -го столбца показывает в свою очередь уровень интерференции, причиной которой является i -ый МТ.

Таким образом, нахождение максимальной суммы строк, а также максимальной суммы столбцов позволяет нам выявить наиболее сильно интерферирующего либо интерферируемого (страдающего от интерференционных помех) пользователя и удалить его из активного соединения. Итеративный алгоритм, удалив минимально возможное количество соединений, находит максимально возможную матрицу \mathbf{Z} , где для оставшихся активных соединений (МТ-БС) отношение полезного сигнала к уровню соканальной интерференции равно C/I_{\min} .

Алгоритм обладает хорошей сходимостью и обычно ограничивается удалением одного соединения.

В итоге имеем вектор \mathbf{P} с положительными элементами, каждый из которых соответствует уровню мощности на выходе БС в «спасенных» активных соединениях.

Численные результаты и выводы

Результаты моделирования представлены в виде графиков для четырех различных C/I_{\min} : 8, 10, 13 и 17 dB.

На рис. 1 изображена вероятность прерывания соединения из-за плохого качества связи на примере двух алгоритмов. Предложенный в работе алгоритм централизованной адаптации выходной мощности БС с использованием нового критерия распределения МТ (novel algorithm) в сети была сравнена с традиционным алгоритмом регулирования мощности БС (eigenvalue-based PC) [4].

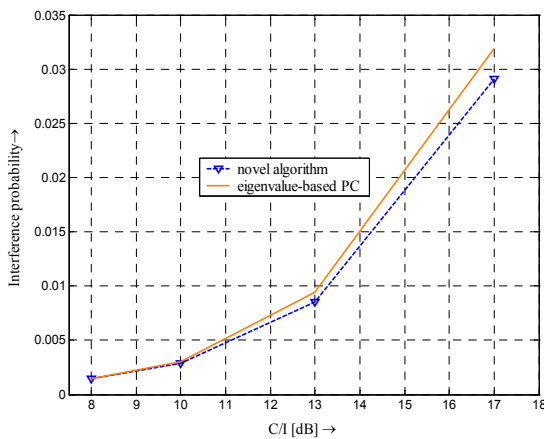


Рис. 1. Сравнение вероятности прерывания соединения на примере двух алгоритмов

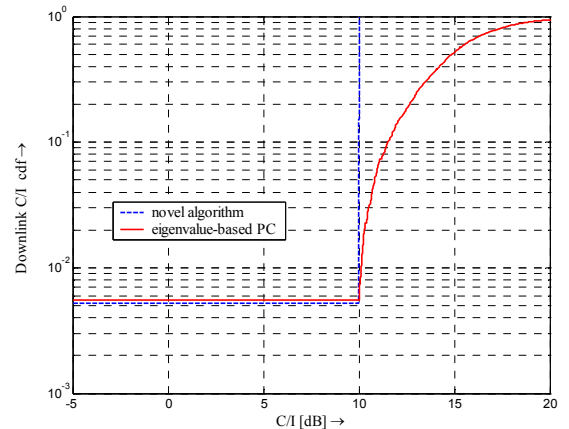


Рис. 2. Функция распределения (cdf) C/I , $(C/I)_{\min}=10$ dB

Численные результаты наглядно демонстрируют преимущества предложенного в работе алгоритма, он позволяет значительно понизить уровень соканальной интерференции в сети.

Кроме того, по сравнению с [4], где основной целью являлось достижение максимально возможного C/I , предложенный в работе рекурсивный алгоритм всегда находит минимальный C/I для всех активных соединений (см. рис. 2), что является более релевантным с практической точки зрения, т.е., рассматриваемый нами алгоритм позволяет избежать излишнего расхода мощности на выходе БС.

Литература

1. Specifications of Third Generation Mobile System, 3GPP GSM PH2 04.13.
2. Попова Л., „Метод управления мощностью передатчиков для оптимизации соотношения емкости сети мобильной связи и качества обслуживания абонентов“, 59-я Научная сессия посвященная дню радио, 2004.
3. R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1985.
4. J. Zander, "Performance of optimum transmitter power control in cellular radio systems", IEEE Trans. of Veh. Technol., vol. 41, no. 1, Feb. 1992.

IMPROVED PERFORMANCE BOUND FOR OPTIMAL CENTRALIZED POWER CONTROL IN CELLULAR MOBILE RADIO SYSTEMS

Popova L., Gerstacker W., Koch W.

Institute for Mobile Communications
University of Erlangen-Nürnberg

This paper deals with the transmit power allocation problem for cellular wireless networks. In order to improve the estimation of the theoretical optimum performance in cellular radio systems meaning maximum capacity for given constraints on the carrier-to-interference ratio and minimum energy consumption of the mobile station, an iterative centralized power control scheme with novel criterion for user allocation is presented and promising results are shown.

All the analysis presented in this paper are based on simulation results using the Radio Network Simulator RUNE (Rudimentary Network Emulator, developed by Ericsson Research) and implemented in MATLAB.

Keywords—mobile communications; cochannel interference; capacity improvement; power control; optimum criterion.



УСТОЙЧИВЫЕ ПРОЦЕДУРЫ ОЦЕНКИ ОБЪЕМА ТЕЛЕКОММУНИКАЦИОННОГО РЕСУРСА, ДОСТАТОЧНОГО ДЛЯ ОБСЛУЖИВАНИЯ ИЗВЕСТНОГО ПОТОКА ЗАЯВОК С ЗАДАННЫМ КАЧЕСТВОМ

Лагутин В.С.¹, Костров В.О.¹, Степанов С.Н.¹, Iversen V.B.²

¹ЗАО «Система Телеком», ²COM Center, Technical University of Denmark

Введение

Одной из важнейших задач, которую приходится решать администрации сетей связи, является определение величины телекоммуникационного ресурса, достаточного для обслуживания известного потока заявок с заданными (нормированными) характеристиками обслуживания. Обычно под телекоммуникационным ресурсом понимается каналный ресурс, выраженный: числом соединительных линий; битовой скоростью, представленной через целое число основных передаточных единиц; количеством радиочастот, используемых для передачи; числом временных слотов и т.д. В других постановках это может быть величина буфера, число операторов, количество управляющих устройств и т.д. Достаточность ресурса оценивается сравнением характеристик качества обслуживания с их нормированными показателями, которые задаются соответствующими регламентирующими документами. Обычно в качестве характеристики используется доля отказов в общем числе обращений пользователей к предоставлению телекоммуникационного ресурса. Но в зависимости от постановки задачи это может быть и время задержки или какой-то иной интегральный показатель, учитывающий, например, стоимостные аспекты обслуживания абонентов.

Какова общая схема решения сформулированной проблемы? Вначале определяются показатели обслуживания абонента, используемые для оценки достаточности телекоммуникационного ресурса. Затем с привлечением понятий теории вероятностей на основе функциональной схемы исследуемой системы связи или ее фрагмента строится математическая модель, которая обеспечивает оценку требуемых характеристик. Для нахождения численных значений характеристик разрабатываются точные и приближенные расчетные алгоритмы. Эта часть задачи решается с помощью основных и специальных методов вычислительной математики. На заключительном шаге методом перебора находится искомое значение объема телекоммуникационного ресурса.

Важность решения сформулированной проблемы для оптимального распределения телекоммуникационного ресурса современных мультисервисных сетей связи заставляет искать способы уменьшения объема вычислительных работ при нахождении ответа [1-4]. Изложение соответствующих возможностей будет показано на примере модели звена мультисервисной сети связи с произвольным числом пуассоновских потоков нагрузки. Потоки различаются интенсивностью поступления заявок, числом единиц используемого ресурса и временем его удерживания при обслуживании одной заявки.

Основной результат

Рассмотрим модель сети, в которой имеется одно звено со скоростью k основных передаточных единиц. Предположим, что звено обслуживает заявки на предоставление каналной емкости от n групп пользователей. Будем считать, что процесс поступления заявок от каждой группы пользователей подчиняется закону

Пуассона с интенсивностью λ_s , $s = 1, \dots, n$. Сообщение s -го потока использует b_s единиц ресурса линии в течение времени соединения. Без потери общности будем считать, что время удерживания ресурса имеет экспоненциальное распределение с параметром равным единице. Известно, что для данной модели расчетные алгоритмы остаются теми же и в случае произвольного распределения функции времени обслуживания, если среднее значение времени обслуживания фиксировано. Более того, среднее значение времени удерживания ресурса для каждого потока может быть и индивидуальным.

Обозначим через $i_s(t)$ число сообщений s -го потока, находящихся на обслуживании в момент времени t . Построенная модель описывается n -мерным марковским процессом с компонентами $r(t) = \{i_1(t), i_2(t), \dots, i_n(t)\}$, определенном на пространстве состояний S , состоящем из векторов (i_1, \dots, i_n) , где i_s - число сообщений s -го потока, находящихся в стационарном состоянии на обслуживании. Пространство состояний S определено следующим образом $(i_1, \dots, i_n) \in S$, $i_s \geq 0$, $s = 1, \dots, n$, $\sum_{s=1}^n i_s b_s \leq k$.

Обозначим через $P(i_1, \dots, i_n)$ ненормированные значения стационарных вероятностей $r(t)$, а через $p(i_1, \dots, i_n)$ их нормированные значения. После нормировки вероятность $p(i_1, \dots, i_n)$ получает интерпретацию доли времени пребывания модели в состоянии, когда установлены i_1 соединений для первого потока заявок, i_2 соединений для второго потока заявок и т.д. Обозначим для состояния (i_1, \dots, i_n) через i общее число занятых единиц ресурса $i = i_1 b_1 + \dots + i_n b_n$. Процесс обслуживания сообщений s -го потока характеризуется вероятностью отказа в установлении требуемого соединения π_s , $s = 1, \dots, n$.

$$\pi_s = \sum_{i+b_s > k} p(i_1, \dots, i_n). \quad (1)$$

Наиболее эффективная схема вычисления введенных стационарных характеристик была получена в работах [5]. Этот результат позднее был повторен в [6], [7]. Алгоритм основан на возможности оценки характеристик (1) с помощью вероятностей $p(i)$ пребывания модели в состоянии, когда заняты ровно i передаточных единиц

$$p(i) = \sum_{i_1 b_1 + \dots + i_n b_n = i} p(i_1, \dots, i_n).$$

Соответствующие расчетные формулы имеют вид

$$\pi_s = \sum_{i=k-b_s+1}^k p(i), \quad s = 1, 2, \dots, n. \quad (2)$$

Ненормированные значения $P(i)$ могут быть найдены из рекурсии

$$P(i) = \frac{1}{i} \sum_{s=1}^n \lambda_s b_s P(i - b_s), \quad i = 1, 2, \dots, k, \quad (3)$$

где обычно $P(0) = 1$. Будем использовать нижний индекс k для обозначения числа передаточных единиц, при которых был произведен расчет нормированных значений соответствующих вероятностей. Тогда из (3) нетрудно получить

$$p_k(k) = \frac{\frac{1}{k} \sum_{s=1}^n \lambda_s b_s p_{k-1}(k - b_s)}{1 + \frac{1}{k} \sum_{s=1}^n \lambda_s b_s p_{k-1}(k - b_s)}. \quad (4)$$

Алгоритм определения величины достаточности канального ресурса звена, оцениваемый с помощью показателей качества обслуживания нагрузки (2), который на каждом шаге своей реализации оперирует только с нормированными значениями стационарных вероятностей, имеет следующий вид:

Шаг 1. Положим $p_0(0) = 1$.

Шаг 2. Пусть $b = \max(b_s)$. Для фиксированных $k = 1, 2, \dots$ найти нормированные значения $p_k(i)$, $i = 0 \leq i \leq k$,

$k, k-1, \dots, \max(k-b, 0)$, используя соотношения (4) и условие нормировки

$$p_k(k) = \frac{\frac{1}{k} \sum_{s=1}^n \lambda_s b_s p_{k-1}(k-b_s)}{1 + \frac{1}{k} \sum_{s=1}^n \lambda_s b_s p_{k-1}(k-b_s)},$$

$$p_k(i) = \frac{p_{k-1}(i)}{1 + \frac{1}{k} \sum_{s=1}^n \lambda_s b_s p_{k-1}(k-b_s)}, \quad i = k-1, k-2, \dots, \max(k-b_s, 0).$$

Шаг 3. Находятся значения характеристик, определяемые (2), проверяется условие достаточности ресурса и в случае его невыполнения значение канального ресурса увеличивается на единицу и расчеты повторяются заново. Использование построенной версии рекурсивного алгоритма требует хранения в памяти ЭВМ массивов объема $O\{b\}$. При этом необходимо выполнить число операций, оцениваемое величиной $O\{(n+b)k\}$.

Заключение

В работе описана общая схема построения стабильных вычислительных алгоритмов оценки объема телекоммуникационного ресурса, достаточного для обслуживания известного потока нагрузки с заданным качеством. Полученные результаты основаны на идеях [3] и обобщают для случая многопоточковых моделей известную рекуррентную схему вычисления значения формулы Эрланга

$$E_k(\lambda) = \frac{\frac{\lambda}{k} E_{k-1}(\lambda)}{1 + \frac{\lambda}{k} E_{k-1}(\lambda)}, \quad k = 1, 2, \dots, \quad E_0(\lambda) = 1, \quad \text{где } E_k(\lambda) - \text{формула Эрланга для числа обслуживающих}$$

устройств k и интенсивности входной нагрузки λ , выраженной в Эрлангах.

Литература

1. Лагутин В.С., Костров В.О., Степанов С.Н. Повышение эффективности использования цифровых линий на сетях связи. // Документальная электросвязь. 2001. № 6, с.7-14.
2. Лагутин В.С., Степанов С.Н. Телетрафик мультисервисных сетей связи. М.: Радио и связь, 2000. - 320 с.
3. Stepanov, S.N., Iversen, V.B., Kostrov, V.O. Optimised Dimensioning of Large Bandwidth Resources Under MPLS Technology. Proc. of St. Petersburg Regional International Teletraffic Seminar on Telecommunication Network and Teletraffic Theory. St. Petersburg, Russia. January 2002, pp. 50-63.
4. Nilsson, A., Perry, M., Gersht, A., Iversen, V.B. On Multi-Rate Erlang-B Computations. ITC 16, Sixteenth International Teletraffic Congress, Edinburgh, UK, June 1999. Elsevier 1999. pp. 1051-1060.
5. Fortet, R., Grandjean, Ch. Congestion in a Loss System when Some Calls want Several Devices Simultaneously. Electrical Communications, Vol. 39 (1964):4, 513-526. Paper presented at ITC-4, Fourth International Teletraffic Congress, London, UK, 15-21 July 1964.
6. Kaufman, J.S. Blocking in a Shared Resource Environment. IEEE Transactions on Communications, Vol. COM-29 (1981):10, pp. 1474-1481.
7. Roberts, J.W. A Service System with Heterogeneous User Requirements - Applications to Multi-Service Telecommunication Systems. Pages 423-431 in Performance of Data Communication Systems and their Applications. G. Pujolle (editor), North-Holland Publ. Co. 1981.

THE STABLE PROCEDURES OF DIMENSIONING TELECOMMUNICATION SYSTEMS WITH PRESCRIBED LEVEL OF SERVICE AND GIVEN INPUT TRAFFIC FLOWS

The unified approach for teletraffic models to convert recursions for global state probabilities into stable form is derived. At each step of recurrence we deal with normalised values of global state probabilities used for estimation of main stationary performance measures. This makes it very well suited for solving problems of dimensioning. In addition to being numerically stable, the main positive features of the suggested approach are its simplicity and small storage and computational requirements.



МОДЕЛЬ ПРИЕМА ЗАЯВОК НА ВЫДЕЛЕНИЕ КАНАЛЬНОГО РЕСУРСА ДЛЯ ОТДЕЛЬНОЙ СОТЫ СЕТИ ПОДВИЖНОЙ СВЯЗИ С УЧЕТОМ ЧИСЛА ПОЛЬЗОВАТЕЛЕЙ, НАХОДЯЩИХСЯ НА ОБСЛУЖИВАНИИ

Введение

Основной задачей систем подвижной связи 3-го поколения является создание технических условий для предоставления мультисервисных услуг пользователю с помощью мобильного терминала, имеющего единый номер. Мультисервисный характер нагрузки обусловлен широким спектром услуг, которые сеть подвижной связи будет предоставлять пользователю уже в ближайшем будущем. Только таким образом оператор может решить проблему уменьшения показателя доходности на абонента ARPU, характерную для современных сетей подвижной связи. Предполагаемый рост объемов трафика потребовал разработки новых стандартов, позволяющих значительно повысить скорость передачи информации. В результате проведенных исследований для систем подвижной связи 3-го поколения основным технологическим принципом, положенным в основу стандарта связи, было выбрано кодовое разделение каналов связи (CDMA), позволяющее, в отличие от временного разделения каналов TDMA (основа стандарта GSM), более эффективно использовать имеющийся спектр частот и увеличить пропускную способность сотовых сетей. Международный союз электросвязи принял 5 стандартов 3G интерфейса. Наиболее перспективными являются WCDMA и CDMA2000 (IMT-MC). Оба стандарта основаны на принципе широкополосного множественного доступа с кодовым разделением каналов. Европейская схема миграции с GSM на WCDMA, носит название UMTS [1-7].

Особенностью кодового разделения каналов является возможность многим пользователям вести передачу в одной и той же полосе частот. Частотный канал разбивается на несколько десятков логических каналов, передача и приём по которым осуществляется в одной полосе частот и одновременно. Теоретически число пользователей ограничивается числом кодовых комбинаций и может быть очень большим. Практически оно ограничивается уровнем помех, который возрастает с ростом числа пользователей, находящихся на обслуживании в соте. Каждый раз, когда происходит прием заявки от нового абонента, отношение сигнал/шум для остальных пользователей, находящихся на обслуживании в данной соте, уменьшается. Поскольку качество связи не должно быть ниже фиксированного уровня, то с ростом числа пользователей, находящихся на обслуживании в исследуемой соте, уменьшается вероятность приема на обслуживание нового абонента. При проведении математического моделирования рассматриваемого процесса можно использовать модели с внутренними блокировками, которые активно применялись в теории телетрафика при анализе неполнодоступных схем. В данной работе будет построена математическая модель процесса приема заявок на выделение канального ресурса для отдельной соты сети подвижной связи при наличии внутренней блокировки, обусловленной учетом числа обслуживаемых пользователей, и рассмотрена схема оценки основных характеристик способности соты к приему абонента к обслуживанию.

Описание модели и алгоритм оценки характеристик

Допустим, что в отсутствие шума максимальная пропускная способность соты задается целочисленным параметром v , где единичное значение ресурса - это минимальное значение пропускной способности соты, выделяемое одному пользователю. Обычно в качестве такового выступает ресурс, необходимый для формирования одного речевого канала. Предположим, что сота обслуживает заявки на предоставление канального ресурса от n групп пользователей. Допустим, что процесс поступления заявок от каждой группы пользователей подчиняется закону Пуассона с интенсивностью λ_k , $k = 1, \dots, n$. Сообщение k -го потока использует b_k единиц пропускной способности соты в течение времени соединения. С тем, чтобы описать динамику изменения состояний модели с течением времени, будем предполагать, что время удерживания канального ресурса имеет экспоненциальное распределение со средним значением $1/\mu_k$.

В дополнение к приведенному выше стандартному описанию процесса функционирования звена мультисервисной сети связи предположим, что заявка k -го потока ($k = 1, 2, \dots, n$), поступившая в то состояние модели, когда ровно i единиц канальной емкости занято на обслуживание с вероятностью $\varphi_{k,i}$ получит отказ в выделении канальной емкости. Предположим, что существует целое значение $b > 0$ такое, что для всех $k = 1, 2, \dots, n$, $\varphi_{k,i} = 0$ для $i = 0, 1, \dots, v-b$. Выбор конкретных значений вероятностей $0 \leq \varphi_{k,i} \leq 1$ позволяет учесть разные формы зависимости приема заявок на выделение канального ресурса от числа единиц канального ресурса уже занятого передачей информационных сообщений. Введение показателя b объясняется естественным предположением о наличии уровня занятости канального ресурса величины b , до достижения которого влиянием взаимных помех можно пренебречь.

Построенная модель описывает процесс занятия каналов в одной соте. Ее можно обобщить и на случай сети с учетом возможности перехода абонентов из соты в соту. Кроме того, введенная модель может быть использована для анализа различных схем резервирования канального ресурса, зависящего от числа занятых единиц канального ресурса сообщениями всех видов в момент поступления рассматриваемой заявки.

Обозначим через $i_k(t)$ число сообщений k -го потока, находящихся на обслуживании в момент времени t . Построенная модель описывается n -мерным марковским процессом с компонентами $r(t) = \{i_1(t), i_2(t), \dots,$

$i_n(t)$, определенном на пространстве состояний S , состоящем из векторов (i_1, \dots, i_n) , где i_k - число сообщений k -го потока, находящихся в стационарном состоянии на обслуживании. Пространство состояний S состоит из векторов (i_1, \dots, i_n) с целочисленными компонентами i_1, \dots, i_n , в совокупности удовлетворяющие неравенствам $(i_1, \dots, i_n) \in S, i_k \geq 0, k = 1, \dots, n, \sum_{k=1}^n i_k b_k \leq v$.

Процесс обслуживания поступающих заявок k -го потока на выделение канального ресурса оценивается долей отказов π_k , определяемой из соотношения

$$\pi_k = \sum_{i=v-b+1}^v p(i) \varphi_{k,i}, \quad k = 1, 2, \dots, n, \quad (1)$$

где

$$p(i) = \sum_{i_1 b_1 + \dots + i_n b_n = i} p(i_1, \dots, i_n)$$

доля времени пребывания модели в состоянии, когда ровно i единиц канального ресурса занято на обслуживание.

Для вычисления введенной характеристики можно составить и решить систему уравнений статистического равновесия. Делается это с помощью известных стандартных средств теории телетрафика. Однако увеличение размерности пространства состояний позволяет выполнить указанную процедуру только для небольших значений n (1-3) и v (до 10). Увеличение данных параметров сверх указанных пределов приводит вначале к большим вычислительным сложностям, а затем и к невозможности решить поставленную задачу. В силу указанных причин, единственным реальным способом оценить значения введенных показателей приема заявок к обслуживанию остается использование достаточно обоснованных приближенных алгоритмов.

Для оценки значений $p(i)$ предлагается использовать следующую рекурсивную формулу

$$P(i) = \begin{cases} 0 & \text{для } i < 0 \\ a & \text{для } i = 0 \\ \frac{1}{i} \sum_{k=1}^n \frac{\lambda_k}{\mu_k} b_k (1 - \varphi_{k,i-b_k}) P(i - b_k) & \text{для } i = 1, 2, \dots, v \end{cases} \quad (2)$$

Реализация соотношений (2) при определении $P(i)$ не вызывает затруднений. Истинные значения $P(i)$ находятся после нормировки. Пусть $a_k = \frac{\lambda_k}{\mu_k}$, интенсивность предложенной нагрузки, выраженная в Эрландах.

Перечислим шаги рекурсивного алгоритма.

1. Положим значение $P(0) = 1$.

2. Для i , меняющихся последовательно от 1 до v , с помощью (2) находим ненормированные значения вероятностей $P(i)$

$$P(i) = \frac{1}{i} \sum_{k=1}^n a_k b_k P(i - b_k) (1 - \varphi_{k,i-b_k}).$$

3. Вычисляем значение нормировочной константы

$$N = \sum_{i=0}^v P(i).$$

4. Определяем нормированные значения вероятностей $P(i)$

$$P(i) = \frac{P(i)}{N}, \quad i = 0, 1, \dots, v.$$

Находим значения характеристик качества передачи каждого из n информационных потоков

$$\pi_k = \sum_{i=v-b+1}^v p(i) \varphi_{k,i}, \quad k = 1, \dots, n.$$

Заключение

В работе построена модель приема заявок на выделение канального ресурса для отдельной соты сети подвижной связи. В модели учитывается рост вероятности отказа с увеличением числа пользователей, уже находящихся на обслуживании в данной соте. Построен приближенный алгоритм оценки доли заявок, получивших отказ.

Литература

1. Лагутин В.С., Степанов С.Н. Телетрафик мультисервисных сетей связи. М.: Радио и связь, 2000. - 320 с.
2. Iversen V.B., Benetis V., Ha N.T., Stepanov S. Evaluation of Multi-service CDMA Networks with Soft Blocking // Proc. 16th ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems. University of Antwerp. Antwerp. Belgium. August 31-September 02. 2004.
3. 3GPP TS 23.002 V5.5.0 (2002-01); 3rd Generation Partnership Project (3GPP); technical specification group services systems aspects; network architecture (Release 5)
4. 3GPP TR 23.107 (2003); Quality of Service (QoS) concept and architecture.
5. Grilo, A.; Nunes, M.; Sergio, G.; Ciulli, N. ; Integration of IP mobility and QoS for heterogeneous wireless access in MOICANE, Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on, Volume:1, 7-10 Sept. 2003, pp. 470-475.
6. Orlik, P.; Rappaport, S.; A model for teletraffic performance and channel holding time characterization in wireless cellular communication with general session and dwell time distributions, IEEE JSAC, 16(5):788-803, June 1998.
7. Orlik, P.; Rappaport, S.; Traffic performance and mobility modeling of cellular communications with mixed platforms and highly variable mobilities, Proc. of the IEEE, 86(7):1464-1479, July 1998.

THE MODEL OF CALL ADMISSION FOR SINGLE CELL WITH TAKING INTO ACCOUNT THE NUMBER OF SUBSCRIBERS BEING ON SERVICE

This paper presents an approach for calculation the capacity of multi-service CDMA networks with soft blocking using multirate teletraffic models with state dependent blocking probabilities. Performance calculations are carried out with help of effective recursive algorithm that is well suited for solving problems of dimensioning. Derived algorithm is very simple and easy in realization.

Сумма элементов i -ой строки, исключая элементы главной диагонали, отражает суммарную интерференцию, которую испытывает i -ый МТ; сумма элементов i -го столбца показывает в свою очередь уровень интерференции, причиной которой является i -ый МТ.

Таким образом, нахождение максимальной суммы строк, а также максимальной суммы столбцов позволяет нам выявить наиболее сильно интерферирующего либо интерферируемого (страдающего от интерференционных помех) пользователя и удалить его из активного соединения. Итеративный алгоритм, удалив минимально возможное количество соединений, находит максимально возможную матрицу Z' , где для оставшихся активных соединений (МТ-БС) отношение полезного сигнала к уровню сокопальной интерференции равно C/I_{\min} .

Алгоритм обладает хорошей сходимостью и обычно ограничивается удалением одного соединения.

В итоге имеем вектор P с положительными элементами, каждый из которых соответствует уровню мощности на выходе БС в «спасенных» активных соединениях.

Численные результаты и выводы

Результаты моделирования представлены в виде графиков для четырех различных C/I_{\min} : 8, 10, 13 и 17 dB.

На рис. 1 изображена вероятность прерывания соединения из-за плохого качества связи на примере двух алгоритмов. Предложенный в работе алгоритм централизованной адаптации выходной мощности БС с использованием нового критерия распределения МТ (novel algorithm) в сети была сравнена с традиционным алгоритмом регулирования мощности БС (eigenvalue-based PC) [4].

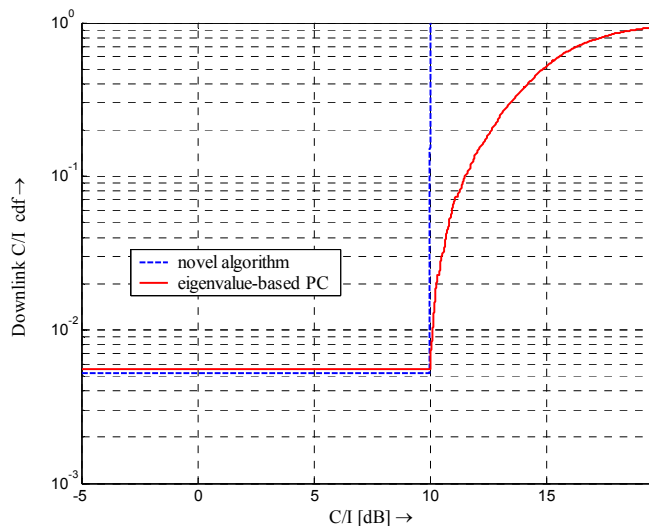
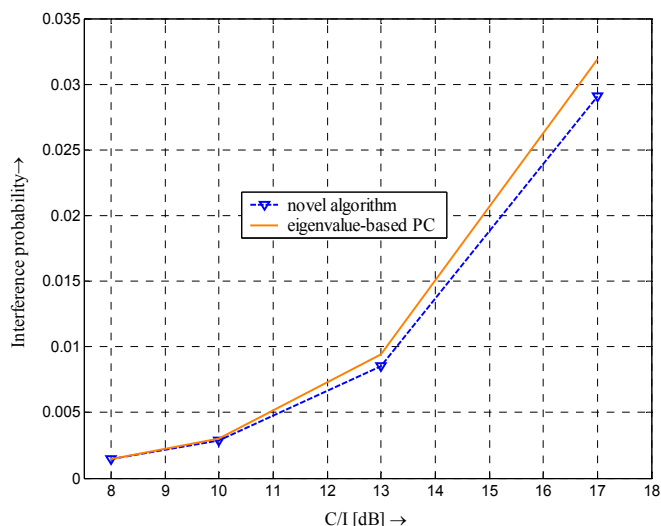


Рис.1. Сравнение вероятности прерывания соединения на примере двух алгоритмов

Рис.2. Функция распределения (cdf) C/I, $(C/I)_{\min}=10$ dB

Численные результаты наглядно демонстрируют преимущества предложенного в работе алгоритма, он позволяет значительно понизить уровень соканальной интерференции в сети.

Кроме того, по сравнению с [4], где основной целью являлось достижение максимально возможного C/I, предложенный в работе рекурсивный алгоритм всегда находит минимальный C/I для всех активных соединений (см. рис.2), что является более релевантным с практической точки зрения, т.е., рассматриваемый нами алгоритм позволяет избежать излишнего расхода мощности на выходе БС.

Литература

1. Specifications of Third Generation Mobile System, 3GPP GSM PH2 04.13.
2. Попова Л., „Метод управления мощностью передатчиков для оптимизации соотношения емкости сети мобильной связи и качества обслуживания абонентов“, 59-я Научная сессия посвященная дню радио, 2004.
3. R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1985.
4. J. Zander, "Performance of optimum transmitter power control in cellular radio systems", IEEE Trans. of Veh. Technol., vol. 41, no. 1, Feb. 1992.

IMPROVED PERFORMANCE BOUND FOR OPTIMAL CENTRALIZED POWER CONTROL IN CELLULAR MOBILE RADIO SYSTEMS

Popova L., Gerstacker W., Koch W.
Institute for Mobile Communications
University of Erlangen-Nürnberg

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ МУЛЬТИСЕРВИСНОЙ СЕТИ СВЯЗИ С УЧЕТОМ ПОВЕДЕНИЯ АБОНЕНТА ПОСЛЕ ОТКАЗА В ОБСЛУЖИВАНИИ

Кокина О.А.

ЗАО «Система Телеком»

Введение

Исследованием показателей обслуживания телефонных сетей занимаются многие специалисты в области телекоммуникаций. Особый интерес вызывает проблема повторных вызовов - изучение поведения абонента в условиях отказа в обслуживании. После получения отказа абонент повторяет вызов и тем самым создает дополнительную нагрузку на сеть. Это является основной причиной возникновения повторных вызовов. К другим причинам возникновения повторных вызовов можно отнести неисправности в телефонной сети: устаревание и износ оборудования, неотчет вызываемого абонента, ограниченное число телефонных станций, которое проектируется на определенное число одновременно поступающих вызовов. Повторные вызовы ухудшают качество обслуживания абонента из-за увеличения времени установления соединения, увеличивают время занятия коммутационного оборудования сети, увеличивают число соединений, не заканчивающихся разговором, что ведет к увеличению нагрузки сети и уменьшению дохода от эксплуатации. В последнее время проблема повторных вызовов приобрела более яркую окраску из-за усовершенствования технических устройств: появились новые цифровые АТС с программным управлением, а также телефонные

аппараты, позволяющие периодически набирать номер до установления соединения с абонентом (функция «автодозвон»).

Век информационных технологий принес с собой значительные изменения в телекоммуникациях - наряду с существующей коммутацией каналов в телефонных сетях появилась коммутация пакетов, которая позволила предложить абоненту ряд новых услуг, таких, к примеру, как: видеоконференции, мультимедиа, TV, видео по требованию. Телефонная сеть расширила спектр предоставляемых услуг и стала мультисервисной. Вследствие этого становится актуальным исследование моделей мультисервисных сетей [1-4]. В данной статье рассмотрена математическая модель мультисервисной сети связи с учетом поведения абонента после отказа в обслуживании. В модели учтены два источника возникновения повторных вызовов – занятость телефонной линии и неотвѣт вызываемого абонента.

Описание модели

Построим модель сети, состоящей из некоторого числа узлов, соединенных между собой J цифровыми линиями и имеющей n информационных потоков сообщений. Пусть v_j - скорость передачи сообщений j -ой цифровой линии, выраженная в единицах ресурса, требуемого для обслуживания поступающих сообщений, $j = 1, \dots, J$, b_k - число единиц ресурса, необходимого для передачи одного сообщения k -ого информационного потока, $k = 1, \dots, n$. Каждый из информационных потоков характеризуется поступлением сообщений двух различных потоков - первичных и повторных вызовов. Поток первичных вызовов поступает по закону Пуассона с интенсивностью λ_k . Получив отказ в обслуживании из-за недостаточности ресурса или неотвѣта вызываемого абонента, вызывающий абонент повторяет вызов через случайное время, имеющее экспоненциальное распределение с параметром μ_k . Поступающий первичный или повторный вызов занимает свободный ресурс на случайное время, распределенное экспоненциально с параметром α_k . Обозначим через $H_{1,k}$ и $H_{2,k}$ соответственно - вероятности повторения вызова для первого и всех последующих отказов в обслуживании из-за нехватки ресурса. Через $H_{s,k}$ обозначим вероятность повторения вызова после неотвѣта абонента, а p_k будем считать вероятностью отказа в обслуживании при наличии свободного ресурса (вероятность неотвѣта абонента).

Частный случай сети - занятие ресурса одной цифровой линией при совместной передаче n информационных потоков сообщений показан на рис. 1.

Пусть $P(j_1, \dots, j_n, i_1, \dots, i_n)$ - вероятность пребывания системы в состоянии $(j_1, \dots, j_n, i_1, \dots, i_n)$. Каждый из потоков сообщений характеризуется маршрутом следования сообщений от источника к получателю R_k , который задается последовательностью номеров соединительных линий, составляющих k -ый маршрут.

Для описания маршрутов следования сообщений в сети, а также величины канального ресурса, используемого каждым соединением, введем понятие маршрутной матрицы $Q = \|q_{j,k}\|$ прямоугольного вида размера $J \times n$ с компонентами:

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3} & \dots & q_{1,n-1} & q_{1,n} \\ q_{2,1} & q_{2,2} & q_{2,3} & \dots & q_{2,n-1} & q_{2,n} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ q_{J,1} & q_{J,2} & q_{J,3} & \dots & q_{J,n-1} & q_{J,n} \end{pmatrix}.$$

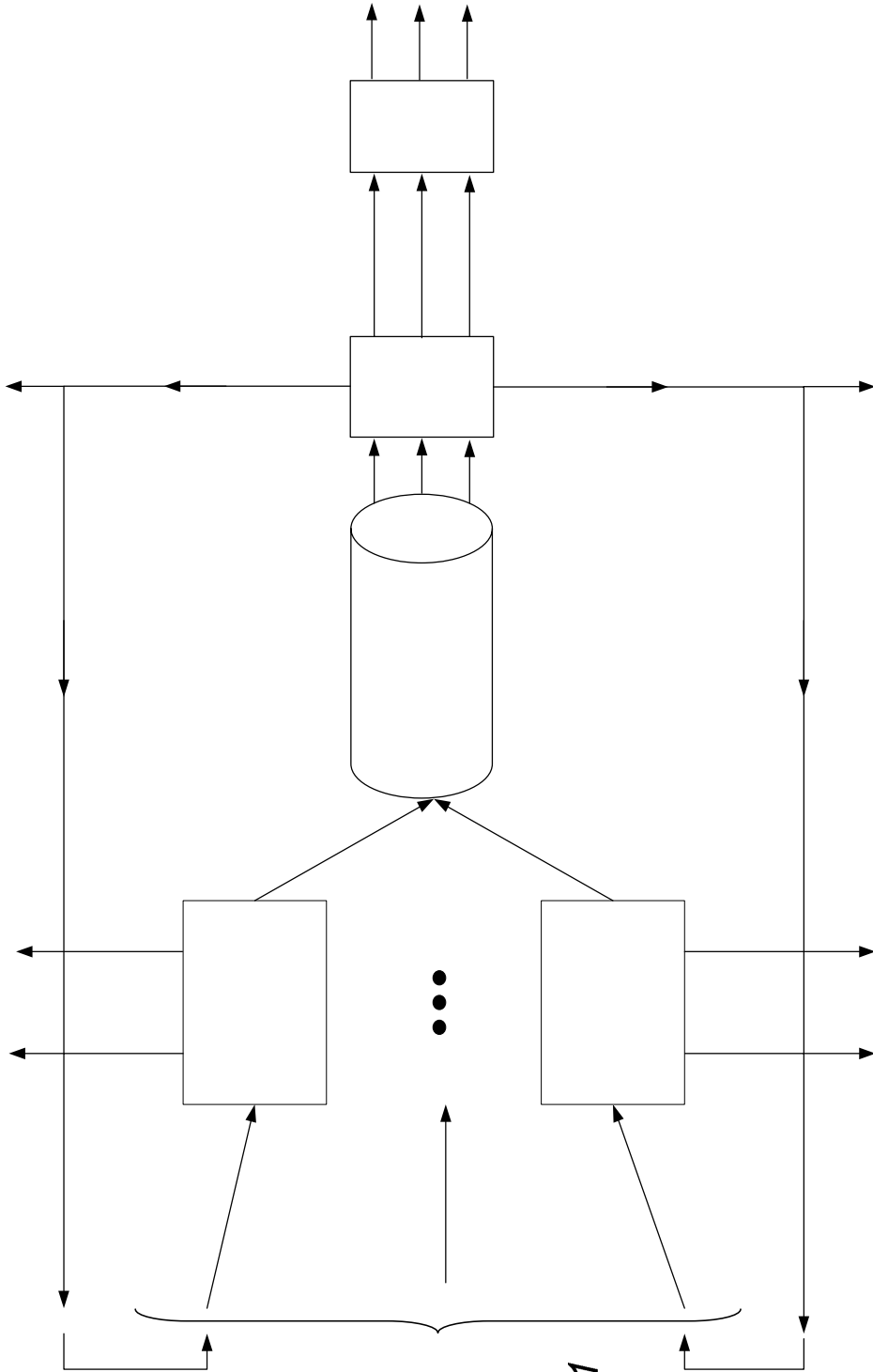


Рис. 1. Схема занятия ресурса цифровой линии при совместной передаче n информационных потоков сообщений

Элемент $q_{j,k}$ матрицы Q информирует, используется или нет j -ая цифровая линия для передачи сообщения k -го потока сообщений, и какое число единиц ресурса цифровой линии занято. Данная информация следует из равенства

ХОД $1-H_{s,1}$ НЕТ P_1

$$q_{j,k} = \begin{cases} b_k, & \text{сообщение } k \text{ – го информационного потока использует } j \text{ – ую} \\ & \text{цифровую линию, занимая для своего обслуживания } b_k \\ & \text{основных передаточных единиц} \\ 0, & \text{сообщение } k \text{ – го информационного потока не использует } j \text{ – ую} \\ & \text{цифровую линию.} \end{cases}$$

Ввиду громоздкости системы уравнений равновесия мы не будем приводить ее в данной статье, ограничившись подробным описанием модели.

Заключение

В работе приведено описание математической модели мультисервисной сети связи с учетом поведения абонента после отказа в обслуживании.

Литература

1. Лагутин В.С., Степанов С.Н. Телетрафик мультисервисных сетей связи. М.: Радио и связь, 2000. - 320 с.
2. Степанов С.Н. Материалы курса лекций "Основы теории моделирования сетей и систем телетрафика". М.: МТУСИ, 2000.
3. Степанов С.Н. Численные методы анализа систем с повторными вызовами. М.: Наука, 1983. – 20 с.
4. Клещев Н.Т., Федулов А.А., Симонов В.М. и др. Телекоммуникации. Мир и Россия. Состояние и тенденции развития. М.: Радио и связь, 1999. – 480 с.

MATHEMATICAL MODEL OF MULTISERVICE NETWORK WITH TAKING INTO ACCOUNT THE SUBSCRIBE BEHAVIOUR IN CASE REFUSAL IN SERVICE

The problem of repeated calls was and remains actual in telephone systems. New digital automatic telephone exchanges with program management, and also the telephone sets allowing periodically to dial the number up to an establishment of connection with the subscriber (function "call-back") have appeared. The telephone system has expanded a spectrum of given services and became multiservice. Thereof there is actual a research of models of multiservice networks. This paper presents an mathematical model of multiservice network with taking into account the subscribe behaviour in case refusal in service.



РАСЧЕТ ПУЧКОВ СОЕДИНИТЕЛЬНЫХ ЛИНИЙ ДЕКАДНО-ШАГОВЫХ И КООРДИНАТНЫХ АТС С УЧЕТОМ ЭФФЕКТА ПОВТОРНЫХ ВЫЗОВОВ

Степанов С.Н.¹, Мелик-Гайказова Э.И., Наумова Е.О.²

¹ЗАО «Система Телеком», ²ИППИ РАН

Введение

Повторные вызовы оказывают существенное влияние на работу систем телефонного сообщения. Причин этому несколько:

- Из экономических соображений количество коммутационного оборудования телефонных станций и узлов проектируется исходя из предположения об ограниченном числе вызовов (гораздо меньшем теоретически возможного), которые поступят в течение среднего времени занятия. Такое предположение для больших телефонных сетей в некоторых ее участках может не выполняться из-за возможных перераспределений нагрузки, вызванных строительством новых телефонных станций или появлением зон повышенного интереса абонентов, например справочных служб различного рода учреждений. Это ведет к возникновению в сети "узких мест", обладающих низкой пропускной способностью и являющихся причиной образования сильного потока повторных вызовов, часто по интенсивности превосходящего поток первичных вызовов.
- Необходимо учитывать, что для телефонной сети характерна неравномерность развития ее отдельных частей, например, вызываемая устаревшим оборудованием. Указанная диспропорция в развитии сети ведет к образованию повторных вызовов на стыках ее частей, что особенно сильно сказывается в междугородной телефонной связи
- Источником возникновения повторных вызовов на телефонных сетях часто является сам абонент: из-за ошибок в наборе номера, неответа или занятости вызываемого абонента и т.д.

Одной из основных причин появления повторных вызовов является нехватка пропускной способности системы связи. Чтобы в этой ситуации оценить необходимое по нагрузке число каналов, требуется построить алгоритм вычисления доли потерянных заявок на установление соединения в условиях, когда поступающий поток вызовов представлен как первичными, так и повторными заявками. В работе будет рассмотрено решение соответствующей проблемы для пучков соединительных линий декадно-шаговых и координатных АТС.

Основной результат

Рассмотрим модель пучка соединительных линий (СЛ) декадной или координатной АТС в условиях, когда получив отказ в обслуживании из-за занятости всех доступных каналов в требуемом направлении, абонент с вероятностью H повторяет вызовов через случайное время, имеющее экспоненциальное распределение с параметром, равным μ . Время занятия канала имеет экспоненциальное распределение с параметром, равным единице. Оценка характеристик пучков СЛ декадно-шаговых и координатных АТС с учетом повторных вызовов будет вестись с использованием метода замены потока повторных вызовов на пуассоновский. Методика соответствующей замены подробно изложена в [1]. Пусть x - неизвестная интенсивность потока повторных вызовов. Для ее определения воспользуемся предположением о совпадении законов сохранения для модели с повторными вызовами и ее аналога, получаемого после пуассоновской замены. В соответствии с законом сохранения для исходной модели, находящейся в стационарном равновесии, выполняется равенство

интенсивность потока повторных вызовов	=	интенсивность потока событий, приводящих к появлению источников повторных вызовов
--	---	---

Правая часть приведенного равенства в соответствии с правилами построения исследуемой модели может быть записана как сумма

интенсивность потока заблокированных первичных попыток с вероятностью H , приводящих к появлению источника повторных вызовов	+	интенсивность потока заблокированных повторных попыток с вероятностью H , приводящих к появлению источника повторных вызовов
--	---	--

Пусть $P(\lambda+x)$ - доля времени занятости доступных каналов для упрощенного аналога анализируемого пучка СЛ (декадного или координатного) [2], зависящая от неизвестной пуассоновской интенсивности поступления повторных вызовов x . Тогда из сформулированного выше тождества получаем следующее равенство

$$x = \lambda HP(\lambda + x) + xHP(\lambda + x).$$

Оно и дает неявное уравнение для определения x

$$x = \frac{\lambda HP(\lambda + x)}{1 - HP(\lambda + x)}.$$

Аналогично тому, как это было сделано в [1], можно исследовать вопросы существования решения и его единственности. Для нахождения решения можно использовать рекуррентную формулу

$$x_k = \frac{\lambda HP(\lambda + x_{k-1})}{1 - HP(\lambda + x_{k-1})}, \quad k = 1, 2, \dots$$

с начальным значением $x_0 = 0$.

Таким образом, для оценки доли времени занятости каналов в направлении - Р и пропущенной нагрузки - у необходимо рассчитать

$$\frac{\lambda HP(\lambda + x_k)}{1 - HP(\lambda + x_k)}$$

для некоторой последовательности x_k , сходящейся к решению x , и воспользоваться формулами

$$P = P(\lambda + x), \quad y = (\lambda + x)(1 - P(\lambda + x)).$$

С вычислительной точки зрения этот приближенный алгоритм расчета исходной модели не вызывает никаких затруднений и сводится к многократному использованию процедур оценки вероятности потерь, разработанных без учета повторных вызовов для соответствующих моделей пучков СЛ декадно-шаговых и координатных АТС [2].

Значение интенсивности первичной нагрузки λ определяется по измеренным значениям обслуженной нагрузки y . Для этого можно использовать расчетные модули, в которых по значению y находится соответствующая величина потерь P . Далее, используя полученные неявные уравнения, получаем соотношение для интенсивности первичной нагрузки

$$\lambda = \frac{y(1-PH)}{1-P}.$$

Заключение

В работе получен простой алгоритм оценки доли потерянных заявок на выделение канальной емкости в пучке СЛ декадно-шаговых и координатных АТС с учетом эффекта повторных вызовов.

Литература

1. Степанов С.Н. Численные методы анализа систем с повторными вызовами. М. Наука, 1983.
2. Корнышев Ю.Н., Пшеничников А.П., Харкевич А.Д. Теория телетрафика. М.: Радио и связь, 1996.

THE SOLVING OF DIMENSIONING PROBLEM FOR DECADE-STEP AND CROSSBAR SYSTEMS WITH REPEATED CALLS

Simple algorithm for estimation of main performance measures for decade-step and crossbar systems with taking into account the subscriber behavior in form of retrials is derived.

РАЗРАБОТКА МОДЕЛИ СЕТИ СОТОВОЙ СВЯЗИ ПРИ СОВМЕСТНОМ ОБСЛУЖИВАНИИ ПОТОКОВ РЕЧИ И ДАННЫХ

Болдырев М.В.

Московский Институт Радиотехники Электроники и Автоматики

Введение

Одной из основных тенденций развития мирового рынка сотовой связи в последние 10 лет является завоевание господствующих позиций цифровыми стандартами сотовой связи и, конечно, наиболее популярным из них – стандартом GSM. Количество абонентов данного стандарта на конец 2004 года по оценкам аналитиков составило 73% от общего количества абонентов сотовой связи в мире. Такую популярность данный стандарт завоевал по ряду объективных причин и, не в последнюю очередь, из-за наличия в его основе как сервисов, основанных как на технологии коммутации каналов, так и технологии пакетной коммутации - GPRS (General Radio Packet Service) и ее дальнейшим развитием технологии EDGE (Enhanced Data for Global Evolution). Активное внедрение на сетях GSM технологий пакетной коммутации началось с 2000 года и к настоящему моменту, практически, все операторы в мире предоставляют своим абонентам услуги на их основе.

Технология GPRS базируется на принципах множественного доступа с временным разделением каналов (TDMA) и позволяет абонентам в многослотовом режиме получать доступ к многочисленным сервисам. Наиболее популярными сервисами являются доступ в публичный Internet и корпоративные сети Intranet, к сервисам WAP, передача и прием MMS, а в последнее время и новые высокоскоростные сервисы такие, как доступ к мобильному Video-контенту, а также сервисы передачи голоса по пакетным сетям (технология Push To Talk). Столь внушительное количество предоставляемых сервисов и их разнородность по предъявляемым требованиям к полосе пропускания и возможным задержкам, с одной стороны, и ограниченность радиоресурсов, имеющихся у сотовых операторов, с другой стороны, не позволяет предоставить абонентам пакетных услуг гарантированного качества сервиса. Этот факт заставляет операторов мобильной связи искать методы расчета и планирования своих сетей для сбалансированного предоставления своим абонентам как основных услуг, основанных на технологии коммутации каналов, так и дополнительных при предоставлении услуг пакетной коммутации.

Важность решения сформулированной проблемы для оптимального распределения телекоммуникационного ресурса современных сетей стандарта GSM/GPRS/EDGE заставляет искать новые методы для оценки основных показателей совместной передачи разнородных потоков речи и данных в мобильных сетях. Описание процесса занятия и освобождения радиоресурса различными пользовательскими потоками в сетях стандарта GSM/GPRS и создание функциональной модели на основе методов и подходов теории телетрафика будет проведено в основной части работы.

Основной результат

Рассмотрим процесс занятия канального ресурса соты в мобильной сети, применив подходы теории телетрафика и построим схему функционирования такой модели.

Стандарт GSM основан на технологии временного разделения каналов TDMA, что позволяет на одной несущей частоте получить восемь независимых каналов для обслуживания полезной нагрузки (голос или данные). Ресурс в один временной интервал (Time Slot) будем рассматривать как основную передаточную единицу. Теперь опишем схему доступа к канальным ресурсам в соте мобильной сети, имеющие V передаточных единиц.

При рассмотрении цифровой сети стандарта GSM можно выделить два основных потока нагрузки: поток речевых сообщений (GSM вызов) и поток пакетов данных (GPRS вызов). Речевое сообщение имеет абсолютный приоритет в получении передаточного ресурса. Поток речевых сообщений использует для передачи единичный ресурс линии и обслуживается также, как и пакеты данных в рамках модели с потерями. Для обслуживания пакета данных может требоваться ресурс больше одной передаточной единицы, обозначим его k . При этом, если требуемого ресурса нет, то для передачи данного потока данных предоставляется максимально возможный ресурс из имеющегося в настоящий момент. В случае поступления приоритетного речевого сообщения, когда весь свободный ресурс занят обслуживанием потока данных, единичный ресурс цифровой линии будет переопределен для обслуживания речевого потока, а поток данных продолжит обслуживание с уменьшенным на единицу зарезервированным для его передачи ресурсом цифровой линии. При описании функционирования модели необходимо также учесть следующее:

1. Существует возможность использования единичного ресурса несколькими потоками данных дискретно во времени, что значительно повышает пропускную способность цифровой линии связи. Поскольку использование передаточного потоками данных происходит неравномерно и характеризуется значительными временными задержками, это не приводит к значительному ухудшению качества сервиса для конечного пользователя. Максимальное значение для использования единичного ресурса несколькими потоками данных обозначим m .

2. Для повышения качества доступа потоков данных к ресурсам цифровой линии связи на ней может быть зарезервирован определенный ресурс под передачу пакетов данных, который ни при каких условиях не будет предоставлен поступающим на обслуживание речевым сообщениям. Число зарезервированных под передачу пакетов данных передаточных единиц обозначим l .

Построенная схема введенной модели совместного обслуживания потоков речи и данных в сетях стандарта GSM/GPRS представлена на рис.1.

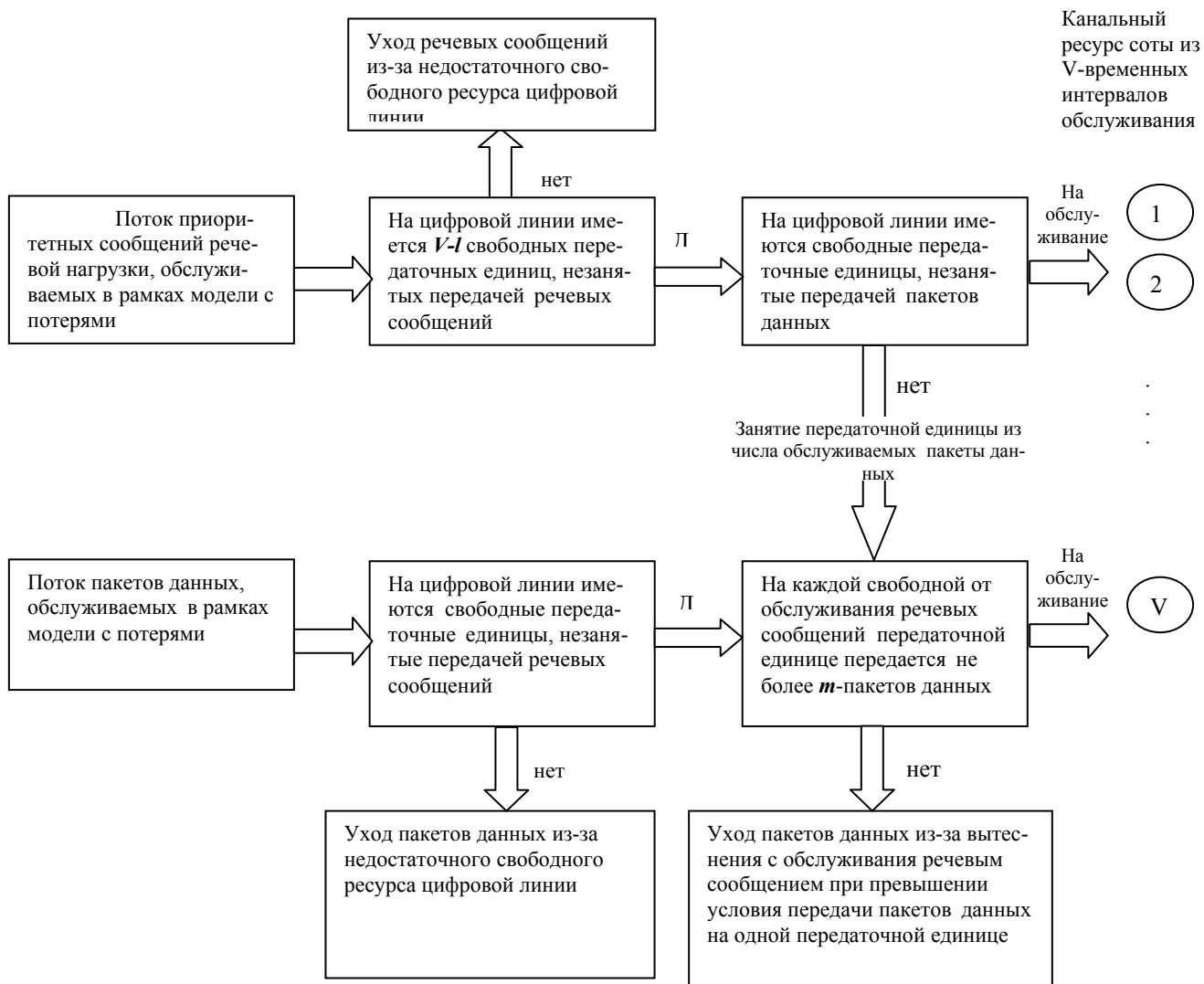


Рис.1 Схема функционирования модели совместного обслуживания потоков речи и данных в мобильных сетях стандарта GSM/GPRS.

На этапе проведенной формализации процедуры занятия канального ресурса соты в сетях мобильной связи при одновременном обслуживании потоков речи и данных были определены основные характеристики, характерные для моделей теории телетрафика, такие как: основная передаточная единица, скорость цифровой линии, объем передаточного ресурса для обслуживания потоков речи и данных. Используя эти данные, нам удалось создать схему функционирования описанной модели.

Данная в работе интерпретация позволяет вести оценку характеристик качества совместной передачи сообщений поток речи и данных в сетях мобильной связи второго поколения с помощью класса моделей теории телетрафика с потерями. Построенная модель даст возможность получить оценки основных показателей совместной передачи потоков речи и данных в мобильных сетях и определить количество канального ресурса соты, достаточную для передачи сообщений с заданным качеством.

Заключение

В работе описаны основные проблемы, с которыми в последнее время сталкиваются сотовые операторы для обеспечения оптимального распределения телекоммуникационного ресурса в современных сетях мобильной связи. Проведено исследование и формализация процесса занятия канального ресурса соты в сетях стандарта GSM/GPRS при одновременном обслуживании потоков речевых сообщений и данных и построена схема функционирования введенной модели.

Используя данный подход, можно создать методы получения оценок основных показателей совместной передачи потоков речи и данных в мобильных сетях, среди которых доля потерянных сообщений речевой нагрузки и пакетов данных, средняя величина ресурса цифровой линии, занятого передачей разных типов нагрузки, и среднее время нахождения пакетов данных и сообщений речевой нагрузки на передаче.

Литература

1. Лагутин В.С., Степанов С.Н. Телетрафик мультисервисных сетей связи. М.: Радио и связь, 2000. – 320с.
2. Карташевский В.Г., Семенов С.Н., Фирстова Т.В. Сети подвижной связи. М.: Эко-Трендз, 2001. – 300с.
3. Ратынский М.В. Сотовая связь как система массового обслуживания.// Мобильные системы, 1997, №2, с.16-18.
4. Гершман И.Р. Модели и методы расчета абонентской нагрузки в сотовых сетях.// Международная конференция по теории телетрафика. Санкт-Петербург, ЛОНИИС, 01-07 июня 1998, с.293.
5. Синчуков А. Сотовая связь: что нового?// Радио, 2004, №5, с.70-71.
6. Гольшко А. Системы GSM промежуточного поколения.// Радио, 2001, №7, с.68-69.
7. Афанасьев В., Развитие сетей подвижной связи в мире: тенденции, перспективы. Тезисы выступления на 5-м юбилейном ежегодном Российском форуме коммуникаций 15 ноября 2004 г., г. Москва.

DEVELOPMENT OF THE CELLULAR NETWORK MODEL AT THE TIME OF THE JOINT SERVICE OF THE SPEECH AND DATA FLOWS

The problem of optimum distribution of a telecommunication resource in GSM/GPRS mobile networks, forces to search for new methods of estimation the basic parameters of the joint service of the diverse traffic flows. It is possible to create the circuit of functioning of a cellular network model having applied approaches of the teletraffic theory. This model will enable to dimension of the cells channel resource with prescribed level of service and given input traffic flows.

ПОСТРОЕНИЕ УПРОЩЕННОЙ МОДЕЛИ РАБОТЫ CALL-ЦЕНТРА

Серова С.А.

МТУСИ

Проведенный анализ работы современных Call-центров позволил учесть их специфические особенности. Это возможность обработки вызовов различных типов (голосовые, Интернет-обращения), интеллектуальная маршрутизация вызовов, наличие нескольких групп операторов, возможность обработки вызовов без привлечения операторов (наличие IVR) и пр. В результате была построена модель, на основании которой была разработана упрощенная модель для оценки характеристик работы Call-центра.

В качестве одного из упрощений рассмотрим модель работы Call-центра, в которой имеются только две группы операторов, разделенных по типу обслуживаемых вызовов. А также для начала рассмотрим модель без повторных вызовов, что удобно для первоначального анализа. В результате получили модель следующего вида.

На систему поступает пуассоновский поток первичных вызовов с интенсивностью λ . Вызов, поступивший на обслуживание в Call-центр, попадает на систему распределения вызовов, которая в зависимости от номера вызывающего абонента, набираемого им номера или идентификационного номера, направляет вызов в требуемую группу операторов (в нашем случае в первую или вторую). Число операторов в первой и второй группах, обслуживающих вызовы заданного типа, составляет соответственно v_1 и v_2 . Перед каждой группой имеется буфер (очередь) определенной длины для ожидания обслуживания: l_1 – количество мест для ожидания обслуживания первой группой операторов и l_2 – количество мест для ожидания обслуживания второй группой операторов. Время ожидания в очереди ограничено постоянной величиной ρ .

С вероятностью P_1 вызов попадает в первую группу операторов, а с вероятностью $P_2 = 1 - P_1$ – во вторую группу операторов. Если все операторы в первой группе заняты, то вызов становится в очередь к этой группе операторов. Если все операторы первой группы заняты и в очереди этой группы нет свободных мест, то вызов направляется в группу операторов, имеющую вторым приоритетом обслуживание вызовов этого типа, т.е. во вторую группу операторов. Вызов также может быть перенаправлен во вторую группу операторов, если время пребывания этого вызова в очереди превосходит допустимое - ρ . Если все операторы второй группы заняты и в очереди этой группы нет свободных мест для ожидания обслуживания или время ожидания в очереди превосходит допустимое, то вызов покидает систему необслуженным.

Процесс обслуживания вызовов, попавших с вероятностью P_2 во вторую группу операторов проходит аналогично вышеизложенному. Если все операторы во второй группе заняты, то вызов становится в очередь к этой группе операторов. Если все операторы второй группы заняты и в очереди этой группы нет свободных мест, то вызов направляется в группу операторов, имеющую вторым приоритетом обслуживание вызовов этого типа, т.е. в первую группу операторов. Вызов также может быть перенаправлен в первую группу операторов, если время пребывания этого вызова в очереди превосходит допустимое - ρ . Если все

операторы первой группы заняты и в очереди этой группы нет свободных мест для ожидания обслуживания или время ожидания в очереди превосходит допустимое, то вызов покидает систему необслуженным.

В соответствии с введенным описанием будет составлена система уравнений равновесия. Построенная модель будет использована для оценки показателей обслуживания вызовов Call-центром.

DESIGNS OF CALL CENTRE SIMPLIFIED MODEL

Call center simplified model is considered in this paper. In this model only two operator's groups with their waiting queues and are not present repeated calls. Incoming stream of primary calls distribute between two operator's groups with specified probability. If one of operator's groups cannot serve a call it is sent to second group on a priority and only after refuse in service is lost. The constructed model will allow evaluation of call center characteristics.

МОДЕЛИ ТРАФИКА

Денисова Т.Б.

Поволжская государственная академия телекоммуникаций и информатики

Выбор типа модели трафика существенно зависит от цели построения модели, т.е. от того, для чего и как будет использоваться модель трафика. Модель, хорошо решающая один класс задач (например, предназначенная для прогнозирования результатов измерений трафика на определенный период), часто оказывается непригодной для решения другого класса задач (например, для оценки основных характеристик качества обслуживания трафика). Для прогнозирования трафика можно использовать модели временных рядов. Для оценки таких характеристик качества обслуживания трафика, как задержка передачи пакета, вариация задержки передачи пакета, доля потерянных пакетов, используется теория массового обслуживания, в которой центральное место занимают Марковские процессы. В этом случае желательной моделью трафика является модель Марковского потока. Фрактальные регрессионные модели и модели потоков, управляемые Марковскими процессами, используются в прикладных задачах оценивания сетевой производительности. Эти модели, они являются развитием традиционных регрессионного и Марковского подходов для аппроксимации самоподобных процессов.

Для краткосрочного и долгосрочного прогнозирования трафика по результатам измерений могут быть использованы различные модели временных рядов: детерминированная модель, модель авторегрессии (АР), модель авторегрессии и скользящего среднего (АРСС), модель интегрированной авторегрессии и скользящего среднего (ИАРСС).

В общем случае модель наблюдаемого процесса $y(t)$ может быть задана уравнением

$$A(\beta)y(t) = G(\beta)u(t-1) + \psi(t-1) + B(\beta)\xi(t), \quad (1)$$

где $u(t)$ - входной процесс (управление),

$\psi(t)$ - детерминированный тренд,

$\xi(\cdot)$ - последовательность одинаково распределенных случайных величин с нулевым средним,

β - оператор единичной задержки, определяемый соотношением $\beta(y(t)) = y(t-1)$,

$A(\beta)$, $B(\beta)$, $G(\beta)$ - полиномы степени m_1 , m_2 , m_3 соответственно.

Обычно процесс АРСС задается моделью

$$A(\beta)y(t) = B(\beta)\xi(t).$$

Если полином $A(\beta)$ имеет по крайней мере хотя бы один корень на единичной окружности, демонстрирующий нестационарность $y(t)$, то $y(t)$ описывается моделью ИАРСС. Определим оператор задержки как $\beta^d y(t) = y(t-d)$. Тогда процесс ИАРСС задается моделью

$$A_1(\beta)(1-\beta)^d y(t) = B(\beta)\xi(t), \quad (2)$$

где $A_1(\beta)(1-\beta)^d = A(\beta)$.

Параметр d в процессе ИАРСС рассматривается только как целочисленный. Для фрактальной интегрированной авторегрессии и скользящего среднего (ФИАРСС) параметр d принимает дробные значения, $d \in (-0.5, 0.5)$. Модели ФИАР и ФИАРСС описывают долговременные зависимости.

Проблема точного расчета характеристик систем обслуживания реальных потоков потребовала создания удобной математической модели этих потоков. Такой математической моделью является модель в виде ВМАР-потока (Batch Markovian Arrival Process, *групповой Марковский входной поток*). Модель ВМАР-потока аппроксимирует самоподобный процесс с помощью управляющего Марковского процесса.

ВМАР-поток полностью характеризуется размерностью $W + 1$ управляющего потока, интенсивностями λ_v времен пребывания этого процесса в своих состояниях, $v = \overline{0, W}$, и набором переходных вероятностей $p_k(v, v')$, $v, v' = \overline{0, W}$, $k \geq 0$. Эту информацию удобно хранить в виде набора матриц D_k , $k \geq 0$, размера $(W + 1) \times (W + 1)$, определяемых следующим образом.

$(D_k)_{v,v'}$ - (v, v') -й элемент матрицы D_k - имеет вид:

$$(D_k)_{v,v'} = \lambda_v \cdot p_k(v, v'), \quad v, v' = \overline{0, W}, \quad k \geq 1, \quad (3)$$

$$(D_0)_{v,v'} = \begin{cases} \lambda_v \cdot p_0(v, v'), & v \neq v', v, v' = \overline{0, W}, \\ -\lambda_v, & v = v', v = \overline{0, W}. \end{cases} \quad (4)$$

Модель ВМАР-потока обобщает многие известные ранее модели потоков: МАР (Markovian Arrival Process – Марковский входной поток), ММРР (Markov Modulated Poisson Process - Марковский модулированный пуассоновский поток), РН (Phase Type – поток фазового типа), ВММРР (Batch Markovian Modulated Poisson Process – групповой Марковский модулированный пуассоновский).

Рассмотрим подход к построению моделей реальных потоков по измерительной информации с использованием кусочно-непрерывных временных рядов. В качестве модели потока примем модель ММРР потока, которая хорошо описывает многие реальные потоки и имеет прозрачную физическую трактовку пуассоновского потока с несколькими уровнями интенсивностей.

Будем полагать, что осуществляется измерение трафика (регистрация случайного процесса). Пусть измеряется либо количество пакетов (количество событий) в последовательные равные интервалы времени, либо регистрируются моменты поступления пакетов (наступления событий). Для построения ММРР модели необходимо выделить квазистационарные состояния случайного процесса, определить интенсивности пребывания процесса в выделенных состояниях и определить вероятности перехода между выделенными состояниями. Таким образом, управляющий процесс ММРР потока можно представить в виде кусочно-непрерывной модели временного ряда, участки непрерывности (однородности) которой чередуются в определенном порядке и соответствуют квазистационарным состояниям процесса. Кусочно-непрерывная модель описывает выборочную последовательность состояний управляющего процесса ММРР потока. Имея достаточно большую выборку состояний процесса, можно оценить переходные и стационарные вероятности состояний управляющего процесса.

Пусть мы располагаем исходными статистическими данными – выборкой временного ряда:

$$(y_1, t_1), (y_2, t_2), \dots, (y_N, t_N),$$

где t_i - момент регистрации трафика,

y_i - интенсивность трафика на интервале (t_{i-1}, t_i) .

Алгоритм помехоустойчивого оценивания параметров кусочно-непрерывного тренда

На каждом участке непрерывности решается задача помехоустойчивого оценивания параметров θ тренда. В предлагаемом алгоритме в качестве критерия ошибки приближения используется критерий статистического исключения. Алгоритм использует рекуррентную процедуру с усреднением, которая не использует матричных операций. Статистическая М-устойчивость оценок параметров θ достигается путем использования функции Тьюки.

Алгоритм имеет вид:

$$\theta^{m+1} = \theta^m + \gamma_m \varphi_m \psi(r_m) \quad (5)$$

$$r_m = y_m - \theta_m^T \varphi_m \quad (6)$$

$$\hat{\theta}^{m+1} = \hat{\theta}^m + (\theta^m - \hat{\theta}^m) / (m + 1) \quad (7)$$

$$s_m = s_{m-1} - \frac{1}{2m} \cdot \left(s_{m-1} - \frac{\psi(r_m) \cdot r_m}{s_{m-1}} \right) \quad (8)$$

$$\psi(r_m) = \begin{cases} r_m, & \text{если } |r_m| \leq c \\ \pm c, & \text{если } |r_m| > c, \quad c = s_m \cdot k(\varepsilon) \end{cases} \quad (9)$$

$$\gamma_m = \begin{cases} \gamma_0 \cdot (m - m_0)^{-0.5}, & \text{если } m > m_0 \\ \gamma_0, & \text{если } m \leq m_0 \end{cases} \quad (10)$$

Алгоритм построения вероятностной модели управляющего процесса ММРР потока

Алгоритм построения вероятностной модели управляющего процесса ММРР потока включает следующие шаги:

1. На начальном интервале каждого участка однородности по ограниченной выборке (объемом $n_0 = 10$) с использованием алгоритма (5)-(10) оцениваются параметры θ , описывающие трафик на однородном участке. Подсчитывается число переходов $m_{ij} = m_{ij} + 1$ из i - квазистационарного состояния (i - участка) в j - состояние. Подсчитываются среднее время пребывания в i -состоянии τ_i и среднее время между повторениями i -состояний T_i .

2. Для последующих точек участка проверяется их принадлежность данному участку или следующему однородному участку временного ряда. Считается, что произошла разладка временного ряда, если несколько (более пяти) подряд измерений трафика отклоняются от прогноза более чем на контрольную величину

$$\left| y_i - \hat{f}(\theta_i, \sigma_i, t_i) \right| > k \cdot \sigma_i, \quad (11)$$

где σ_i - помехоустойчивая оценка масштаба, $k = 1.5$.

Если не обнаружена разладка временного ряда, то выполняется рекуррентная процедура (5)-(10) для последующих точек участка, и уточняются параметры θ .

3. По построенной последовательности уровней θ управляющего процесса формируется матрица вероятностей переходов

$$P = \{P_{ij}\}, \quad P_{ij} = \frac{m_{ij}}{\sum_j m_{ij}} \quad (12)$$

и вектор стационарного распределения состояний

$$\pi = \{\pi_i\}, \quad \pi_i = \frac{\tau_i}{T_i} \quad (13)$$

ММРР модель трафика задается матрицами D_0 и D_1 , которые определяются следующим образом: $D_0 = \Lambda$, $D_1 = \Phi(P - I) - \Lambda$, где $\Lambda = \text{diag}\{\lambda_0, \dots, \lambda_W\}$, т.е. диагональная матрица с диагональными элементами λ_v , $v = \overline{0, W}$, $\Phi = \text{diag}\{\varphi_0, \dots, \varphi_W\}$, P - стохастическая матрица. Если кусочно-непрерывная модель описывает интенсивность трафика, то $\lambda_v = \theta_v$, W - число однородных участков модели, длительность каждого участка имеет показательное распределение с параметром $\varphi_v = 1/\tau_v$.

Выводы

1. Трафик телекоммуникационных сетей представляет собой нестационарный случайный процесс, для описания которого в теории массового обслуживания используются модель ВМАР потока (3-4). В то же время для прогноза трафика на практике используются модели временных рядов (1-2).

2. Для построения модели ММРР потока по результатам измерений трафика предложены алгоритм помехоустойчивого оценивания параметров кусочно-непрерывного тренда (5)-(10) и алгоритм построения вероятностной модели управляющего процесса ММРР потока (11)-(13).

3. Предложенные алгоритмы позволяют по результатам измерений строить модели реального трафика телекоммуникационных сетей с целью применения моделей для различных прикладных задач: для прогноза, для оценки характеристик качества обслуживания в сетях связи, для определения оптимальной пороговой политики управления системами.

MODELS OF TRAFFIC

Denisova T.

We take up compliance of models of traffic to the applied tasks of evaluation and management in the communication nets. The recurrent algorithm of building the piece-unbroken model of current and the algorithm of building the probabilistic model of directing process of the current are supposed. The algorithm of building of the piece-unbroken model of current has an asymptotic speed of likeness and provides the statistic steadiness of the parameters of the model. The algorithm of the building of the probabilistic model of the managing process uses the evaluation of the frequency of the conversions of the process of reconstruction.

АЛГОРИТМ ОБЕСПЕЧЕНИЯ КАЧЕСТВА ОБСЛУЖИВАНИЯ В УСЛОВИЯХ ВЛИЯНИЯ ЭФФЕКТА САМОПОДОБИЯ ТЕЛЕТРАФИКА

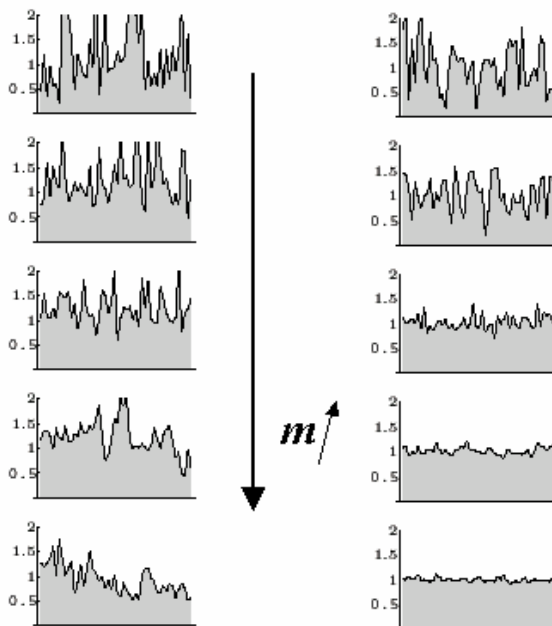
Богатырев Е.А., Петров В.В.
(vpetroff@nm.ru)

Московский Энергетический Институт (ТУ)

Современные информационные системы предъявляют повышенные требования к уровням потерь, задержек информации, джиттеру при обработке информационных потоков (трафика) в телекоммуникационной системе. Другими словами – требуют обеспечения определенного качества обслуживания. Основные предложенные механизмы для решения данной задачи в системах с пакетной передачей данных разработаны с учетом предположения, что трафик в таких сетях хорошо описывается простейшим (пуассоновском) потоком. Тем не менее, многочисленные исследования в области изучения телетрафика утверждают о том, что в современных системах распределения информации с пакетной передачей трафик является пачечным и обладает так называемым свойством самоподобия [1]. Коэффициент пачечности (пачечность) для заданного потока соответствует отношению пиковой интенсивности процесса поступления заявок на обслуживание к его среднему значению. Другими словами, на качественном уровне пачечный процесс выглядит сильно изменчивым: при сравнительно небольшом среднем значении в реализации присутствует достаточное количество относительно больших выбросов. Такая форма трафика для процесса обработки достаточно неудобна, поскольку для обеспечения заданного уровня потерь необходимо увеличивать пропускную способность каналов, мирясь при этом с существенным недоиспользованием ресурсов системы (низкой утилизацией).

Свойство самоподобия (масштабной инвариантности) еще больше усложняет дело, поскольку пачечный трафик, обладающий этим свойством, сохраняет свою пачечность при агрегировании и сложении (суперпозиции) нескольких потоков в магистральном канале [2]. Причем агрегирование представляет собой процесс усреднения реализации сетевого трафика по блокам данных одинаковой длительности. Другими словами, процесс в магистральном канале современной телекоммуникационной сети с пакетной передачей данных также является пачечным и недостаточно сглаживается при увеличении количества потоков. Присутствие данного явления можно заметить по графикам, приведенным на рис.1. Анализируя данный рисунок, легко видеть, что с увеличением уровня агрегирования (сверху - вниз) самоподобный трафик (слева) сглаживается намного хуже, чем несамоподобный трафик (справа). Как уже было сказано выше, статическое выделение пропускной способности для такого пачечного трафика приводит к ее неэффективному использованию.

Для разрешения описанного выше естественного противоречия предлагается реализовать новый алгоритм, который динамически (адаптивно) выделяет пропускную способность общего канала для заданного пачечного потока по требованию. Предполагается, что при этом данному потоку необходимо обеспечить высокий приоритет. Оставшиеся ресурсы общего канала распределятся между остальными источниками.



5.

Рис. 1. Временные реализации реального сетевого (самоподобного) трафика (слева) и традиционной несамоподобной (Пуассоновской) модели телетрафика (справа) при различных масштабах временной оси (из D. Wischik. Implication of long-range dependence, 2001). Сверху-вниз масштаб временной оси укрупняется

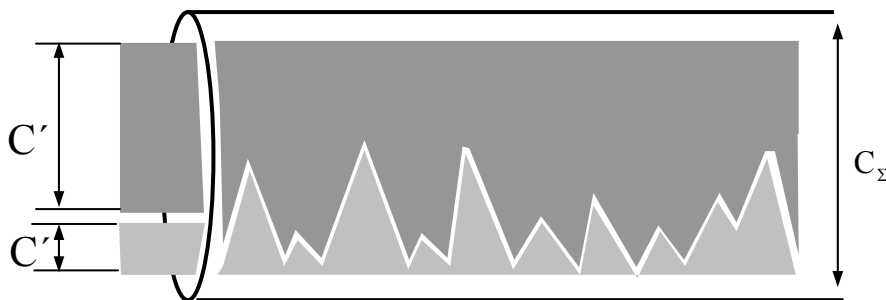


Рис. 2. Динамическое распределение пропускной способности физического канала между двумя логическими

Данная идея проиллюстрирована на рис. 2. Здесь общая пропускная способность C_{Σ} канала распределяется следующим образом: для пачечного трафика, имеющего в данном случае высокий приоритет, выделяется динамически изменяющаяся пропускная способность C'_1 . Сэкономленный при этом ресурс канала в виде пропускной способности C'_2 выделяется для остальных (низкоприоритетных) потоков. При этом данный алгоритм обеспечивает высокое качество обслуживания для выбранного потока трафика и, кроме того, экономит ресурсы пропускной способности канала.

Для реализации описанного алгоритма в каждый момент времени необходима информация о потребностях высокоприоритетного потока в ресурсах пропускной способности в будущем. Эта информация требуется для того, чтобы заблаговременно перераспределить ресурс пропускной способности между потоками. Другими словами необходимо уметь прогнозировать интенсивность трафика.

Для предсказания сетевого трафика возможно использовать авторегрессионные модели различных порядков и порядков такие как $AR(p)$, $ARMA(p,q)$, $ARIMA(p,q)$, $FARIMA(p,d,q)$ и др. [3] Наиболее перспективной моделью в целях прогнозирования самоподобного трафика видится модель $FARIMA(p,d,q)$, поскольку она обладает самоподобными свойствами, как и сетевой трафик.

Для реализации функции обеспечения качества обслуживания при влиянии эффекта самоподобия трафика нами разработан специальный алгоритм [4], представленный на рис.3. Данный алгоритм основан на принципе “Token Bucket” (корзина маркеров) и использует метод динамического управления пропускной способностью на основе прогнозирования трафика. Здесь входящий поток пакетов одновременно с поступлением в систему попадает на схему агрегирования по временным интервалам Δ . После чего n последних отсчетов получившегося временного ряда поступают на предсказатель, который выдает прогностическую оценку следующего отсчета $\hat{x}(n+1)$ агрегированного ряда x на время Δ вперед. Исходя из требований к

величине возможных потерь оценивается пропускная способность системы $\hat{C}(n+1)$. В результате размер буфера жетонов устанавливается $V_c = \hat{C}(n+1) \cdot T_c$ на время Δ вперед.

В результате исследований функционирования данного алгоритма в среде сетевого эмулятора ns-2 с использованием трафика реальных систем в качестве входящего потока данных мы нашли, что эффективность данного алгоритма (даже при предсказателе по последнему известному значению) характеризуется уменьшением потерь и одновременным увеличением степени использования сетевых ресурсов при обработке самоподобного пачечного трафика примерно на 10% по сравнению с методом статического распределения пропускной способности.

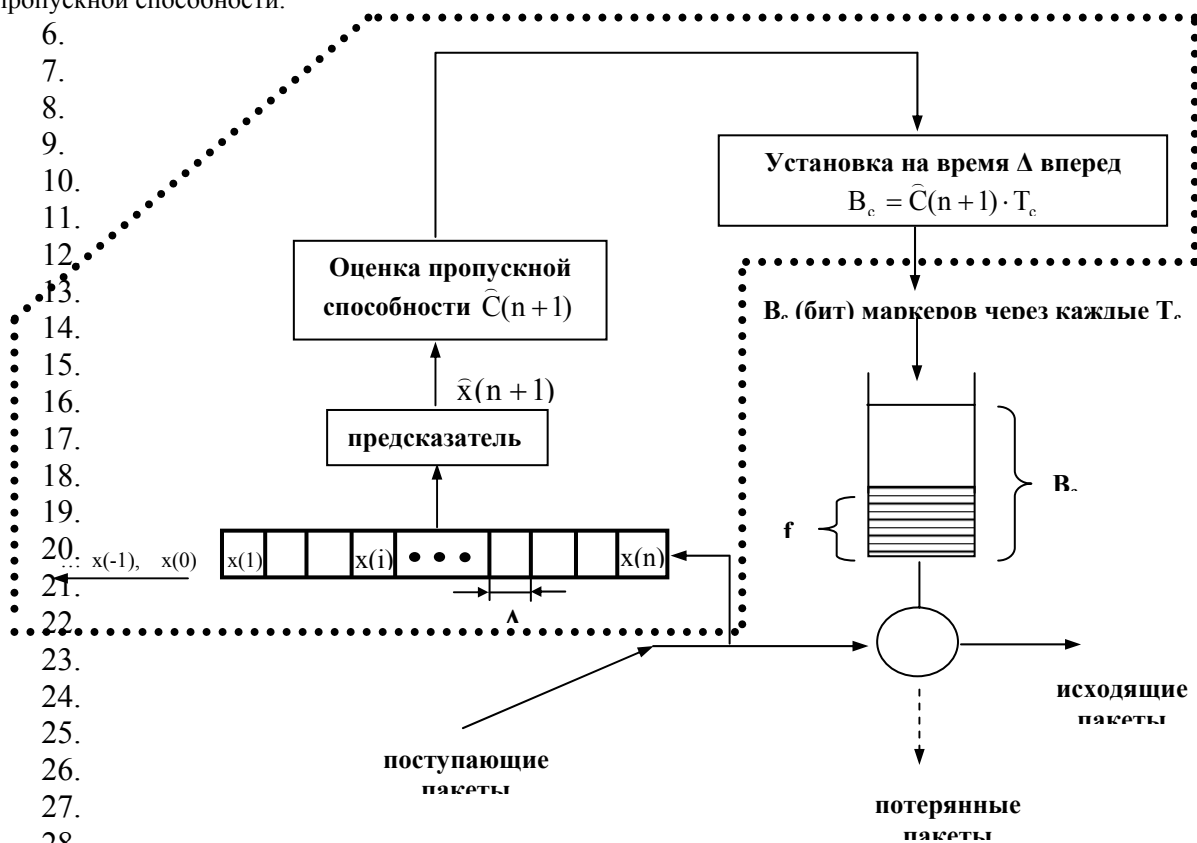


Рис. 3. Алгоритм обеспечения качества обслуживания при влиянии эффекта самоподобия трафика

Необходимо отметить, что разработанный метод не является универсальным средством при решении задач, связанных с проблемой самоподобия. Область его применения ограничивается случаями, когда возможно (чаще всего это удастся сделать) разделить трафик на две категории: приоритетный и не приоритетный, а также когда используется техника виртуальных каналов (в отличие от статистического мультиплексирования потоков). При этом если приоритетный трафик оказывается самоподобным и обладает высоким коэффициентом пачечности, то разработанный алгоритм будет эффективен. В противном случае его эффективность окажется не хуже эффективности метода статического распределения пропускной способности.

Литература

1. Leland W.E., Taqu M.S., Willinger W., and Wilson D.V. On the self-similar nature of ethernet traffic // IEEE/ACM Transactions of Networking, 2(1), 1994. – P. 1-15.
2. Willinger W., Taqu M.S., Sherman R., and Wilson D.V. Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level // IEEE/ACM Transactions on Networking, Vol. 5, No. 1, 1997 - P. 71-86.
3. Шелухин О.И., Тенякшев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях. Монография: Под ред. О.И. Шелухина. – М.: Радиотехника, 2003. – 480 с.
4. Петров В.В. Структура телетрафика и алгоритм обеспечения качества обслуживания при влиянии эффекта самоподобия. Диссертация на соискание ученой степени кандидата технических наук – Москва, 2004.

It is known, that the traffic of today’s networks has a “self-similar” nature. As result, the burstiness of traffic does not disappear at large scales. This property of traffic is at the heart of important problem – Quality of Service

(QoS) provisioning in networks with self-similar traffic. In this article we propose new algorithm of QoS for the case of burstiness and self-similar traffic.

МНОГОЛИНЕЙНАЯ СИСТЕМА ОБСЛУЖИВАНИЯ С ПОВТОРНЫМИ ВЫЗОВАМИ И ДВУМЯ ТИПАМИ ЗАПРОСОВ

Клименок В.И., Орловский Д.С.

Белорусский государственный университет, Минск

Системы массового обслуживания (СМО) с повторными вызовами являются хорошими математическими моделями процессов передачи информации в реальных сетях телекоммуникаций. Характерной чертой этих систем является то, что поступающий запрос, заставший обслуживающее устройство занятым, не становится в очередь и не покидает систему навсегда, а повторяет попытки попасть на обслуживание через случайные интервалы времени.

Различные практические задачи, появляющиеся в телефонных сетях, мобильных сотовых сетях связи и т.д. могут быть успешно решены путём аналитического моделирования соответствующих процессов передачи информации в терминах многолинейных СМО с повторными вызовами и двумя типами запросов различных приоритетов.

В [2] рассмотрена одна из таких СМО — система $M_1, M_2 | M | c$ с бесконечной очередью для приоритетных и бесконечной орбитой для неприоритетных запросов. Для такой системы получено условие эргодичности и некоторые средние характеристики её производительности.

В данной работе рассматривается N -линейная СМО с буфером объёма $R < \infty$ и орбитой неограниченного размера. В систему поступает два независимых стационарных пуассоновских потока. Запрос из приоритетного потока (с параметром $\lambda_1 < \infty$) немедленно начинает обслуживаться одним из свободных приборов. Если свободных приборов нет, но есть свободные места в буфере, то запрос становится в буфер. Если же буфер в момент поступления запроса из приоритетного потока полностью заполнен, то с вероятностью q ($0 \leq q \leq 1$) этот запрос идёт на орбиту, а с вероятностью $1 - q$ навсегда покидает систему. Запрос из неприоритетного потока (с параметром $\lambda_2 < \infty$) немедленно начинает обслуживаться одним из свободных приборов, если таковой есть в момент поступления запроса. В противном случае запрос уходит на орбиту. Предполагается, что суммарная интенсивность повторных попыток с орбиты является функцией α_i числа запросов i на орбите такой, что $\alpha_0 = 0, \lim_{i \rightarrow \infty} \alpha_i = \infty$. Время обслуживания любого запроса распределено по экспоненциальному закону с параметром $\mu, 0 < \mu < \infty$.

Функционирование системы описывается случайным процессом $\xi_t = \{i_t, n_t\}, t \geq 0$, где i_t — число запросов на орбите, n_t — общее число запросов на приборах и в буфере в момент времени t , $i_t \geq 0, n_t = \overline{0, N + R}$. Этот процесс является асимптотически квазитеплицевой цепью Маркова ([3]). С использованием результатов [3] исследовано стационарное поведение этой цепи. Достаточным условием эргодичности $\xi_t, t \geq 0$ является выполнение следующего неравенства:

$$\rho_2 \frac{1 - \rho_1^{R+1}}{1 - \rho_1} + q \rho_1^{R+1} < 1, \text{ где } \rho_k = \lambda_k / N\mu, k = 1, 2.$$

Для нахождения стационарного распределения цепи Маркова $\xi_t, t \geq 0$ используется алгоритм, описанный в [3]. Алгоритм реализован при помощи средств ППП «SIRIUS++» [1].

Литература

1. Dudin A.N., Klimenok V.I. and Tsarenkov G.V. Software “SIRIUS” for performance evaluation of modern communication networks // Modelling and Simulation 2002. 16th European Simulation Multi-conference. Darmstadt. P. 489-493.
2. Falin G.I. and J.G.C. Templeton. Retrial Queues. Chapman&Hall, London, 1997.
3. Klimenok V.I., Dudin A.N. Continuous-time asymptotically quasi-toeplitz Markov chain // Queues: Flows, Systems, Networks, 2005.- V. 18.- P. 77-85.

A multi-server retrial queue with finite buffer, infinite orbit and priority customers is considered. The ergodicity condition and algorithm for calculating the stationary distribution is presented.

СИСТЕМА ОБСЛУЖИВАНИЯ ВМАР|РН|N|0 С РАНДОМИЗИРОВАННОЙ СТРАТЕГИЕЙ ПРИНЯТИЯ ПАКЕТОВ ЗАЯВОК

Орловский Д.С.

Белорусский государственный университет, Минск

Системы массового обслуживания со стационарным пуассоновским входным потоком долгое время являлись основным средством моделирования реальных телекоммуникационных сетей и их фрагментов. Среди результатов их исследования важное место занимает полученная Эрлангом для системы $M|M|N|0$ формула расчёта вероятности отказа в обслуживании. Однако, с внедрением цифровых технологий потоки данных в современных сетях уже нельзя считать ни ординарными, ни стационарными, ни потоками без последовательности. В таких условиях формула Эрланга не обеспечивает достаточную точность, что вынуждает изучать более соответствующие действительности системы.

Обобщения системы Эрланга на случай коррелированных входных потоков, характерных для современных телекоммуникационных сетей, исследовались, например, в [1,2]. В [1] был предложен алгоритм для вычисления стационарных вероятностей состояний системы и вероятности отказа для системы ВМАР|РН|N|0 с групповым марковским входным потоком (Batch Markovian Arrival Process, или ВМАР, подробнее см., например, [3]) и фазовым (Phase-type или РН-type, см., например, [4]) распределением времени обслуживания. Рассматривались три возможные стратегии принятия групп заявок: частичное принятие, полное принятие и полный отказ.

Для построения процесса изменений состояний системы при этом отслеживался номер фазы РН-обслуживания, на которой в текущий момент находится каждый из занятых приборов. Максимальный порядок участвующих в вычислениях матриц в [1] совпадает с порядком наибольшего из блоков инфинитезимального генератора (ИГ) системы и равен $\bar{K} = \bar{W}M^N$, где \bar{W} — число состояний управляющего процесса ВМАР, M — число фаз РН-обслуживания. При $M > 1$ это делает невозможным расчёты при сколь-либо существенных N : при $\bar{W} = 3, M = 3, N = 25$ мы имеем $\bar{K} \approx 2.5 \cdot 10^{12}$, а при $N = 30$, соответственно, $\bar{K} \approx 6.2 \cdot 10^{14}$. Вычисления с матрицами такой размерности практически неосуществимы.

Альтернативный подход к исследованию систем с идентичными обслуживающими приборами, предложен Рамасвами в работе [5]. При этом во всякий момент для каждой из возможных фаз РН-обслуживания отслеживается число занятых приборов, находящихся на этой фазе.

В настоящей работе с использованием подхода Рамасвами рассматривается следующая система.

Входящий ВМАР-поток, определяется неприводимой цепью Маркова с непрерывным временем (ЦМНВ) $\nu, t \geq 0$ с пространством состояний $\{0, \dots, W\}$ (т.н. управляющим процессом), и набором матриц $D_k, k \geq 0$, элементы которых характеризуют интенсивности переходов $\nu, t \geq 0$, приводящих к генерации пакетов из k заявок. Пусть $D(z) = \sum_{k=0}^{\infty} D_k z^k, |z| \leq 1$. Матрица $D(1)$ — ИГ процесса $\nu, t \geq 0$. Здесь и далее \vec{e} — вектор-столбец, состоящий из единиц, размер которого определяется из контекста.

В системе имеется N идентичных обслуживающих приборов. Распределение времени обслуживания — фазовое, определяется управляющим процессом $m, t \geq 0$, ЦМНВ с пространством состояний $\{0, \dots, M\}$, где состояние 0 — поглощающее. Завершению обслуживания соответствует попадание в поглощающее состояние. Начальное состояние определяется вероятностным вектором-строкой $(0, \vec{\beta})$. ИГ процесса $m, t \geq 0$

есть матрица $\tilde{S} = \begin{pmatrix} 0 & 0 \\ \vec{S}_0 & S \end{pmatrix}$, где $\vec{S}_0 = -S\vec{e}$.

Если в момент поступления группы заявок все приборы заняты, группа покидает систему. Если имеется достаточное количество свободных приборов, каждая из заявок группы занимает произвольный свободный прибор. Если свободные приборы есть, но их недостаточно для приёма всей группы, с некоторой наперёд заданной вероятностью \tilde{p} вся группа покидает систему, и с вероятностью $1 - \tilde{p}$ часть заявок группы занимает все свободные приборы, а остальные заявки остаются в системе ожидать обслуживания. Отметим, что данная система ведёт себя как рассмотренная в [1] система со стратегией полного принятия при $\tilde{p} = 0$, и полного отказа при $\tilde{p} = 1$.

Под состоянием системы в момент t будем понимать вектор (i, v, \vec{h}_t) , где i — число заявок в системе, v — состояние управляющего процесса ВМАР, $\vec{h}_t = (h_t^{(1)}, \dots, h_t^{(M)})$ — вектор, компонента $h_t^{(m)}$ которого есть число занятых приборов, находящихся на m -й фазе обслуживания в момент t . Процесс изменения состояний системы есть многомерная ЦМНВ $\xi_t = (i_t, v_t, \vec{h}_t)$, $i_t \geq 0, v_t = \overline{0, W}, m = \overline{1, M}, h_t^{(1)} + \dots + h_t^{(M)} = i_t, t \geq 0$.

Пусть $p(i, v, h^{(1)}, \dots, h^{(M)}) = \lim_{t \rightarrow \infty} p(i_t, v_t, h_t^{(1)}, \dots, h_t^{(M)})$. Упорядочим состояния $\xi_t, t \geq 0$ в лексикографическом порядке по компонентам i, v . Состояния с равными значениями i и v в свою очередь доупорядочим в обратном лексикографическом порядке по компонентам $h^{(m)}$ вектора \vec{h} . Такое совмещённое использование прямого и обратного лексикографического порядков необходимо, чтобы с минимальными изменениями задействовать алгоритм из [1], и при этом иметь возможность использовать результаты работ [5,6]. Введём вектор-строки $\vec{p}_i, i \geq 0$ стационарных вероятностей, соответствующих значению i первой компоненты процесса $\xi_t, t \geq 0$ и вектор $\vec{p} = (\vec{p}_0, \vec{p}_1, \dots)$. Вектор \vec{p} удовлетворяет системе уравнений баланса: $\vec{p}Q = \vec{0}, \vec{p}\vec{e} = 1$, где Q — ИГ ЦМ $\xi_t, t \geq 0$.

Найдём для состояния $z = (i, v, h^{(1)}, \dots, h^{(M)})$ процесса $\xi_t, t \geq 0$ его номер p_z при таком упорядочении. Нумерацию начинаем с нуля. Пусть $L(\xi, \epsilon, \bar{h}^{(1)}, \dots, \bar{h}^{(k)})$ — число состояний вида $(\xi, \epsilon, \bar{h}^{(1)}, \dots, \bar{h}^{(k)}, h^{(k+1)}, \dots, h^{(M)})$, $\xi \geq 0, \epsilon = \overline{0, W}, k = \overline{0, M-1}, \bar{h}^{(1)} + \dots + \bar{h}^{(k)} \leq \min\{\xi, N\}$. Можно показать, что эта величина не зависит от ϵ и равна $L(\xi, \epsilon, \bar{h}^{(1)}, \dots, \bar{h}^{(k)}) = L(\xi, \bar{h}^{(1)}, \dots, \bar{h}^{(k)}) = C_{\min\{\xi, N\} - (\bar{h}^{(1)} + \dots + \bar{h}^{(k)})}^{M-k-1}$, тогда $p_z = \vec{W} (C_{\min\{i, N\} + M - 1}^M + \max\{i - 1 - N, 0\} C_{N + M - 2}^M) + v C_{\min\{i, N\} + M - 1}^{M-1} + \sum_{k=1}^{M-1} C_{\min\{i, N\} - (h_1 + \dots + h_k + 1) + M - k}^{M-k}$.

Определим интенсивности переходов между состояниями процесса $\xi_t, t \geq 0$. Пусть в момент t он находится в состоянии $z_t = (i_t, v_t, \vec{h}_t)$. Рассмотрим всевозможные события, которые могут произойти за малый промежуток времени $[t, t + \delta]$ с вероятностью, превосходящей $o(\delta)$: (1) переход управляющего процесса ВМАР-потока v из состояния v в состояние v' с генерацией пакета из k заявок. (2) Переход процесса v из состояния v в состояние v' без генерации заявок. (3) Изменения номера фазы РН-обслуживания на одном из занятых приборов с m на m' без завершения обслуживания. (4) Завершение обслуживания прибором, бывшим на m -й фазе РН-обслуживания.

Интенсивности этих событий и устанавливающиеся в результате их наступления состояния приведены в таблице 1. Символом \vec{e}_l обозначается вектор-строка размера M , l -й элемент которого равен 1, а остальные элементы — нули.

Таблица 1. Интенсивности возможных переходов

событие	$z_{t+\Delta}$	интенсивность перехода	Условие
1	$(i+k, v', \vec{h} + \vec{e}_1 + \dots + \vec{e}_k)$ $(i+k, v', \vec{h} + \vec{e}_1 + \dots + \vec{e}_{N-i})$	$(D_k)_{v, v'} \vec{\beta}_1 \dots \vec{\beta}_k$ $(1 - \tilde{p})(D_k)_{v, v'} \vec{\beta}_1 \dots \vec{\beta}_{N-i}$	$i < N, 1 \leq k \leq N - i$ $i < N, k > N - i$
1,2	(i, v', \vec{h}) (i, v', \vec{h})	$(D_0 + \tilde{p} \sum_{k=N-i+1}^{\infty} D_k)_{v, v'}$ $(D(1))_{v, v'}$	$i < N, v \neq v'$ $i \geq N, v \neq v'$
3	$(i, v', \vec{h} - \vec{e}_m + \vec{e}_{m'})$	$\vec{h}^{(m)} S_{m, m'}$	$m \neq m', \vec{h} \vec{e}_m \geq 1$
4	$(i-1, v, \vec{h} - \vec{e}_m)$ $(i-1, v, \vec{h} - \vec{e}_m + \vec{e}_{m'})$ $(i+1, v, \vec{h})$	$\vec{h}^{(m)} (\vec{S}_0)_m$ $\vec{h}^{(m)} (\vec{S}_0)_m \vec{\beta}_m$ $\sum_{r=1}^M \vec{h}^{(r)} (\vec{S}_0)_r \vec{\beta}_r$	$1 \leq i \leq N, \vec{h} \vec{e}_m \geq 1$ $i > N, \vec{h} \vec{e}_m \geq 1, m \neq m'$ $i > N$

Опираясь на таблицу 1, ИГ Q процесса $\xi_t, t \geq 0$ можно построить поэлементно, однако удобнее привлечь для этого предложенный Лукантони и Рамасвами в работах [5,6] алгоритм.

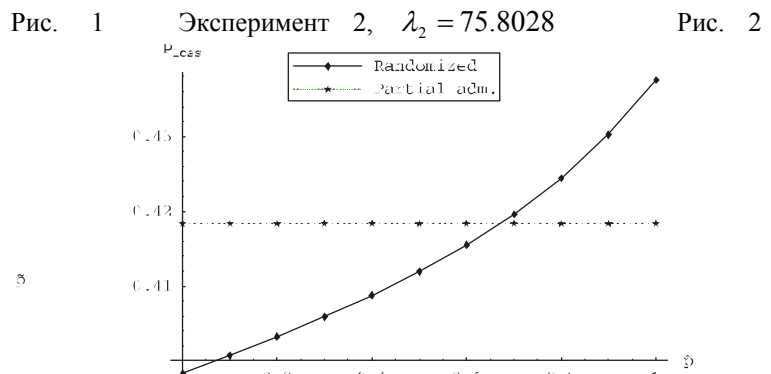
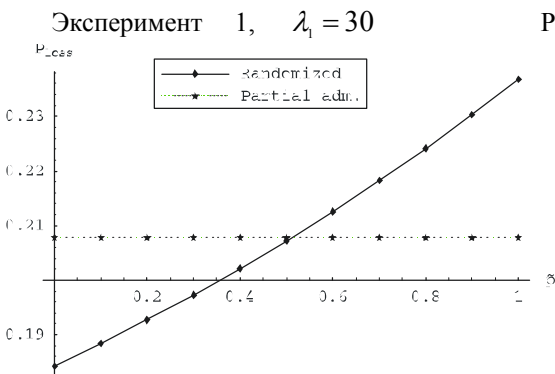
Лемма. ИГ процесса $\xi_t, t \geq 0$ имеет блочную структуру $Q = (Q_{ij}), i, j \geq 0$. Ненулевые блоки Q_{ij} размера $\overline{WL}(i) \times \overline{WL}(j)$ определяются следующим образом:

$$\begin{aligned} Q_{i,i-1} &= I_{\overline{w}} \otimes L_{N-i}(N, \tilde{S}), i = \overline{1, N}, & Q_{i,i-1} &= I_{\overline{w}} \otimes Q(N, \tilde{S}_0, \tilde{\beta}), i \geq N+1, \\ Q_{ii} &= \left(D_0 + \tilde{p} \sum_{k=0}^{\infty} D_{N-i+1+k} \right) \oplus A(i, S) + \Delta_i, i = \overline{0, N-1}, & Q_{ii} &= D(1) \oplus A(N, S) + \Delta_{\min\{i, N+1\}}, i \geq N, \\ Q_{ij} &= D_{j-i} \otimes P_{i,j}(\tilde{\beta}), i = \overline{1, N-1}, j = \overline{i+1, N}, & Q_{ij} &= (1 - \tilde{p}) D_{j-i} \otimes P_{i,N}(\tilde{\beta}), i = \overline{1, N-1}, j \geq N+1, \end{aligned}$$

где символы \otimes и \oplus обозначают кронекерово произведение и сумму матриц, $P_{i,j}(\tilde{\beta}) = P_i(\tilde{\beta}) \cdot \dots \cdot P_{j-1}(\tilde{\beta}), j \geq i$. Матрицы $A(c, T), Q(c, T), L_k(c, T)$ вводятся в [5], а $P_k(T)$ — в [6].

После того, как построен ИГ Q процесса $\xi_t, t \geq 0$, систему уравнений баланса можно решить, например, при помощи предложенного в [1] алгоритма. Вероятность потери заявки P_{loss} определяется на основании стационарных вероятностей $\tilde{p}_i, i \geq 0$ по формуле: $P_{loss} = 1 - \lambda^{-1} \sum_{i=1}^{\infty} \tilde{p}_i Q_{i,i-1} \tilde{e}$.

Были произведены следующие два вычислительных эксперимента. Зафиксировано распределение Эрланга второго порядка времени обслуживания, имеющее интенсивность $\mu = 2$, ВМАР с максимальной длиной группы заявок, равной 50, тремя состояниями управляющего процесса, коэффициентом корреляции $c_{corr} = 0.3$ и коэффициентом вариации $c_{var} = 2$. Число приборов $N = 50$. Интенсивность ВМАР в первом эксперименте равна $\lambda_1 = 30$, во втором $\lambda_2 = 75.8028$. При расчёте по формуле Эрланга с учётом лишь числа приборов и загрузки, для этих систем были бы получены вероятности отказа $P_{Erl}^1 \approx 6.4 \cdot 10^{-13}$ и $P_{Erl}^2 = 0.01$ соответственно. Вероятность отказа группе в обслуживании \tilde{p} изменялась от 0 до 1 с шагом 0.1. Результаты экспериментов представлены на рис. 1,2. Горизонтальные линии на графиках — вероятности отказа, полученные ранее для этих же систем в случае дисциплины частичного принятия $P_{PA}^1 = 0.207768$ и $P_{PA}^2 = 0.4183518$.



Литература

1. Klimenok V.I., Kim C.S., Orlovsky D.S., Dudin A.N. Lack of invariant property of Erlang BMAP|PH|N|0 model // Queueing Systems (accepted 10.04).
2. Krieger U.R., Naumov V. Analysis Of A Versatile Queuing Model With State-Dependent Service Times // MMB, 1999.- P. 121-135.
3. Lucantoni D. New results on the single server queue with a batch Markovian arrival process // Comm. Statist.-Stochastic Models, 1991.- V. 7.- P. 1-46.
4. Neuts M. Matrix-geometric solutions in stochastic models. Baltimore: John Hopkins University Press, 1981.
5. Ramaswami V. Independent Markov Processes in parallel // Comm. Statist.-Stochastic Models, 1985.- V. 1(3).- P. 419-432.
6. Ramaswami V. and Lucantoni D.M. Algorithms for the multi-server queue with phase-type service. Independent Markov Processes in parallel // Comm. Statist.-Stochastic Models, 1(3), 1985.- V. 1(3).- P. 393-417.

The BMAP|PH|N|0 model with the Batch Markovian Arrival Process (BMAP) and randomized discipline of admission is investigated. Effective algorithm for calculating the loss probability is presented. Impact of the group rejection probability on loss probability is analyzed numerically.

 ВРЕМЯ ПРЕБЫВАНИЯ ТРЕБОВАНИЯ В СИСТЕМЕ SM/MSP/1

Казимирский А.В.

Белорусский государственный университет, г. Минск

1. Математическая модель

Рассмотрим однолинейную систему массового обслуживания с бесконечным буфером. Поступающий в систему поток требований описывается следующим образом. Пусть имеется полумарковский случайный процесс $n_t, t \geq 0$ с пространством состояний $\{1, \dots, N\}$, называемый управляющим процессом входящего потока. Время пребывания процесса n_t в состоянии n является случайной величиной с функцией распределения $A_n(t), n = 1, \dots, N, t \geq 0$ с преобразованием Лапласа-Стилтьеса $A_n^*(s), \operatorname{Re} s > 0$ и математическим ожиданием $a_n^{(1)} = \int_0^\infty t dA_n(t)$. После окончания времени пребывания процесса n_t в состоянии n в систему приходит требование и с вероятностью $p(n, n'), n' = 1, \dots, N$ состояние процесса n_t становится равным n' . Введём следующие матрицы:

$$A(t) = \operatorname{diag}\{A_n(t), n = 1, \dots, N\}, t \geq 0, \quad A^*(s) = \operatorname{diag}\{A_n^*(s), n = 1, \dots, N\}, \operatorname{Re} s > 0,$$

$$A^{(1)} = \operatorname{diag}\{a_n^{(1)}, n = 1, \dots, N\}, \quad P = \begin{pmatrix} p(1,1) & \dots & p(1,N) \\ \dots & \dots & \dots \\ p(N,1) & \dots & p(N,N) \end{pmatrix},$$

Предполагается, что матрица P является стохастической, неприводимой и непериодической. Средняя интенсивность входящего потока требований λ вычисляется следующим образом:

$$\lambda^{-1} = \mathbf{1}A^{(1)}\mathbf{1},$$

где $\mathbf{1}$ - вектор-столбец соответствующей размерности, состоящий из единиц, $\mathbf{1}$ - вектор-строка, являющийся единственным решением следующей системы линейных уравнений:

$$\mathbf{1}P = \mathbf{1}, \quad \mathbf{1}\mathbf{1} = 1.$$

Заданный таким образом входящий поток требований можно называть “полумарковским входящим потоком требований”.

Процесс обслуживания задаётся следующим образом. Пусть имеется случайный процесс $m_t, t \geq 0$ с пространством состояний $\{1, \dots, M\}$, называемый управляющим процессом обслуживания. Время пребывания процесса m_t в состоянии $m = 1, \dots, M$ имеет показательное распределение. Поведение процесса m_t зависит от того, является ли в момент времени t система пустой или нет. Так, если система не является пустой, то время пребывания процесса m_t в состоянии m имеет показательное распределение с параметром $\lambda_m^{(service)}$, и по его окончании с вероятностью $p_k^{(service)}(m, m'), m = 1, \dots, M$ систему покидает k , $k = 0, 1$ требований, а следующим состоянием процесса m_t становится $m', m' = 1, \dots, M$. Иначе, если система является пустой в момент времени t , то время пребывания процесса m_t в состоянии m имеет показательное распределение с параметром $\lambda_m^{(idle)}$, и по его окончании с вероятностью $p^{(idle)}(m, m'), m = 1, \dots, M$ следующим состоянием процесса m_t становится $m', m' = 1, \dots, M$. Предполагается, что $p_0^{(service)}(m, m) = 0, p^{(idle)}(m, m) = 0, m = 1, \dots, M$. Информацию о параметрах обслуживания удобно хранить в квадратных матрицах $S_k^{(service)}, k = 0, 1$ и $S^{(idle)}$ размерности M , элементы которых задаются следующим образом:

$$\begin{aligned} (S_0^{(service)})_{m,m} &= -\lambda_m^{(service)}, (S^{(idle)})_{m,m} = -\lambda_m^{(idle)}, m = 1, \dots, M, \\ (S_0^{(service)})_{m,m'} &= \lambda_m^{(service)} p_0^{(service)}(m, m'), m \neq m', m, m' = 1, \dots, M, \\ (S^{(idle)})_{m,m'} &= \lambda_m^{(idle)} p^{(idle)}(m, m'), m \neq m', m, m' = 1, \dots, M, \\ (S_1^{(service)})_{m,m'} &= \lambda_m^{(service)} p_1^{(service)}(m, m'), m, m' = 1, \dots, M. \end{aligned}$$

Введённый выше процесс обслуживания является одним из возможных вариантов “марковского процесса обслуживания” (Markovian Service Process – MSP).

2. Распределение числа требований в системе

Для дальнейшего исследования понадобятся следующие вероятности:

$$f_t(i, n, m, x) = \Pr\{i_t = i, n_t = n, m_t = m, x_t \leq x\}, t \geq 0, i \geq 0, n = 1, \dots, N, m = 1, \dots, M, x \geq 0,$$

где $i_t \geq 0$ - число требований в системе в момент времени t , $n_t = 1, \dots, N$ - состояние управляющего процесса входящего потока в момент времени t , $m_t = 1, \dots, M$ - состояние управляющего процесса обслуживания в момент времени t , $x_t \geq 0$ - время оставшееся до окончания пребывания процесса n_t в состоянии n в момент времени t .

Введённые вероятности образуют следующие векторы:

$$\mathbf{f}_t(i, n, x) = (f_t(i, n, 1, x), \dots, f_t(i, n, M, x)), \mathbf{f}_t(i, x) = (\mathbf{f}_t(i, 1, x), \dots, \mathbf{f}_t(i, N, x)), i \geq 0, x \geq 0.$$

Предположим, что в рассматриваемой системе существует стационарное распределение. Обозначим векторы стационарных вероятностей следующим образом:

$$\mathbf{f}(i, x) = \lim_{t \rightarrow \infty} \mathbf{f}_t(i, x), i \geq 0, x \geq 0.$$

Также введём следующие обозначения:

$$\mathbf{f}'(i, x) = \frac{d\mathbf{f}(i, x)}{dx}, x \geq 0, \mathbf{f}^*(i, s) = \int_0^{\infty} d\mathbf{f}(i, x) e^{-sx}, \text{Re } s > 0,$$

$$\mathbf{p}_i = \lim_{x \rightarrow \infty} \mathbf{f}(i, x) = \mathbf{f}^*(i, s)|_{s=0}, i \geq 0, \mathbf{F}^*(z, s) = \sum_{i=0}^{\infty} \mathbf{f}^*(i, s) z^i, \mathbf{F}'(z, 0) = \sum_{i=0}^{\infty} \mathbf{f}'(i, 0) z^i, |z| \leq 1.$$

Используя Δt -метод для марковских процессов (см., например, [1]), составим систему дифференциальных уравнений для вероятностей $\mathbf{f}(i, x)$, $i \geq 0$, $x \geq 0$. После применения к полученным уравнениям преобразования Лапласа-Стилтьеса, умножения каждого из них на z^i , $|z| \leq 1$, $i \geq 1$ и суммирования, получим следующее уравнение.

Лемма 1. Векторные производящие функции $\mathbf{F}^*(z, s)$, $\mathbf{F}'(z, 0)$ и вектор $\mathbf{f}^*(0, s)$ удовлетворяют следующему уравнению:

$$s\mathbf{F}^*(z, s) + \mathbf{f}^*(0, s) \tilde{S}^{(idle)} + (\mathbf{F}^*(z, s) - \mathbf{f}^*(0, s)) \left(z^{-1} \tilde{S}_1^{(service)} + \tilde{S}_0^{(service)} \right) + \mathbf{F}'(z, 0) (z \tilde{P} \tilde{A}^*(s) - I_{NM}) = \mathbf{0}, \quad (1)$$

где

$$\tilde{S}^{(idle)} = I_N \otimes S^{(idle)}, \quad \tilde{S}_k^{(service)} = I_N \otimes S_k^{(service)}, \quad k = 0, 1, \quad \tilde{A}^*(s) = A^*(s) \otimes I_M,$$

$$\tilde{P} = P \otimes I_M.$$

\otimes - символ кронекерова произведения матриц, I - единичная матрица соответствующей размерности, $\mathbf{0}$ - вектор-строка соответствующей размерности, состоящий из нулей.

Следствие 1. Выполняется следующее соотношение:

$$\lambda = \mathbf{F}'(1, 0) \mathbf{1} = \sum_{i=1}^{\infty} \mathbf{p}_i \tilde{S}_1^{(service)} \mathbf{1}. \quad (2)$$

3. Распределение времени пребывания требования в системе

Для дальнейшего исследования понадобятся следующие вероятности:

$$w_t(m, x) = \Pr\{m_{t+\phi_t} = m, \phi_t \leq x\}, \quad m = 1, \dots, M, \quad x \geq 0, \quad t \geq 0,$$

где $\phi_t, t \geq 0$ - случайная величина времени пребывания требования, прибывшего в систему в момент времени t . Введённые выше вероятности образуют следующие векторы:

$$\mathbf{w}_t(x) = (\mathbf{w}_t(1, x), \dots, \mathbf{w}_t(M, x)), \quad t \geq 0, \quad \mathbf{w}(x) = \lim_{t \rightarrow \infty} \mathbf{w}_t(x), \quad x \geq 0.$$

Будет исследоваться следующее преобразование Лапласа-Стилтьеса:

$$\mathbf{W}^*(s) = \int_0^{\infty} d\mathbf{w}(x) e^{-sx}, \quad \operatorname{Re} s > 0.$$

Используя метод катастроф (см., например, [2]), и учитывая (2) можно получить следующее выражение для нахождения преобразования Лапласа-Стилтьеса времени пребывания $\mathbf{W}^*(s)$.

Лемма 2. Преобразование Лапласа-Стилтьеса $\mathbf{W}^*(s), \operatorname{Re} s > 0$ вычисляется следующим образом:

$$\lambda \mathbf{W}^*(s) = \sum_{i=0}^{\infty} \mathbf{f}'(i, 0) \sum_{k=1}^{\infty} P_k(\mathbf{1}_N \otimes I_M) \sum_{l=1}^k (Z(s))^{i+l}, \quad \operatorname{Re} s > 0. \quad (3)$$

Учитывая (1), после некоторых преобразований выражение (3) может быть приведено к следующему виду.

Теорема. Преобразование Лапласа-Стилтьеса времени пребывания требования вычисляется следующим образом:

$$\lambda \mathbf{W}^*(s) \mathbf{1} = \lambda - s \sum_{i=1}^{\infty} \mathbf{p}_i (\mathbf{1}_N \otimes I_M) \sum_{j=0}^{i-1} (Z(s))^j \mathbf{1}, \quad (4)$$

где $Z(s) = (sI_M - S_0^{(service)})^{-1} S_1^{(service)}, \operatorname{Re} s > 0$.

Из (4) непосредственно вытекает следующее соотношение.

Следствие 2. Справедливо следующее соотношение:

$$\lambda W = L,$$

где $W = -\left. \frac{d\mathbf{W}^*(s)}{ds} \right|_{s=0} \mathbf{1}$ - среднее время пребывания требования в системе,

$L = \sum_{i=1}^{\infty} i \mathbf{p}_i \mathbf{1}$ - среднее число требований в системе в произвольный момент времени.

Литература

1. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания // М.: Наука. – 1987. – С. 336.
2. Климов Г. П. Стохастические системы обслуживания // М.: Наука. – 1966. – С. 244.

SOJOURN TIME IN SM/MSP/1 SYSTEM

Kazimirsky A.

Belarusian State University, Minsk, Republic of Belarus.

In the given article the SM/MSP/1 system is considered. For the considered system, the Laplace-Stieltjes transform of sojourn time is got, based on which the proof of Little's formula can be easily derived.



ДВОИЧНЫЕ ГИПЕРКУБИЧЕСКИЕ СЕТИ С ДИАГОНАЛЯМИ

Гармаш В.А., Сергеева О.Ф

Институт проблем передачи информации РАН

Одним из видов сетей, используемых в информационных системах, являются гиперкубические сети B_n [1]. В гиперкубической сети источники информации располагаются в вершинах куба, а ребра, соединяющие вершины, являются каналами связи. В частном случае двоичная гиперкубическая сеть является булевым n -мерным кубом.

Основные предположения о работе сети B_n являются следующими:

1. Каждый узел, входящий в сеть B_n , с одинаковой вероятностью может вступать в обмен информацией с любым другим узлом сети, включая и самого себя.
2. Каждый узел может играть посредника при обмене информацией между другими узлами. В этом случае он получает информацию от соседнего узла и передает ее другому узлу по смежному каналу. При этом один и тот же узел может одновременно получать и передавать информацию по нескольким каналам.
3. По каждому каналу информация может передаваться в обоих направлениях.
4. По одному каналу одновременно может передаваться информация только между двумя узлами.

Математическая модель сети. Математической моделью является граф, вершины которого соответствуют источникам информации, а ребра между ними – каналам. Число соединений между любыми чередующимися вершинами A и B называется расстоянием $\rho(A, B)$. Расстояние между вершинами $A(\alpha_1, \alpha_2, \dots, \alpha_n)$ и

$B(\beta_1, \beta_2, \dots, \beta_n)$ равно $c(A, B) = \sum_{i=1}^n |\alpha_i - \beta_i|$. Диаметром сети называется наибольшее расстояние между вершинами A и B . Он равен $d(B_n) = n$. Среднее расстояние между вершинами равно $l(B_n) = n/2$.

Функционирование сети B_n и алгоритм установления соединений. Функционирование сети состоит в том, что в дискретные моменты времени возникают требования на переход от одного имеющегося состояния C к новому состоянию C' , в котором либо отсутствует одно указанное в требовании соединения состояния C , либо к соединениям состояния C должно быть добавлено заданное соединение. Одной из важнейших характеристик функционирования сети является время перехода из одного состояния в соседнее, причем время перехода зависит от длины соединения. Поэтому эти длины должны быть минимальными, т.е. равны расстояниям между вершинами. Между двумя вершинами гиперкуба, расстояние между которыми равно l , существует $l!$ соединений длины l . Алгоритм построения соединений должен быть однозначным, т.е. должен строить одно из них. Но при этом могут возникать коллизии, в которых строящееся соединение пересекает одно из имеющихся соединений состояния C . В таком случае алгоритм должен принимать решение об отказе в построении соединения т.е. соединение должно быть заблокировано.

Значения средней длины соединения в сетях B_n и B_n^* при некоторых значениях n функционирование гиперкубической сети с диагоналями B_n^* . Рассматривается сеть B_n^* , полученная из гиперкубической сети B_n

введением дополнительных каналов по диагоналям между каждыми двумя противоположными вершинами $A = (\alpha_1, \alpha_2, \dots, \alpha_n)$ и $A' = (\alpha'_1, \alpha'_2, \dots, \alpha'_n)$, где $\alpha'_i = 1 - \alpha_i, i = 1, \dots, n$.

От алгоритма построения соединений в сети B_n^* потребуем, чтобы он обладал теми же качествами, что и алгоритм для сети B_n . Это значит, что он должен действовать в любом состоянии сети, строить одно определенное соединение, длина которого равна расстоянию, если же такое соединение относительно его заблокировано, сообщить об этом и прекратить работу.

Алгоритм для сети B_n^* работает следующим образом. Пусть в состоянии C сеть B_n^* находится в состоянии $A(\alpha_1, \alpha_2, \dots, \alpha_i \dots \alpha_n)$ и $B(\beta_1, \beta_2, \dots, \beta_i, \dots, \beta_n)$. Находится значение $\delta = \sum_{j=1}^n |\alpha_j - \beta_j|$. Если $\delta > n/2$, то установка соединений осуществляется по алгоритму для сети B_n , если $\delta \leq n/2$, то соединение производится по диагонали.

Двоичная гиперкубическая сеть B_n^* обладает следующими свойствами.

1. Множество всех вершин сети B_n^* по отношению к заданной вершине A можно разбить на два непересекающихся подмножества M_1 и M_2 : если построенное по алгоритму соединение вершины A с вершиной B не проходит через диагональ (A, A_1) , то $B \in M_1$, в противном случае $B \in M_2$.

2. Длина соединения L , устанавливаемого алгоритмом, выражается формулой:

$$l(L) = \min \left(\sum_{j=1}^n |\alpha_j - \beta_j|, 1 + \sum_{j=1}^n |1 - \alpha_j - \beta_j| \right)$$

3. Наибольшая длина соединений, устанавливаемых алгоритмом, $\leq \left\lfloor \frac{n}{2} \right\rfloor$.

4. Множество M_l всех вершин сети, расстояние которых от данной вершины A равно l , состоит из двух непересекающихся подмножеств $M_{l1} = M_l \cap M_1$ и $M_{l2} = M_l \cap M_2$.

5. Среднее значение $l_{cp}(B_n^*)$ равно

$$l_{cp}(B_n^*) = \frac{n+1}{2} \left(1 - 2^{-n} C_n^{\lfloor \frac{n}{2} \rfloor} \right)$$

Воспользовавшись формулой Стирлинга, получим:

$$l_{cp}(B_n^*) = \frac{n}{2} - \frac{1}{2} \left(\sqrt{\frac{2n}{\pi}} - 1 \right) + O\left(\frac{1}{\sqrt{n}}\right)$$

Фактически относительная погрешность уже при $n \geq 5$ не превосходит 4%, а при $n \geq 10$ – 2%. Значения средней длины соединения при некоторых значениях n :

N	3	4	5	6	7	8	9	10	20
$l_{cp}(B_n)$	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	10,0
$l_{cp}(B_n^*)$	1,3	1,6	2,1	2,4	2,9	3,3	3,8	4,2	8,6

6. Вероятности использования в соединениях некоторого ребра e и некоторой диагонали d (P_e и P_d , соответственно) равны $P_e = 2^{-n} - 2^{-(2n-1)} C_{n-1}^{\lfloor \frac{n}{2} \rfloor - 1}$, $P_d = 2^{-n} - 2^{-2n} C_n^{\frac{1}{2}n}$ (при четном n) и

$P_d = 2^{-n} - 2^{-(2n-1)} C_n^{\frac{n-1}{2}}$ (при нечетном n). Отсюда получаем (например, для n четного):

$$\frac{P_d}{P_e} = 1 - \frac{C_{n-1}^{\frac{n-1}{2}}}{2^{n-1} - C_{n-1}^{\frac{n-1}{2}}}$$

С помощью формулы Стирлинга имеем:

$$\frac{P_d}{P_e} \cong 1 - \sqrt{\frac{2}{\pi(n-1)}}. \text{ При нечетном } n \lim_{n \rightarrow \infty} \frac{P_d}{P_e} = 1.$$

Заключение

1. Максимальная длина соединений в сети B_n^* вдвое меньше, чем в сети B_n
2. Средняя длина соединений в сети B_n^* при реальных значениях n на 15-20% меньше, чем в сети B_n .
3. Вероятность использования канала в сети B_n^* несколько меньше, чем в сети B_n . При четных n все каналы сети B_n^* используются с равными вероятностями. При нечетных n вероятность использования диагональных каналов, хотя и не на многое, но меньше, чем вероятность каждого из остальных, и они асимптотически равны.
4. Дальнейшее обобщение результатов для двоичных гиперкубических сетей м.б. достигнуто за счет добавления диагональных каналов в последовательных подкубах, составляющих основной гиперкуб.

Литература

1. Wittie L.D. Commutation Structures for Large Networks of Microcomputers // IEEE Trans. Comput. 1981. V. 30, N 4. P. 264-273

BINARY HYPERCUBIC NETS WITH DIAGONALS

Hypercubic nets and binary hypercubic nets with diagonals are discussed. Peculiarities of functioning of such nets, algorithms of setting of commutations are described. The main characteristics of the sets are obtained: the commutation path length, the probabilities of use of commutation edges and diagonals. The efficiency of use of binary hypercubic nets with diagonals is marked.

