В. А. Малышев, О математических моделях сетей обслуживания, *Автомат. и телемех.*, 2009, выпуск 12, 9–15

Автоматика и телемеханика, № 12, 2009

Некоторые направления развития теории очередей

PACS 89.75.Fb

© 2009 г. В.А. МАЛЫШЕВ, д-р физ.-мат. наук
(Московский государственный университет им. М.В. Ломоносова)

О МАТЕМАТИЧЕСКИХ МОДЕЛЯХ СЕТЕЙ ОБСЛУЖИВАНИЯ

Теория очередей как область теории вероятностей имеет классические постановки задач и сложившийся математический аппарат. В этом кратком обзоре прослеживаются ее связи с другими областями математики и другими прикладными задачами.

1. Введение

Начиная со знаменитой статьи Эрланга [1] теория очередей воспринимается как область теории вероятностей, однако помимо математического содержания теория очередей имеет свою интуицию и идеологию. Поэтому можно говорить о взаимном влиянии математики и теории очередей друг на друга. Без математики не было бы теории очередей, а без теории очередей вряд ли так быстро развились некоторые области математики. Основные математические подходы к теории очередей и ее связи с такими областями, как, например, теория надежности, хорошо известны и вошли в учебники. Здесь кратко описываются в доступной форме менее известные связи теории очередей с другими областями математики и важными прикладными областями.

Проблема связи математики с приложениями только одна: рекламируемая практическая бизнес-полезность математической работы обычно обратно пропорциональна ее математическим достоинствам. Однако это вовсе не значит, что математикам не надо вникать в практическую суть проблемы. Прежде всего, это надо для самой математики (нельзя вариться в собственном соку), кроме того, при участии математиков возникают формальные языки, качественные и количественные методы, которые затем перевариваются для практических нужд. Сети обслуживания (очередей, связи и т.д.) встречаются в таком числе прикладных задач, что естественно искать математический язык для их описания, не привязанный ни к какой прикладной задаче. Такая цель в полном объеме, конечно, не выполнима, так как каждая область имеет свою индивидуальность, но частично объединяющий язык должен быть.

Теория очередей начинается с простейшей ситуации, когда за малое время dt в очередь длины n прибывает одно требование с вероятностью $\lambda_n dt$ и убывает (обслуживается) одно требование с вероятностью $\mu_n dt$, $n > 0$, что соответствует простому неоднородному случайному блужданию на Z_+. Положение точки в момент t соответствует длине очереди. Однако если требования перенумеровать в порядке поступления, то возникают новые задачи, связанные, например, с временем ожидания требований. И ответ будет зависеть от дисциплины обслуживания. Привычными и банальными (что не означает, конечно, простоты этих задач) обобщениями этой ситуации является замена экспоненциальных распределений на более общие.

С математической точки зрения все задачи теории очередей можно разделить на три большие группы: 1) классика теории очередей – одна очередь, примерами недавних больших исследований являются [2–4], уже эти работы показывают математическое богатство одноканальных систем, а работ в действительности много больше, 2) малое число очередей или узлов обслуживания, 3) большое число узлов обслуживания, по которым циркулирует большое число требований.

Внутри этих классов должна быть классификация соответственно структуре сетей, дисциплинам обслуживания и тому, что в технике связи называется протоколами.

2. Малые системы

Это значит, что число S типов требований и число N очередей невелики, так что не надо делать предельные переходы N, $S \to \infty$. Для удобного описания множества частных случаев можно использовать привычные математику языки. Так, N очередей с одним типом требований удобно описывать на языке случайных блужданий в ортанте Z_+^N. Состоянием системы является целочисленный вектор $(n_1, \dots, n_N) \in Z_+^N$ длин очередей. Любой переход

$$(n_1, \dots, n_N) \to (n'_1, \dots, n'_N)$$

может быть проинтерпретирован на языке очередей. Например, для $N = 2$ переходы

$$(n_1, n_2) \to (n_1 + 1, n_2 + 1), \quad (n_1, n_2) \to (n_1 - 1, n_2 - 1)$$

означают одновременное поступление (обслуживание) в каждой из очередей. Математические подходы к таким блужданиям разбиваются на две четко очерченные области: аналитические методы и качественные вероятностные методы.

Явные решения – обратимость. Для изучения простого одномерного блуждания на Z_+ можно использовать его обратимость. Понятие обратимости и его обобщения были использованы для построения интересных модельных примеров для S, $N > 1$, см. [5–7].

Явные решения – задача Римана-Гильберта и алгебра очередей. Для одной очереди с одним типом требований и независимыми поступлениями в середине прошлого века был очень популярен подход, связанный с граничной задачей Римана-Гильберта или уравнениями Винера-Хопфа.

Однако до работ автора не было никаких продвижений для двухканальных систем. Здесь возникло два разных подхода. Первый основан на исследовании уравнений для производящих функций в комплексной области. Этот метод позволил многое, если не все, понять для простых блужданий в Z_+^2. И, главное, понять почему есть такая разница между одной и двумя очередями. Именно этот случай явился лабораторией для получения информации о размерности, большей 1. А точнее, были рассмотрены задачи явного нахождения стационарных вероятностей, их асимптотика, большие уклонения, граница Мартина, скорость сходимости к равновесию, см. обзор [8], книги [9, 10], а также статьи [11–13].

Второму аналитическому подходу в одноканальной ситуации соответствует факторизационное тождество Спицера, работы Кингмана-Венделя и др. В больших размерностях пока единственно возможным является использование рекуррентной процедуры факторизации операторных функций со значениями в алгебрах теплицевых или подобных операторов, см. обзор [8], что сводит задачу к решению уравнений вида $(1 + K)f = g$, где K – компактный оператор.

Однако никакие аналитические методы не позволяют получать качественной информации в размерности 3 и выше, хотя и дают процедуру численного решения.

Функции Ляпунова и динамические системы. Для получения качественной информации – условий эргодичности, устойчивости, т.е. непрерывной или аналитической зависимости от параметров, больших уклонений, хорошо приспособлен вероятностный подход, основанный на геометрической идее построения функции Ляпунова. Эта идея идет как от функций Ляпунова для обыкновенных дифференциальных уравнений, так и от критерия Фостера эргодичности счетной цепи Маркова, см. введение в эти методы в [14]. Однако теория пошла гораздо дальше. Если построение функций Ляпунова в размерности 2 и 3, а также для частных случаев произвольной размерности [16] геометрично и достаточно просто [15], то уже в размерности 4 это потребовало развития специальной теории. Такая теория была предложена в [17]. Она основана на соответствии между случайными блужданиями в N-мерном ортанте и детерминированными динамическими системами на сфере размерности $N-2$. Это и позволило получить полную классификацию для блужданий в Z_+^4, см. [18], используя классификацию Пуанкаре-Бендиксона динамических систем на двумерной сфере.

Задача о больших уклонениях в двумерном случае в нестационарном режиме также доступна вероятностным методам, см. [19].

Несколько типов. Если есть $S > 1$ типов требований, то все очень сильно зависит от дисциплины обслуживания. Например, ситуация, когда есть две очереди и два типа требований, одно из которых имеет абсолютный приоритет, сводится к блужданию в Z_+^4, где есть полная классификация. Однако уже случай $N = 1$, $S > 1$ с дисциплинами FIFO или LIFO не сводится к блужданиям. Здесь очередь является словом из S символов, т.е.

$$(1) \qquad \alpha = s_1 \dots s_n, \quad s_i \in \{1, \dots, S\}.$$

При дисциплинах LIFO и FIFO может приходить и уходить s_1 или s_n. При более сложных протоколах обслуживаемое требование может случайно выбираться из очереди. При параллельных вычислительных процедурах часть слова, т.е. некоторое подслово, может заменяться на другое слово. Поэтому естественное обобщение приводит к понятию случайной грамматики.

Случайная грамматика определяется множеством слов (1) и множеством допустимых подстановок $\gamma_i \to \delta_i$, $i = 1, \dots, R$ для некоторых слов γ_i, δ_i малой длины. Дисциплине LIFO соответствует так называемая линейная грамматика. Теория случайных грамматик была построена в [21–27].

3. Большие сети

Слово "сети" содержит в себе слишком много разных моделей. Какие модели и какие задачи для них считать связанными с теорией очередей? Там, где возникает проблема пробок, или более общо, временные задержки и перебои с функционированием. А пробки и задержки возникают не только в очередях, сетях связи, но и на автодорогах, вычислительных сетях и даже в живом организме.

К большим сетям относятся все задачи и результаты, где $N \to \infty$. Отсюда тесная связь с задачами статистической физики, где частицы (которые можно трактовать как пакеты, требования, машины и т.д.) движутся от узла к узлу по некоторым законам. Чрезвычайно популярны так называемые процессы с запретами (exclusion processes). В них в каждом узле может быть не более одной частицы, причем частица прыгает с некоторой интенсивностью в один из соседних узлов, если он пуст. Эти процессы моделируют явления во многих прикладных областях, см. последние ссылки в обзоре [28]. Хорошо изучены они пока только на одномерной решетке Z, см. однако модель течения жидкости в трубе в [29].

Сейчас приведем несколько моделей больших сетей. Основная их цель не конкретное практическое применение, а выявление качественных эффектов, которые могут возникнуть в более сложных системах.

Имеется цепочка обслуживающих приборов, занумерованных числами от $-N$ до N. В дискретные моменты времени на каждый прибор независимо поступает требование с вероятностью p. Каждая из этих $2N+1$ очередей обслуживается в дискретные моменты с вероятностью α, если оба соседних прибора заняты, и с вероятностью α' в остальных случаях. Если $\alpha = \alpha'$, то приборы работают независимо и условие эргодичности $p < \alpha$. Для случая $\alpha \neq \alpha'$ в [30] получены условия эргодичности, равномерные по N.

Есть два естественных способа передачи сообщений в сети. В первом (circuit-switching или коммутация каналов) для всего сообщения выделяется на время определенный канал, т.е. путь на графе. Во втором (packet-switching или коммутация пакетов [31]) сообщение разбивается на куски (пакеты), которые далее идут к получателю независимо по разным путям. Второй способ сейчас является основным, но математически менее изучен.

Простейшая сеть типа circuit-switching определяется так. На каждую вершину v графа G с некоторой интенсивностью приходят потоки требований $(v, 1), \ldots$ $\ldots, (v, n), \ldots$, каждому из которых приписан маршрут, т.е. путь $\Gamma_n(v, v_n)$ из v в некоторую другую вершину v_n, а также случайное время его обслуживания $S(\Gamma_n(v, v_n))$. В [32] получены примеры условий эргодичности для таких систем, равномерные по размеру графа G.

В [33] рассматривается сеть с потерями и circuit-switching дисциплиной на графе G. Каждое ребро l графа имеет некоторую пропускную способность $h(l)$. Каждое приходящее требование на маршрут Γ, состоящий из последовательности ребер $l_1, l_2, \ldots, l_k$, забирает некоторую часть пропускной способности на каждом из этих ребер. Новое требование теряется, если на каком-то из его ребер пропускной способности не хватает. Получены формулы для вероятностей потери требования в условиях малой нагрузки.

В [34] рассматриваются сети с потерями другого типа. Именно, требование прибывает в вершину v некоторого графа и обслуживается, если вершина свободна с интенсивностью μ, а если эта вершина занята, то требование по некоторому алгоритму (маршрутизации) переправляется в другую вершину v', где обслуживается с меньшей интенсивностью $\mu(v, v') < \mu$ либо теряется, если v' также занята. Много оптимизационных задач по сетям с потерями рассматривается в [35].

Для больших сетей есть много общих глобальных проблем. К таковым относится например известная проблема о пуассоновости потоков в больших сетях. Этой задаче посвящена серия работ Рыбко, Шлосмана, Владимирова [36–39] и др. (см. ссылки в этих работах). Основные модели в этих работах основаны на известном в статистической физики приближении среднего поля.

Локальные сети и вычислительные системы. При параллельных вычислениях большую роль играет синхронизация работы разных процессоров. В [40] рассматривается такая модель, возникшая для обоснования некоторых экспериментов в AT&T Bell Labs. Обзор других вероятностных моделей см. в [41].

Автодороги и другие транспортные сети. Представим сеть улиц в виде графа, где вершины – перекрестки, ребра – улицы. Пусть будет N перекрестков и M машин. Заданы средние времена ожидания μ_i^{-1} на перекрестке i и матрица "вероятностей" $P = (p_{ij})$ того, куда едет машина от перекрестка i – прямо, налево, направо. После ожидания они немедленно попадают на другой перекресток. В [42] рассматривается асимптотика $M, N \to \infty$ причем предел $\lim \dfrac{M}{N} = \lambda$ существует. Дан метод нахождения λ_{cr} такого, что при $\lambda > \lambda_{cr}$ хотя бы на одном перекрестке будет пробка.

Надо решить уравнение маршрутизации

$$\pi P = \pi$$

и ввести относительные интенсивности использования перекрестков и их эмпирическую меру

$$r_i = C_N \frac{\pi_i}{\mu_i}, \quad C_N = \max \frac{\mu_i}{\pi_i},$$

$$I_N(A) = \frac{1}{N} \#\{i : r_i \in A \subset [0,1]\}.$$

Тогда

$$\lambda_{cr} = \lim_{z \to 1-} h(z), \quad h(z) = \int\limits_0^1 \frac{r}{1 - rz} dI(r).$$

Пока не решенная задача – учет времени движения машин от перекрестка к перекрестку. Оптимизировать движение без перестройки дорог можно влияя на матрицу маршрутизации P так, чтобы сделать λ_{cr} максимальным.

Химические и биологические сети. Математических моделей передачи локальной и глобальной информации в живом организме известно мало. Основой функционирования организма являются химические и электрические сети. Химическая сеть состоит из многих подсетей химических реакций, сконцентрированных в компартментах, например клетках или ее частях. Компартменты связаны между собой. Стандартные математические модели функционирования химической сети в одном компартменте – классическая химическая кинетика, химическая термодинамика, микромоделью которых является стохастическая химическая кинетика. Это модели типа среднего поля, см. [7, 43, 44] и ссылки в этих работах.

Для изучения взаимодействий между разными компартментами вводится транспорт между ними. Одна из задач теории сетей – как ведут себя внутренние потоки в сети, если входы сильно флуктуируют. Этот вопрос был разобран в [45] на примере сетей химических реакций, где доказано, что флуктуации в данном узле убывают в зависимости от расстояния узла от множества входных узлов.

Нейронные сети. Теория нейронных сетей чаще всего рассматривается как модель для обучения и дальнейшего распознавания. Литература по ним огромна, как математическая, физическая так и вычислительная, см. [46–48]. Однако конкретные электрические сети в организме изучены существенно меньше, даже на физическом уровне.

Рост сетей. Сети могут расти. Это относится как к абстрактным графам [49], сетям нейронов [50,51], так и к сетям обслуживания [52], например к Интернету. При этом рост возможен в разных направлениях – математически это означает разнообразие финальных событий в соответствующей транзиентной цепи Маркова. Другая модель роста – рост пробки перед светофором – имеется в [53].

СПИСОК ЛИТЕРАТУРЫ

1. *Erlang A.K.* The theory of probabilities and telephone conversations // Nyt Tidsskrift Matematik. 1909. V. 20. P. 33–39.
2. *Ganesh A., O'Connell N., Wischik D.* Big Queues (Lect. Notes in Math. V. 1838). Berlin: Springer, 2004 (reprinted 2008).
3. *Yashkov S.F., Yashkova A.S.* Processor sharing: a survey of the mathematical theory // Autom. Remote Control. 2007. V. 68. № 9. P. 1662–1731.
4. *Yashkov S.F.* Processor-sharing queues: Some progress in analysis // Queueing Syst. 1987. V. 2. № 1. P. 1–17.

5. *Kelly F.* Reversibility and Stochastic Networks. Chichester: Wiley, 1979 (reprinted 1987, 1994).
6. *Nelson R.* The mathematics of product form queueing networks // ACM Comput. Surveys. 1995. V. 25. № 3. P. 339–369.
7. *Малышев В.А., Пирогов С.А.* Обратимость и необратимость в стохастической химической кинетике // Успехи мат. наук. 2008. Т. 63. № 1. С. 3–36.
8. *Малышев В.А.* Уравнения Винера–Хопфа и их применения в теории вероятностей // Итоги науки и техники. Сер. Теория вероятностей. Мат. статистика. Теорет. кибернетика. М.: ВИНИТИ, 1976. Т. 13. С. 5–35.
9. *Малышев В.А.* Случайные блуждания. Уравнения Винера-Хопфа в четверти-плоскости. Автоморфизмы Галуа. М.: МГУ, 1970.
10. *Fayolle G., Malyshev V., Iasnogorodski R.* Random Walks in the Quarter Plane (Algebraic methods, boundary value problems and applications to queueing systems). Berlin: Springer, 1999 (reprinted 2008).
11. *Kurkova I.A., Malyshev V.A.* Martin boundary and elliptic curves // Markov Proc. Relat. Fields. 1998. V. 4. № 2. P. 203–272.
12. *Kurkova I.A., Suhov Yu.M.* Malyshev's theory and JS-queues. Asymptotics of stationary probabilities // Ann. Appl. Probab. 2003. V. 13. № 4. P. 1313–1354.
13. *Malyshev V., Spieksma F.* Intrinsic convergence rates of countable Markov chains // Markov Proc. Relat. Fields. 1995. V. 1. № 2. P. 203–266.
14. *Fayolle G., Malyshev V., Menshikov M.* Topics in Constructive Theory of Countable Markov Chains. Cambridge: Cambridge University Press, 1995.
15. *Малышев В.А., Меньшиков М.В.* Эргодичность, непрерывность и аналитичность счетных цепей Маркова // Тр. Моск. мат. общ-ва. 1979. Т. 39. С. 3–48.
16. *Fayolle G., Malyshev V., Menshikov M., Sidorenko A.* Lyapounov functions for Jackson networks // Math. Oper. Res. 1993. V. 18. № 4. P. 916–927.
17. *Malyshev V.* Networks and dynamical systems // Adv. Appl. Prob. 1993. V. 25. P. 140–175.
18. *Malyshev V.A., Ignatyuk I.A.* Classification of random walks in Z_+^4 // Selecta Math. Sov. 1993. V. 12. № 2. P. 129–194.
19. *Игнатюк И.А., Малышев В.А., Щербаков В.В.* Влияние границ в задачах о больших уклонениях // Успехи мат. наук. 1994. Т. 49. № 2. С. 43–102.
20. *Gajrat A., Hordijk A., Malyshev V., Spieksma F.* Fluid approximation of decision Markov processes // Markov Proc. Relat. Fields. 1997. V. 3. № 1. P. 129–150.
21. *Малышев В.А.* Законы стабилизации при эволюции случайной струны // Пробл. передачи информ. 1994. Т. 30. № 3. С. 87–103.
22. *Gairat A., Iasnogorodski R., Malyshev V.* Null recurrent random string // Markov Proc. Relat. Fields. 1996. V. 2. № 3. P. 427–460.
23. *Malyshev V., Gairat A., Zamyatin A.* Two-sided evolution of a random string // Markov Proc. Relat. Fields, 1995. V. 1. № 2. P. 281–316.
24. *Малышев В., Гайрат А., Меньшиков М., Пелих К.* Классификация марковских цепей, описывающих эволюцию случайных струн // Успехи мат. наук. 1995. Т. 50. № 2. С. 5–24.
25. *Малышев В.А.* Взаимодействующие цепочки символов // Успехи мат. наук. 1997. Т. 52. № 2. С. 59–86.
26. *Малышев В.А.* Случайные грамматики // Успехи мат. наук. 1998. Т. 53. № 2. С. 107–134.
27. *Karpelevich F., Malyshev V., Petrov A., Pirogov S., Rybko A.* Context Free Evolution of Words / Analytic Methods in Applied Probability. In Memory of Fridrih Karpelevich. Ed. Yu. Suhov. Providence: Amer. Math. Soc. Transl., Ser. 2. 2002. V. 207. P. 91–115.
28. *Blythe R.A., Evans M.R.* Nonequilibrium steady states of matrix–product form: a solver's guide // J. Phys. A: Math. Theor. 2007. V. 40. P. R333–R441 (arXiv: 0706.1678. 2007. P. 1–127).
29. *Малышев В., Манита А.* Стохастическая микромодель течения Куэтта // Теория вероятностей и ее применения. 2008. Т. 53. № 4. С. 798–809.

30. *Малышев В., Цареградский И.* Система массового обслуживания с локальным взаимодействием // Теория вероятностей и ее применения. 1982. Т. 27. № 3. С. 579–583.
31. *Lee T.* The mathematical parallels between packet switching and parallel transmission. Preprint. 2006. arXiv:0610.050. P. 1–21.
32. *Berezner S.A., Malyshev V.A.* The stability of infinite server network with random routing // J. Appl. Probab. 1989. V. 26. P. 363–371.
33. *Botvich D., Fayolle G., Malyshev V.* Loss networks in thermodynamic limit / Lecture Notes in Control and Information Sciences. Berlin: Springer-Verlag. 1994. V. 199. P. 465–489.
34. *Malyshev V., Robert Ph.* Phase transitions in a load sharing model // Ann. Appl. Probab. 1995. V. 5. № 4. P. 1161–1176.
35. *Kelly F.* Loss networks // Ann. Appl. Probab. 1991. V. 1. P. 319–378.
36. *Владимиров А.А., Рыбко А.Н., Шлосман С.Б.* Свойство самоусреднения систем массового обслуживания // Пробл. передачи информ. 2006. Т. 42. № 4. С. 91–103.
37. *Rybko A., Shlosman S., Vladimirov A.* Absence of breakdown of the Poisson hypothesis I. Closed networks at low load. Preprint. 2008. arXiv:0811.3577. P. 1–18.
38. *Rybko A., Shlosman S., Vladimirov A.* Spontaneous resonances and the coherent states of the queueing networks. Preprint. 2007. arXiv:0708.3073. P. 1–53.
39. *Rybko A., Shlosman S., Vladimirov A.* Spontaneous resonances and the coherent states of the queueing networks // J. Statist. Physics. 2009. V. 134. № 1. P. 67–104 (arXiv:0708.3073).
40. *Greenberg A., Malyshev V., Popov S.* Stochastic model of massively parallel simulation // Markov Proc. Relat. Fields. 1995. V. 1. № 4. P. 473–490.
41. *Манита А.Д.* Коллективное поведение в многомерных вероятностных моделях синхронизации // Обозрение прикл. промышл. матем. 2007. Т. 14. № 6. С. 1001–1021.
42. *Malyshev V., Yakovlev A.* Condensation in large closed Jackson networks // Ann. Appl. Probab. 1996. V. 6. № 1. P. 92–115.
43. *Malyshev V., Pirogov S., Rybko A.* Random walks and chemical networks // Moscow Math. J. 2004. № 2. P. 441–453.
44. *Malyshev V.A.* Microscopic models for chemical thermodynamics // J. Statist. Phys. 2005. V. 119. № 5/6. P. 997–1026.
45. *Замятин А.А., Малышев В.А., Манита А.Д.* Явление гомеостаза в сетях химических реакций // Теория вероятностей и ее применения. 2006. Т. 51. № 4. С. 793–801.
46. *Arbib M. (Ed.)* The handbook of brain theory and neural networks. MIT Press, 2003. Second Edition.
47. *Хайкин С.* Нейронные сети: полный курс. М.: Вильямс, 2006. (пер. на рус. яз. со второго американского издания 1999 г.)
48. *Dotsenko V.* Introduction to the theory of spin glasses and neural networks. Singapore: World Scientific, 1994.
49. *Малышев В.А.* Случайные графы и грамматики на графах // Дискрет. мат. и ее применения. 1998. Т. 10. № 2. С. 30–44.
50. *Karpelevich F., Malyshev V., Rybko A.* Stochastic evolution of neural networks // Markov Proc. Relat. Fields. 1995. V. 1. № 1. P. 141–161.
51. *Malyshev V., Turova T.* Gibbs measures on attractors in biological neural networks // Markov Proc. Relat. Fields. 1997. V. 3. № 4. P. 443–464.
52. *Shcherbakov V., Volkov S.* Queueing with neighbours. Preprint. 2009. arXiv:0907.1826. P. 1–16.
53. *Замятин А.А., Малышев В.А.* Накопление на границе для одномерной стохастической системы частиц // Пробл. передачи информ. 2007. Т. 43. № 4. С. 68–82.

Статья представлена к публикации членом редколлегии С.Ф. Яшковым.

Поступила в редакцию 12.05.2009