

© 2026 г. О.В. СЕМЁНОВА, д-р физ.-мат. наук (olga.semenova@ipu.ru),
В.М. ВИШНЕВСКИЙ, д-р техн. наук (vishn@inbox.ru)
(Институт проблем управления им. В.А. Трапезникова РАН, Москва)

СТОХАСТИЧЕСКИЕ СИСТЕМЫ ПОЛЛИНГА: РАЗВИТИЕ И НОВЫЕ ПРИЛОЖЕНИЯ

Приводится обзор новых результатов, полученных в области исследования стохастических систем поллинга. Систематизируются основные направления практического применения моделей поллинга для оценки производительности и проектирования беспроводных WMAN-сетей с централизованным механизмом управления, сотовых сетей 5G/6G и сетей Интернета вещей, транспортных и медицинских систем и т.д. Обсуждаются перспективы дальнейшего развития прикладных исследований этого важного направления теории очередей.

Ключевые слова: системы поллинга, телекоммуникационные сети, 5G/6G, Интернет вещей, транспортные потоки, здравоохранение.

DOI: 10.7868/S2413977726010013

1. Введение

Математические модели стохастического поллинга нашли широкое применение при проектировании систем связи, производственных и транспортных систем, медицинских систем и т.д. Особенно эффективно эти модели применяются для оценки производительности и проектирования существующих и перспективных протоколов телекоммуникационных сетей с централизованным механизмом управления. Работы по исследованию систем стохастического поллинга и обобщение результатов, полученных до 1995 г. в этой области теории очередей, нашли отражение в монографиях [1, 2]. Благодаря широкому практическому применению интерес к исследованию систем поллинга не ослабевает до настоящего времени. В последние два десятилетия появился ряд обзорных работ по системам поллинга [3–8], посвященных в основном прогрессу в разработке математических методов исследования систем поллинга. Исключением является обзор [9], опубликованный в 2011 г., в котором систематизированы направления практического применения моделей систем стохастического поллинга.

Интенсивное развитие и быстрая смена поколений систем сотовой связи, широкополосных беспроводных сетей, систем промышленной робототехники, различных наземных и воздушных беспилотных систем и т.д. потребовало разработки новых моделей поллинга и методов их исследования для оценки

характеристик производительности и проектирования этих систем. К числу таких методов, в частности, относятся методы исследования систем поллинга с коррелированными входными потоками, характерными для современных вычислительных систем и сетей [10, 11], методы машинного обучения, применяемые, когда получение аналитических и численных результатов либо затруднительно, либо вообще невозможно традиционными методами теории очередей [12, 13]. Однако до настоящего времени отсутствуют обзорные работы по систематизации новых направлений практического применения моделей стохастического поллинга, хотя за последние годы появилось множество статей в этой области. Предлагаемый обзор закрывает этот пробел.

Статья организована следующим образом. Во втором разделе приведены основные понятия и описание базовой модели стохастического поллинга, являющейся предметом изучения большинства теоретических и прикладных работ в этой области. В третьем разделе рассматривается применение модели стохастического поллинга для оценки производительности широкополосных беспроводных городских сетей стандарта IEEE 802.11 с централизованным механизмом управления. Четвертый раздел посвящен систематизации работ по применению моделей централизованного опроса в сетях Интернета вещей и сотовых сетях 5G/6G. В пятом разделе рассматривается применение моделей стохастического поллинга на транспорте и в управлении трафиком автомобилей на перекрестках. Шестой раздел описывает модели, используемые в системах здравоохранения при прогнозировании и диагностике заболеваний, а также организации работы медицинского учреждения. В седьмом разделе описаны другие приложения систем стохастического поллинга, в частности в пассивных оптических сетях, вопросах энергосбережения, производственных процессах и т.д. В Заключении обсуждаются направления дальнейших прикладных исследований в области систем стохастического поллинга.

2. Основные понятия и описание базовой модели стохастического поллинга

В настоящем разделе приведено описание основной математической модели, которая является объектом исследования большинства прикладных работ по системам поллинга. В следующих разделах при рассмотрении результатов публикаций, представляющих теоретический и/или прикладной интерес, дается краткое описание математических моделей, если они отличаются от основной модели стохастического поллинга.

Система поллинга состоит из N ($N \geq 2$) очередей (узлов, станций) с неограниченным числом мест для ожидания и общим обслуживающим прибором (сервером). Сервером является физический объект, осуществляющий обработку поступающих запросов, а под обслуживанием подразумевается предоставление некоторого ресурса или услуги (например, услуги доступа к каналу передачи данных) в зависимости от специфики исследуемой модели.

В очередь с номером i (обозначим ее через Q_i) поступает пуассоновский поток заявок с параметром λ_i . Времена обслуживания заявки в очереди Q_i

независимы и одинаково распределены с функцией распределения $B_i(t)$, первым и вторым моментами b_i и $b_i^{(2)}$ соответственно и преобразованием Лапласа–Стилтьеса (ПЛС) $\tilde{B}_i(x) = \int_0^\infty e^{-xt} dB_i(t)$, $i = \overline{1, N}$. Сервер посещает либо предоставляет доступ к требуемому ресурсу в соответствии с определенным порядком опроса (поллинга). Наиболее часто исследуются модели с циклическим порядком опроса. В ходе обслуживания очереди заявки получают доступ к серверу согласно заданной дисциплине обслуживания. Наиболее распространенными дисциплинами являются *исчерпывающая*, при которой сервер обслуживает очередь до тех пор, пока в ней не останется заявок, *шлюзовая*, при которой обслуживание получают лишь заявки, находящиеся в очереди в момент ее опроса (начала обслуживания очереди), *глобально-шлюзовая*, при которой сервер обслуживает лишь заявки, находящиеся в системе в момент начала цикла, и *k_i -ограниченная* дисциплина, при которой сервер может обслужить не более k_i заявок за одно посещение очереди Q_i . Более подробная информация о классификации систем поллинга, видах порядка опроса, дисциплин обслуживания и других параметрах, характеризующих системы поллинга, приведена в обзорах [4, 6].

Время, которое затрачивает сервер на подключение к очереди Q_i , является случайной величиной с функцией распределения $S_i(t)$ с первым s_i и вторым $s_i^{(2)}$ моментами, а также с ПЛС $\tilde{S}_i(x)$, $i, j = \overline{1, N}$. Для систем с циклическим или периодическим опросом обозначим также через s и $s^{(2)}$ соответственно первый и второй моменты общего времени переключения сервера в цикле $s = \sum_{j=1}^N s_j$, $s^{(2)} = s^2 + \sum_{j=1}^N (s_j^{(2)} - s_j^2)$. Под циклом понимается время, которое затрачивает сервер на опрос и обслуживание всех очередей согласно порядку опроса. Для систем с циклическим опросом – это время на обслуживание очередей Q_1, \dots, Q_N . В этом случае среднее время цикла вычисляется как $C = \frac{s}{1-\rho}$, где $\rho = \sum_{i=1}^N \rho_i$ – загрузка системы, а $\rho_i = \lambda_i b_i$ – загрузка очереди Q_i .

Условие существования стационарного режима для систем с исчерпывающим или шлюзовым обслуживанием очередей имеет вид $\rho < 1$. В случае k_i -ограниченного обслуживания условие $\rho < 1$ дополняется неравенством $\lambda_i < \frac{k_i(1-\rho_i)}{s}$ для всех $i = \overline{1, N}$.

Сложность исследования описанной выше модели стохастического поллинга состоит в построении многомерного случайного процесса, адекватно описывающего алгоритм централизованного опроса, отыскании стационарных вероятностей состояний и средних характеристик производительности, а также численных решений в условиях большой размерности.

Для анализа сложных систем поллинга был разработан ряд методов: метод анализа средних [14], метод ветвящихся процессов [15], метод на основе теории разложимых полурегенерирующих процессов [16].

Наиболее часто используемым методом анализа систем поллинга с циклическим и периодическим опросом является метод производящих функций (см., например, [17]). Он позволяет получить распределение стационарных ве-

роятностей вложенного марковского процесса, описывающего состояние системы в моменты опроса сервером очередей. Приведем далее его краткое описание для случая исчерпывающей и шлюзовой дисциплин обслуживания очередей. Метод производящих функций, хорошо известный для анализа систем массового обслуживания с одной очередью, модифицирован для анализа систем с несколькими очередями (систем поллинга) и подробно описан в [17]. Данный метод применяется для циклических систем с исчерпывающим, шлюзовым или глобально-шлюзовым обслуживанием, а также для некоторых систем с нециклическим опросом. В рамках данного обзора представлены дополнения и модификация метода производящих функций в некоторых частных случаях систем поллинга, представляющих практический интерес.

Обозначим через X_i^j число заявок в очереди Q_j в момент опроса очереди Q_i , $i, j = \overline{1, N}$. Пусть $G_i(\mathbf{z}) = \mathbf{E} \left[\prod_{j=1}^N z_j^{X_i^j} \right]$, $i = \overline{1, N}$ – производящие функции стационарных вероятностей случайных величин X_i^j , $i, j = \overline{1, N}$, где $\mathbf{z} = (z_1, \dots, z_N)$.

В случае *шлюзового обслуживания* производящие функции $G_i(\mathbf{z})$, $i = \overline{1, N}$ удовлетворяют функциональным соотношениям

$$(1) \quad G_{i+1}(\mathbf{z}) = G_i \left(z_1, z_2, \dots, z_{i-1}, \tilde{B}_i \left[\sum_{j=1}^N \lambda_j (1 - z_j) \right], z_{i+1}, \dots, z_N \right) \tilde{S}_{i+1} \left[\sum_{j=1}^N \lambda_j (1 - z_j) \right], \\ i = \overline{1, N}.$$

Среднее число заявок $f_i(j) = \mathbf{E}[X_i^j]$ в очереди Q_j в момент опроса очереди Q_i определяется частной производной $f_i(j) = \mathbf{E}(X_i^j) = \left. \frac{\partial G_i(\mathbf{z})}{\partial z_j} \right|_{\mathbf{z}=(1,1,\dots,1)}$ и может быть вычислено как решение системы линейных уравнений

$$(2) \quad f_{i+1}(j) = f_i(j) + \lambda_j b_i f_i(i) + \lambda_j s_{i+1}, \quad j \neq i, \\ f_{i+1}(i) = \lambda_i b_i f_i(i) + \lambda_i s_{i+1}, \quad i = \overline{1, N}.$$

Система уравнений (2) имеет явное решение

$$f_i(j) = \begin{cases} \lambda_j \left(\sum_{k=j}^{i-1} \left[\rho_k \frac{s}{1-\rho} + s_{k+1} \right] \right), & j \neq i, \\ \lambda_i \frac{s}{1-\rho}, & j = i. \end{cases}$$

С помощью повторного дифференцирования производящих функций $G_i(\mathbf{z})$, $i = \overline{1, N}$ получают вторые моменты $f_i(j, k)$, $i, j, k = \overline{1, N}$ случайных величин X_i^j , $i, j = \overline{1, N}$ как решение системы линейных уравнений и да-

лее используют для вычисления среднего времени ожидания $W_i = \frac{(1+\rho_i)f_i(i,i)}{2\lambda_i^2 C}$ в очереди Q_i , $i = \overline{1, N}$.

В случае *исчерпывающего обслуживания* соотношения (1) имеют вид

$$G_{i+1}(\mathbf{z}) = \\ = G_i \left(z_1, \dots, z_{i-1}, \tilde{\theta}_i \left[\sum_{j=1, j \neq i}^N \lambda_j (1 - z_j) \right], z_{i+1}, \dots, z_N \right) \tilde{S}_{i+1} \left[\sum_{j=1}^N \lambda_j (1 - z_j) \right], \\ i = \overline{1, N},$$

где $\tilde{\theta}_i(s)$ – ПЛС периода занятости, порожденного одной заявкой, в системе массового обслуживания с одним сервером, соответствующей очереди Q_i . Функция $\tilde{\theta}_i(s)$ удовлетворяет функциональному уравнению

$$(3) \quad \tilde{\theta}_i(w) = \tilde{B}_i(w + \lambda_i - \lambda_i \tilde{\theta}_i(w)).$$

Аналогично, средние времена ожидания вычисляются как $W_i = \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} + \frac{f_i(i,i)}{2\lambda_i^2(1-\rho_i)C}$, $i = \overline{1, N}$.

Отметим также ряд других методов, не упомянутых выше и применяемых для анализа систем поллинга: метод функциональных вычислений [18], матрично-аналитический подход [19], асимптотика характеристик систем в условиях большой загрузки [20, 21], использование закона псевдосохранения в системах поллинга [22], а также другие методы, информацию о которых можно найти, например, в монографии [5] и обзоре [6].

3. Применение моделей стохастического поллинга для оценки производительности и проектирования широкополосных беспроводных городских сетей

Беспроводная городская сеть (Wireless Metropolitan Area Network, WMAN) – это тип беспроводной сети, предназначенный для телекоммуникационного покрытия обширной географической территории. Сети WMAN состоят из ряда взаимосвязанных беспроводных базовых станций или точек доступа, стратегически расположенных по всей территории мегаполиса. Эти базовые станции взаимодействуют друг с другом и с устройствами конечных пользователей, обеспечивая эффективную и бесперебойную передачу данных.

В беспроводных городских сетях, функционирующих под управлением протоколов стандарта IEEE 802.11, широко используется механизм централизованного опроса. К числу таких протоколов, обеспечивающих реализацию циклического опроса базовой станции (сервером доступа) абонентских терминалов, относятся протокол с функцией координации PCF (Point Coordination Function) и его развитие – протокол с гибридной функцией координацией HCF (Hybrid Coordination Function) и протокол HCCA (HCF Controlled

Channel Access) с контролируемым доступом к каналам HCF [23]. Использование этих протоколов в городских сетях с высокоэтажной застройкой и наличием радиопомех позволяет избежать проблемы скрытых абонентских станций, эффективно планировать порядок доступа станций к беспроводному каналу, гибко управлять работой радиоячейки и изменять ее параметры в соответствии с текущей ситуацией.

В [24–26] проведен сравнительный анализ различных вариантов схем опроса для оценки характеристик производительности городских сетей PCF IEEE 802.11, а также предложена модель адаптивного поллинга, позволяющая не только решить проблему скрытых станций, но и оптимизировать доступ к каналу передачи данных. Отметим также, что впервые схема адаптивного опроса и первая версия протокола были описаны в [27–29].

При адаптивной схеме опроса порядок посещения очередей цикличен, но сервер не посещает (не опрашивает) очереди, которые были пусты на момент опроса в предыдущем цикле. В случае несимметричной системы и ненулевых времен переключения адаптивный опрос сокращает среднее время цикла опроса и, соответственно, повышает производительность беспроводной сети.

Формула для среднего времени цикла из раздела 2 в этом случае принимает вид

$$C = \frac{\sum_{i=1}^N s_i u_i + \beta \prod_{i=1}^N (1 - u_i)}{1 - \rho},$$

где $u_i = \frac{1}{1 + e^{-\lambda_i C}}$ – вероятность того, что в произвольном цикле очередь Q_i посещается сервером, $i = \overline{1, N}$, β – среднее время простоя (ожидания) сервера в случае, если все очереди подряд оказались пусты при опросе и должны быть пропущены. Ожидание сервера здесь вводится с целью устранения пустых циклов опроса.

В случае шлюзовой дисциплины обслуживания система (1) для производящих функций $G_i(\mathbf{z})$, $i = \overline{1, N}$ принимает вид

$$(4) \quad G_i(\mathbf{z}) = (1 - u_i) u_{i-1} \mathcal{M}_{i+1}^{(1)}(\mathbf{z}) + \dots + (1 - u_1) \dots (1 - u_{N-1}) u_N \mathcal{M}_{i+1}^{(N-1)}(\mathbf{z}) + \\ + (1 - u_1) \dots (1 - u_N) \mathcal{M}_{i+1}^{(N)}(\mathbf{z}) + u_i \mathcal{M}_{i+1}^{(0)}(\mathbf{z}),$$

где

$$\mathcal{M}_{i+1}^{(l)}(\mathbf{z}) = G_{i-l} \left(z_1, \dots, z_{i-l}, \tilde{B}_{i-l} \left(\sum_{j=1}^N \lambda_j (1 - z_j) \right), z_{i-l+2}, \dots, z_N \right) \times \\ \times \tilde{S}_{i-l+1} \left[\sum_{j=1}^N \lambda_j (1 - z_j) \right],$$

$$\mathcal{M}_{i+1}^{(N)}(\mathbf{z}) = G_{i-N} \left(z_1, \dots, z_{i-N}, \tilde{B}_{i-N} \left(\sum_{j=1}^N \lambda_j (1 - z_j) \right), z_{i-N+2}, \dots, z_N \right) \times \\ \times \tilde{S}_{i-N+1} \left[\sum_{j=1}^N \lambda_j (1 - z_j) \right] \tilde{H} \left(\sum_{j=1}^N \lambda_j (1 - z_j) \right), \quad l = \overline{0, N-1},$$

а система (2) преобразуется к виду

$$f_{i+1}(j) = u_i [I_{\{i=j\}} f_i(j) + \lambda_j b_i f_i(i) + \lambda_j s_{i+1}] + \\ + (1 - u_i) u_{i-1} [\lambda_j (s_{i+1} + \beta) + I_{\{i-1=j\}} f_{i-1}(j) + \lambda_j b_{i-1} f_{i-1}(i-1) + \lambda_j s_{i+1}] + \dots + \\ + (1 - u_1) \dots (1 - u_N) [I_{\{i-N=j\}} f_{i-N}(j) + \lambda_j b_{i-N} f_{i-N}(i-N)], \quad i, j = \overline{1, N}.$$

Как и в случае обычного циклического опроса, средние времена ожидания для системы с адаптивным опросом определяются как $W_i = \frac{(1+\rho_i) f_i(i, i)}{2\lambda_i^2 C}$, $i = \overline{1, N}$, где вторые моменты $f_i(j, k)$, $i, j, k = \overline{1, N}$ вычисляются как решение системы линейных уравнений, полученных дифференцированием производящих функций $G_i(\mathbf{z})$, $i = \overline{1, N}$. Случай исчерпывающего обслуживания очередей рассматривается аналогичным образом [25].

В [30] исследуется модель поллинга с ограниченным по времени обслуживанием очередей, описывающая передачу данных в сети НССА IEEE 802.11e. Показано, что производящая функция $\omega_i(\mathbf{z})$ числа заявок в системе в моменты завершения обслуживания в очереди Q_i выражается через производящую функцию $\bar{\omega}_i(\mathbf{z})$ числа заявок в системе в моменты опроса очереди Q_i следующим образом:

$$\omega_i(\mathbf{z}) = \frac{\gamma \tilde{B}_i \left(\sum_{k=1}^N \lambda_k (1 - z_k) \right)}{z_i - \tilde{B}_i \left(\sum_{k=1}^N \lambda_k (1 - z_k) \right)} \left[\bar{\omega}_{i-1}(\mathbf{z}) \tilde{S}_i \left(\sum_{k=1}^N \lambda_k (1 - z_k) \right) - \bar{\omega}_i(\mathbf{z}) \right], \\ i = \overline{1, N},$$

где γ – известная константа, определяемая средним временем цикла C и общей интенсивностью поступления заявок в систему. Производящие функции $\bar{\omega}_i(\mathbf{z})$, $i = \overline{1, N}$ находятся из функциональных уравнений численно с помощью преобразования Фурье. ПЛС $W_i(s)$ времени ожидания в очереди Q_i удовлетворяет соотношению

$$W_i(s) = N_i (1 - s/\lambda_i) / \tilde{B}_i(s), \quad i = \overline{1, N},$$

где $N_i(s)$ – производящая функция числа заявок в очереди Q_i в моменты завершения обслуживания, которая вычисляется как $N_i(z) = \omega_i(1, \dots, 1, z_i, 1, \dots, 1) / \omega_i(\mathbf{1})$.

В [31] также рассматривается применение крупномасштабной системы поллинга с дисциплинами обслуживания ветвящегося типа к моделированию сетевого протокола ВАСnet (Building Automation and Control Networks)

для систем автоматизации зданий и сетей управления зданиями. Дисциплины ветвящегося типа определяются следующим образом. Если в некоторый вложенный момент времени, в частности в момент опроса произвольной очереди Q_i , ее длина равна k_i , то каждая из этих k_i заявок за время обслуживания этой очереди замещается некоторым случайным числом заявок, распределение которых задано многомерной производящей функцией $h_i(\mathbf{z})$. Как показано в [32], системы поллинга с дисциплинами ветвящегося типа позволяют проводить точный анализ характеристик производительности, а дисциплины обслуживания, которые не удовлетворяют данному свойству, предполагают лишь применение приближенных методов анализа.

К дисциплинам ветвящегося типа относятся, например, шлюзовая с

$$h_i(z_1, \dots, z_N) = \tilde{B}_i \left(\sum_{k=1}^N \lambda_k (1 - z_k) \right), \quad i = \overline{1, N}$$

и исчерпывающая с $h_i(z_1, \dots, z_N) = \tilde{\theta}_i \left(\sum_{k=1}^N \lambda_k (1 - z_k) \right)$, $i = \overline{1, N}$, где функция $\tilde{\theta}(s)$ определяется из функционального соотношения (3). Ограниченные же дисциплины не являются таковыми, а системы с такими дисциплинами требуют разработки приближенных методов.

Для системы поллинга с дисциплиной ветвящегося типа функциональное соотношение (1) для производящих функций $G_i(\mathbf{z})$, $i = \overline{1, N}$ преобразуется к виду

$$G_{i+1}(\mathbf{z}) = G_i(z_1, z_2, \dots, z_{i-1}, h_i(\mathbf{z}), z_{i+1}, \dots, z_N) \tilde{S}_{i+1} \left[\sum_{j=1}^N \lambda_j (1 - z_j) \right], \quad i = \overline{1, N},$$

который позволяет получить характеристики производительности на основе средних длин очередей в моменты опроса, полученных путем дифференцирования приведенных выше соотношений.

ПЛС времени цикла C определяется как

$$\mathbf{E} [e^{-uC}] = G_1(\tilde{\theta}_1(\gamma_1(\mathbf{u})), \dots, \tilde{\theta}_N(\gamma_N(\mathbf{u}))) \prod_{i=1}^N \tilde{S}_i(\mathbf{u}),$$

где $\mathbf{u} = (u, \dots, u)$, $\gamma_N(\mathbf{z}) = z_N$, $\gamma_i(\mathbf{z}) = z_i + \sum_{j=i+1}^N \lambda_j (1 - \tilde{\theta}_j(\gamma_j(\mathbf{z})))$, $i = \overline{1, N-1}$.

Для дисциплин обслуживания, которые не относятся к классу ветвящихся, таких как k -ограниченное обслуживание, в [31] также предложена так называемая *гибкая k -ограниченная* дисциплина. Суть ее аналогична обычной k -ограниченной дисциплине за исключением случаев, когда сервер в очереди обслуживает меньше k заявок. В этом случае сервер использует неизрасходованный ресурс на обслуживание следующей очереди по порядку. Это

позволяет, например, в случае сильной несимметричности системы поллинга отдавать часть ресурса обслуживания от менее загруженных очередей более загруженным, тем самым снижая среднее время ожидания в системе.

В [33, 34] рассматривается возможность организации городской широкополосной беспроводной сети с использованием привязной высотной беспилотной платформы длительного функционирования [35–37]. Базовая станция PCF IEEE 802.11, установленная на борту беспилотного летательного аппарата (БПЛА) на высоте 100–150 м, позволяет значительно увеличить зону прямой видимости между абонентами и антенной базовой станции, соответственно увеличивает область телекоммуникационного покрытия сети. В [38, 39] исследуется модель поллинга с повторными вызовами, описывающая механизм передачи данных в FANET (Flying Ad Hoc Network) – беспроводных самоорганизующихся сетях на основе БПЛА.

В [40] исследуется случай марковского порядка опроса очередей, моделирующий множественный доступ с предотвращением коллизий и контролем несущей (Carrier-Sense Multiple-Access Collision-Avoidance, CSMA-CA) в сетях IEEE802.11 в режиме случайного доступа к среде. Сервер посещает очереди согласно случайной марковской схеме опроса, описываемой стохастической матрицей $\mathbf{P} = (p_{i,j})_{i,j=\overline{1,N}}$ и вектором вероятностей $\mathbf{q} = (q_1, \dots, q_N)$. Вектор \mathbf{q} вычисляется как решение системы $\mathbf{qP} = \mathbf{q}$, $\sum_{i=1}^N q_i = 1$.

Дисциплина обслуживания очередей может быть произвольной ветвящегося типа, как описано выше. Среднее время цикла для очереди Q_i вычисляется как $C_i = \frac{\sigma}{q_i(1-\rho)}$, где $\sigma = \sum_{j=1}^N q_j \sum_{k=1}^N p_{j,k} \mathbf{E}[S_{j,k}]$; $S_{j,k}$ – случайная величина, описывающая время переключения между очередями Q_j и Q_k , $j, k = \overline{1, N}$.

В данном случае возможно применение метода производящих функций, который приводит к функциональным соотношениям, аналогичным (1):

$$q_j G_j(\mathbf{z}) = \sum_{i=1}^N p_{i,j} q_i G_i(z_1, z_2, \dots, z_{i-1}, h_i(\mathbf{z}), z_{i+1}, \dots, z_N) \tilde{S}_{i+1} \left[\sum_{j=1}^N \lambda_j (1 - z_j) \right],$$

$$j = \overline{1, N}.$$

Первые моменты $f_i(j)$, $i, j = \overline{1, N}$ длин очередей в начале опроса могут быть получены в явном виде, что позволяет вычислить ключевые характеристики производительности системы.

Средняя безусловная длина L_j очереди Q_j вычисляется как

$$L_j = \sum_{i=1}^N \left(\rho_i x_i(j) + \frac{(1-\rho)q_i}{\sigma} \sum_{k=1}^N p_{i,k} \mathbf{E}[S_{i,k}] y_{i,k}(j) \right), \quad j = \overline{1, N},$$

где $x_i(j)$ и $y_{i,k}(j)$ – известные константы. Статья [40] также обобщает закон псевдосохранения для произвольной дисциплины обслуживания ветвящегося типа, который может быть использован для оптимизации производительности

сти системы,

$$\begin{aligned} \sum_{i=1}^N \rho_i W_i &= \rho \frac{\sum_{i=1}^N \lambda_i b_i^{(2)}}{2(1-\rho)} + \frac{1}{\sigma} \sum_{i=1}^N q_i \sum_{k=1}^N p_{i,k} \mathbf{E}[S_{i,k}] \sum_{j=1, j \neq i}^N f_i(j) b_j + \\ &+ \frac{1}{1-\rho} \sum_{i=1}^N \lambda_i \frac{1-\rho_i}{1-h_i(i)} \sum_{k=1}^N p_{i,k} \mathbf{E}[S_{i,k}] \sum_{j=1}^N b_j h_i(j) + \frac{\rho}{2\sigma} \sum_{i=1}^N q_i \sum_{k=1}^N p_{i,k} \mathbf{E}[S_{i,k}^2]. \end{aligned}$$

Модель системы приоритетного опроса с тремя очередями и пороговым обслуживанием используется в [41] для анализа систем коммутации в режиме ATM (Asynchronous Transfer Mode) сетевых стандартов типа DQDB (Distributed Queue Dual Bus) для масштабных городских сетей.

Аналогичное приложение системы поллинга со случайным порядком опроса к описанию схемы множественного доступа в городских сетях WiMAX IEEE802.16 в режиме конкурентного доступа к среде рассмотрено в [42]. В [43] предложен многоуровневый протокол MAC с приоритетным опросом в применении для военных целей. Для системы с тремя очередями и ограниченным обслуживанием, описывающей частный случай системы поллинга с адаптивным механизмом опроса, в [44] представлено условие существования стационарного режима.

В заключение раздела отметим, что в дальнейших версиях стандарта IEEE802.11 были добавлены новые механизмы доступа (например, Triggered Uplink Access, TUA) [45, 46], которые также могут быть эффективно исследованы с использованием моделей стохастического поллинга.

4. Систематизация работ по применению моделей поллинга в сетях Интернета вещей и сотовых сетях 5G/6G

Сети Интернета вещей (IoT, Internet of Things) интенсивно развиваются в последние годы (см. [47–49]). Для оценки характеристик и сравнительного анализа протоколов таких сетей могут эффективно использоваться модели стохастического поллинга. Среди работ, посвященных такому анализу, отметим следующие.

В [50] анализируется MAC-протокол управления доступом к среде в Интернете вещей с использованием модели поллинга с приоритетным опросом типа «звезда». Такой опрос позволяет разделять пользователей с разными приоритетами для снижения задержки и обеспечения равнодоступного обслуживания. Модель представлена системой из $N + 1$ очереди типа $M/G/1$. Очередь Q_h является приоритетной по отношению к остальным N очередям и обслуживается сервером согласно шлюзовой дисциплине каждый раз по завершению обслуживания любой другой очереди. Неприоритетные очереди получают исчерпывающее обслуживание.

Вектор, описывающий число заявок в очередях в момент опроса Q_i , имеет размер $N + 1$, поэтому соответствующая производящая функция стационар-

ных вероятностей имеет дополнительный аргумент z_h . Функциональные соотношения (1) в данном случае связывают производящие функции $G_i(\mathbf{z}, z_h)$ и $G_{ih}(\mathbf{z}, z_h)$ в виде

$$(5) \quad G_{i+1}(\mathbf{z}, z_h) = G_{ih} \left(\mathbf{z}, \tilde{B}_h \left(\sum_{j=1}^N \lambda_j(1 - z_j) + \lambda_h(1 - z_h) \right) \right),$$

$$G_{ih}(\mathbf{z}, z_h) = G_i \left(z_1, \dots, z_{i-1}, \tilde{\theta}_i \left(\sum_{j=1}^N \lambda_j(1 - z_j) + \lambda_h(1 - z_h) \right), z_{i+1}, \dots, z_N, z_h \right) \times$$

$$\times \tilde{S}_i \left[\sum_{j=1}^N \lambda_j(1 - z_j) + \lambda_h(1 - z_h) \right], \quad i = \overline{1, N}.$$

Как показано в разделах 2 и 3, данные функциональные уравнения позволяют вычислять основные характеристики производительности. Отметим, что здесь $\tilde{S}_i(s)$ – это ПЛС времени отключения сервера от очереди Q_i .

В [51] используется алгоритм нейронной сети для прогнозирования и анализа производительности сети в беспроводных сенсорных сетях (WSN). Модель позволяет решать приоритетные задачи в сфере Интернета вещей. Для прогнозирования производительности системы построена нейронная сеть. Показано, как предлагаемая схема приоритетов помогает повысить производительность системы по сравнению с аналогичной системой поллинга с одноуровневым приоритетом.

В [52] изложен новый алгоритм быстрого опроса (поллинга) для беспроводных mesh-сетей в узкополосных системах Интернета вещей. Показано, что алгоритм значительно минимизирует время опроса, требуемое центру управления и сбора данных для получения отклика от всех удаленных телеметрических устройств. В [53] исследуется модель поллинга с двумя очередями и коррелированным входным потоком *ВМАР* (групповой марковский входной поток) для проектирования и оптимизации потоковой передачи видео высокой четкости в сетях Интернета вещей.

Модели стохастического поллинга могут эффективно использоваться для оценки производительности и проектирования протоколов сотовых сетей нового поколения 5G/6G. Развертывание сетей 5G/6G обеспечивает значительное расширение спектра услуг пользователям, повышение емкости сети и скорости передачи данных по сравнению с сетями предыдущих поколений [54]. Для достижения высоких и сверхвысоких скоростей передачи данных на абонентских и транспортных участках в сетях нового поколения используются не только традиционные каналы сантиметрового диапазона радиоволн (до 6ГГц), но и каналы миллиметрового диапазона (до 100ГГц), обладающие широкой полосой пропускания [55, 56].

Однако с учетом особенностей распространения волн миллиметрового диапазона (короткие зоны покрытия) консорциумом 3GPP была предложена тех-

нология интегрированного доступа и транспорта IAB (Integrated Access and Backhaul) [57], которая интегрирует функции доступа и транспорта в одной сети. Технология позволяет организовать экономически эффективную сеть за счет использования недорогих ретрансляционных узлов (узлов IAB) вместо полностью оборудованных базовых станций, что особенно важно при развертывании сетей 5G вдоль транспортных магистралей.

В [58] для анализа производительности конечного узла в сети IAB используется модель поллинга с двумя очередями. Одна очередь хранит данные, переданные по нисходящему каналу от родительского узла сети IAB, вторая – данные, переданные по восходящему каналу от дочерних узлов и пользовательского оборудования, связанного с конечным узлом. В [59–61] граничный узел моделируется системой поллинга типа $M_K/G_K/1$ с циклическим опросом с мгновенными переключениями сервера между очередями ($s_i = 0, i = \overline{1, K}$) за исключением периода времени s_0 в начале цикла, которое можно интерпретировать как время подключения к очереди Q_1 . Предполагается, что заявки могут поступать в очереди лишь в течение времени s_0 , что представляет собой комбинацию глобально-шлюзового и исчерпывающего обслуживания. В этом случае производящая функция (1) числа заявок в моменты опроса очередей имеет вид

$$G_i(\mathbf{z}) = \tilde{S} \left(\sum_{j=i}^K (\lambda_j (1 - z_j)) \right), \quad i = \overline{1, K},$$

где $\tilde{S}(w)$ – ПЛС функции распределения случайной величины s_0 . Далее в явном виде получены моменты произвольного порядка для случайных величин $X_i^j, i, j = \overline{1, K}$, описывающих число заявок в очередях в моменты опроса. ПЛС распределения времени задержки в очереди Q_i с произвольной функцией распределения времени обслуживания имеет вид

$$\tilde{\Delta}_i(w) = s^i \prod_{j=1}^i \frac{1}{s + \lambda_j (1 - \tilde{B}_j(w))}$$

со средним $\Delta_i = \frac{1}{s} \sum_{j=1}^i \lambda_j b_j, i = \overline{1, K}$.

Модели поллинга нашли свое применение при моделировании контроля перегрузки SIP-серверов [62]. SIP (Session Initiation Protocol) – протокол передачи данных, описывающий способ установления и завершения пользовательского сеанса связи в сетях 5G. В [62] предложена схема порогового контроля нагрузки, генерируемой неприоритетными клиентами. В [63] продолжен анализ механизма контроля перегрузки SIP-серверов путем тестирования различных дисциплин обслуживания. Рассмотренная модель поллинга предполагает пороговую дисциплину переключения к приоритетной очереди и две дисциплины обслуживания: исчерпывающую и шлюзовую. Показано, что в некоторых случаях предпочтительнее использовать шлюзовое обслуживание приоритетной очереди.

В заключение данного раздела отметим также статью [64], в которой рассматривается модель с управляемым (динамическим) порядком опроса для достижения требуемого уровня равнодоступности – способности системы обеспечить одинаковое качество обслуживания абонентов в обобщенных алгоритмах планирования в современных беспроводных сетях, и работу [65], в которой исследуются различные политики поллинга (правила опроса и обслуживания) для решения задач масштабируемости стандартного протокола управления передачей данных TCP (Transmission Control Protocol).

5. Применение моделей стохастического поллинга на транспорте и в управлении трафиком автомобилей на перекрестках

Применение моделей поллинга в управлении дорожным движением имеет обширную предысторию и подробно освещено в обзоре [9], опубликованном в 2011 г. Дальнейшее развитие этого направления применения моделей поллинга нашло отражение в целом ряде более поздних статей. В частности, в [66] рассмотрена модель поллинга со случайной длительностью разрешающих сигналов светофора для каждого направления движения в зависимости от его загруженности с целью минимизации времени ожидания проезда на перекрестке. Для моделирования автоматизированного управления светофором, регулирующим движение транспортных средств на перекрестке, в [67, 68] исследуется система поллинга двух очередей с приоритетом порогового типа для переключения сервера между очередями. Светофор определяет приоритет проезда между двумя направлениями пороговым образом: когда количество автомобилей в каком-либо направлении достигает порогового значения, для него загорается зеленый свет. При этом в [68] для получения основных характеристик производительности такой модели применен матрично-аналитический подход. Разновидность пороговых стратегий переключения сервера между очередями, регулирующая движение на перекрестке, также рассматривается в [7].

В [69] предложено точное решение для системы поллинга с двумя очередями и пороговым приоритетом. Результаты моделирования механизма поллинга на полностью автономных перекрестках представлены в [70]. Беспилотные автомобили, подъезжающие к перекрестку, корректируют свою скорость, чтобы проехать перекресток, избегая столкновений.

Для транспортных потоков предположение о независимости потоков, осуществляющих движение на перекрестке, на самом деле, не является реалистичным, поэтому учет коррелированного характера реальных потоков транспорта является весьма актуальной задачей. В [71, 72] для модели поллинга с двумя очередями коррелированный и нестационарный характер входных потоков учитывается с помощью входного марковского (МАР) и маркированного марковского (ММАР) входных потоков. В [71] рассматриваются две очереди типа $МАР/PH/1/N$ с поочередным обслуживанием сервером и фазовым распределением времени обслуживания. Авторами приводится по-

дробный анализ времени ожидания в очереди, который может служить руководством для анализа времени ожидания в других моделях поллинга не только через их средние значения, но и в виде распределений. Например, ПЛС распределения времени ожидания заявок в первой очереди имеет вид

$$\begin{aligned}
 v_1(s) = & \frac{1}{\lambda_1} \sum_{i_1=0}^{N_1-1} \sum_{i_2=0}^{N_2} \left(\pi^{(-1)}(i_1, i_2) \left(I_{M_{-1}} \otimes \hat{D}_1^{(1)} \right) \mathbf{e} v_1^{(-1)}(s, i_1) + \right. \\
 & + \pi^{(-2)}(i_1, i_2) \left(I_{M_{-2}} \otimes \hat{D}_1^{(1)} \right) \mathbf{e} v_1^{(-2)}(s, i_1, i_2) + \\
 & + \sum_{j=1}^{N_1} \pi^{(1)}(j, i_1, i_2) \left(I_{M_1} \otimes \hat{D}_1^{(1)} \right) \mathbf{e} v_1^{(1)}(s, j, i_1, i_2) + \\
 & \left. + \sum_{j=1}^{N_2} \pi^{(2)}(j, i_1, i_2) \left(I_{M_2} \otimes \hat{D}_1^{(1)} \right) \mathbf{e} v_1^{(2)}(s, j, i_1, i_2) \right) + P_1^{(loss)},
 \end{aligned}$$

где $P_1^{(loss)}$ – вероятность потери заявки в очереди Q_1 , $\pi^{(k)}(*)$ – стационарные вероятности, соответствующие состоянию $(*)$ при посещении сервером очереди Q_k , $v_1^{(k)}(s, *)$ – ПЛС распределения условного времени ожидания в очереди Q_1 при условии, что сервер находится в состоянии k , а система – в состоянии $*$, $D_1^{(1)}$ – одна из матриц, описывающих поступления заявок МАР-потока.

В заключение этого раздела отметим и другие применения моделей поллинга в транспортных системах. В [73] исследуется применение технологии радиочастотной идентификации (RFID, Radio Frequency IDentification) [74] для контроля числа транспортных средств на перекрестке с целью адаптации сигналов светофора к текущей дорожной ситуации. Для транспортных систем в [75] анализируется проблема улучшения приоритетной доставки заказов без снижения эффективности доставки обычных заказов. С этой целью предложено назначать некоторые комплекточные машины приоритетными с исчерпывающей дисциплиной обслуживания, остальные – неприоритетными с лимитом обслуживания в один заказ (1-ограниченная дисциплина обслуживания). В [76] предложена модификация дисциплин обслуживания в системе поллинга, описывающая систему доставки заказов, с целью снижения времени сборки нескольких заказов для одного клиента.

6. Применение моделей стохастического поллинга в системах здравоохранения

В настоящее время модели систем поллинга активно применяются в области медицинского мониторинга. Например, в [77] с помощью системы поллинга оценивается производительность беспроводных сетей нательных датчиков, которые представляют собой современные системы мониторинга, использующие беспроводные технологии для прогнозирования и диагностики заболеваний. В статье [78] анализируется схема спящего режима в беспроводных сетях

передачи данных для нательных датчиков (WBAN) стандарта IEEE802.15 посредством модели поллинга с дискретным временем и адаптивным порядком опроса в предположении, что сервер пропускает пустые очереди, если они были пусты в предыдущий момент опроса.

Модель, рассмотренная в [19], описывает работу по первичному приему пациентов в медицинской клинике. Пациентов принимают несколько медсестер, каждая из которых организует свою очередь пациентов, и дежурный врач. Пациент направляется к медсестре с самой короткой очередью для обследования. После этого пациентов принимает врач, каждый раз выбирая пациента из самой длинной очереди. Данная схема предполагает комбинацию двух стратегий поступления и обслуживания заявок: присоединение поступающей заявки к очереди меньшей длины и обслуживание сервером очереди с большим числом заявок. Это позволяет описать поведение соответствующей модели поллинга с двумя очередями двумерной цепью Маркова, первая компонента которого отражает длину очереди Q_1 , а вторая – текущую разницу в длине очередей. Цепь Маркова в данном случае принадлежит классу многомерных процессов гибели и размножения и имеет инфинитезимальный генератор вида

$$Q = \begin{bmatrix} B_1 B_0 0 \dots \dots \dots \\ B_2 A_1 A_0 0 \dots \dots \dots \\ 0 A_2 A_1 A_0 0 \dots \dots \dots \\ \vdots \ddots \ddots \ddots \ddots \ddots \ddots \end{bmatrix},$$

где элементы матриц A_i и B_i – входные параметры системы.

Вектор \mathbf{p}_n стационарных вероятностей состояний системы, соответствующий числу n заявок в очереди Q_1 , определяется равенством $\mathbf{p}_n = \mathbf{p}_1 R^{n-1}$, $n \geq 1$, где матрица R есть минимальное неотрицательное решение матричного уравнения $A_0 + R A_1 + R^2 A_2 = O$. А векторы \mathbf{p}_0 и \mathbf{p}_1 находятся как решение системы уравнений

$$\mathbf{p}_0 B_1 + \mathbf{p}_1 B_2 = \mathbf{0}, \quad \mathbf{p}_0 B_0 + \mathbf{p}_1 (A_1 + R A_2) = \mathbf{0}, \quad \mathbf{p}_0 \mathbf{e} + \mathbf{p}_1 [I - R]^{-1} \mathbf{e} = 1,$$

где $\mathbf{e} = (1, \dots, 1)$, I – единичная матрица. Средняя длина очереди Q_1 вычисляется как $L_1 = \mathbf{p}_1 [I - R]^{-2} \mathbf{e}$.

В [79] предложено использовать схему поллинга для планирования периферийных вычислений с целью адаптации методов планирования для множества сервисов. Периферийный сервер предоставляет услуги информационному центру больницы. Авторы рассматривают задачу максимизации эффективности использования периферийных серверов с помощью схемы приоритетного опроса типа «звезда». Как и в [50], модель имеет N обычных очередей и одну приоритетную очередь. Все очереди получают исчерпывающее обслуживание, а очередь Q_h посещается сервером всякий раз, как только сервер покидает неприоритетную очередь. Производящие функции числа заявок в

очередях в моменты опроса удовлетворяют функциональным соотношениям, аналогичным (5), где ПЛС $\tilde{B}_h(s)$ заменяется на $\tilde{\theta}_h(s)$, определяемый формулой (3).

7. Другие приложения систем поллинга

Системы поллинга используются и в целом ряде других областей, которые не были описаны в предыдущих разделах, например в пассивных оптических сетях, робототехнике, спутниковых системах и производственных процессах.

Пассивные оптические сети, PON (Passive Optical Network), – это телекоммуникационная технология, использующая оптоволоконные каналы связи для предоставления абонентам высокоскоростного доступа в интернет, видео- и голосовых услуг. Термин «пассивный» обусловлен тем, что для распределения оптического сигнала по нескольким конечным точкам используются пассивные оптические разветвители, не требующие активных компонентов. В [80] исследуется вопрос существования стационарного режима в многосерверной системе поллинга с ограниченным обслуживанием, моделирующей процесс передачи данных в сетях на основе технологии PON.

В [81] предложена схема накопления и обслуживания заявок в системах поллинга, моделирующих передачу данных на основе технологии PON. Это так называемая многофазная шлюзовая дисциплина обслуживания. Заявка, поступающая в очередь Q_i , должна ожидать K_i циклов прежде, чем она будет обслужена, $i = \overline{1, N}$. Как отмечают авторы, такая многофазная дисциплина помогает решить проблему монополизации сервера более загруженными очередями за счет выбора подходящих значений (K_1, \dots, K_N) . Несмотря на сложность, данная дисциплина позволяет провести точный анализ системы и получить характеристики производительности в явном виде.

Состояние системы в момент опроса очереди Q_1 (начала цикла) описывается набором случайных величин $X_{i,n}^{(k)}$, $k = \overline{1, K_i}$, $i = \overline{1, N}$, $n \geq 0$, где $X_{i,n}^{(k)}$ – число заявок в очереди Q_i на фазе k в n -й момент опроса очереди Q_1 . Процесс $X^{(n)} = (X_{1,n}^{(1)}, \dots, X_{1,n}^{(K_1)}, \dots, X_{N,n}^{(1)}, \dots, X_{N,n}^{(K_N)})$, $n \geq 0$ – K -мерный многотипный ветвящийся процесс с иммиграцией [32] с производящей функцией числа порождаемых частиц

$$f(\mathbf{s}) = \left(f^{(1,1)}(\mathbf{s}), \dots, f^{(1,K_1)}(\mathbf{s}), \dots, f^{(N,1)}(\mathbf{s}), \dots, f^{(N,K_N)}(\mathbf{s}) \right),$$

где $\mathbf{s} = (s_1^{(1)}, \dots, s_1^{(K_1)}, \dots, s_N^{(1)}, \dots, s_N^{(K_N)})$, для $i = \overline{1, N}$

$$f^{(i,k)}(\mathbf{s}) = s_i^{(k+1)}, \quad k = \overline{1, K_i - 1},$$

$$f^{(i,K_i)}(\mathbf{s}) = \tilde{B}_i \left(\sum_{j=1}^i \lambda_j (1 - s_j^{(1)}) + \sum_{i+1}^N \lambda_j (1 - f^{(j,1)}(\mathbf{s})) \right),$$

а производящая функция иммиграции имеет вид

$$g(\mathbf{s}) = \prod_{i=1}^N \tilde{S}_i \left(\sum_{j=1}^i \lambda_j (1 - s_j^{(1)}) + \sum_{i+1}^N \lambda_j (1 - f^{(j,1)}(\mathbf{s})) \right).$$

Закон псевдосохранения (выражение взвешенной суммы средних времен ожидания в системе) для многофазной плюзовой дисциплины имеет вид

$$\sum_{i=1}^N \rho_i \mathbf{E}[W_i] = \rho \frac{\rho}{1-\rho} \frac{b^{(2)}}{2b} + \rho \frac{s^{(2)}}{2s} + \frac{s}{2(1-\rho)} \left[\rho^2 - \sum_{i=1}^N \rho_i^2 \right] + \sum_{i=1}^N \mathbf{E}[M_i],$$

где M_i – среднее количество работы в очереди Q_i , остающейся в момент, когда сервер покидает очередь Q_i , $b^{(2)} = \sum_{i=1}^N \frac{\lambda_i}{\Lambda} b_i^{(2)}$, $b = \sum_{i=1}^N \frac{\lambda_i}{\Lambda} b_i$, $\Lambda = \sum_{i=1}^N \lambda_i$. В случае многофазной плюзовой дисциплины $\mathbf{E}[M_i] = \rho_i((K_i - 1) + \rho_i) \frac{s}{1-\rho}$, $i = \overline{1, N}$.

В [82] представлена другая модель резервирования и передачи данных в сетях PON: система поллинга с повторными вызовами и так называемыми периодами резервирования (glue periods). Заявки, поступающие в очередь Q_i , могут накапливаться в очереди и затем получить обслуживание лишь в течение так называемого периода резервирования (glue period), имеющего фиксированную длину G_i . Заявки, поступающие вне этого периода, уходят на орбиту и совершают повторные попытки занять место в очереди в период резервирования. Повторные попытки заявок с орбиты очереди Q_i образуют простейший поток с параметром ν_i , $i = \overline{1, N}$.

Производящая функция числа заявок в очередях в момент начала периода резервирования очереди Q_1 имеет явный вид

$$X(\mathbf{z}) = \prod_{m=0}^{\infty} K \left(h_1^{(m)}(\mathbf{z}), h_2^{(m)}(\mathbf{z}), \dots, h_N^{(m)}(\mathbf{z}) \right), \quad i = \overline{1, N},$$

где $h_i^{(0)}(\mathbf{z}) = z_i$, $h_i^{(n)}(\mathbf{z}) = h_i(h_i^{(n-1)}(\mathbf{z}), h_2^{(n-1)}(\mathbf{z}), \dots, h_N^{(i-1)}(\mathbf{z}))$, функции $h_i(\mathbf{z})$, $i = \overline{1, N}$ определяются как

$$h_i(\mathbf{z}) = f_i(z_1, \dots, z_i, h_{i+1}(\mathbf{z}), \dots, h_N(\mathbf{z})), \quad i = \overline{1, N-1}, \quad h_N(\mathbf{z}) = f_N(\mathbf{z}),$$

где $f_i(\mathbf{z}) = (1 - e^{-\nu_i G_i}) \beta_i(\mathbf{z}) + e^{-\nu_i G_i} z_i$, $\beta_i(\mathbf{z}) = \tilde{B}_i \left(\sum_{j=1}^N \lambda_j (1 - z_j) \right)$, $i = \overline{1, N}$, а функция $K(\mathbf{z})$ определяется следующим образом:

$$K(\mathbf{z}) = \prod_{i=1}^N \sigma_i(z_i, \dots, z_i, h_{i+1}(\mathbf{z}), \dots, h_N(\mathbf{z})) \prod_{i=1}^N e^{-G_i D_i(\mathbf{z})},$$

где $D_i(\mathbf{z}) = \sum_{j=1}^{i-1} \lambda_j (1 - z_j) + \lambda_i (1 - \beta_i(z_i, \dots, z_i, h_{i+1}(\mathbf{z}), \dots, h_N(\mathbf{z}))) + \sum_{j=i+1}^N \lambda_j (1 - h_j(\mathbf{z}))$, $\sigma_i(\mathbf{z}) = \tilde{S}_i \left(\sum_{j=1}^N \lambda_j (1 - z_j) \right)$ (здесь имеется в виду

ПЛС времени отключения от очереди Q_i , в отличие от основной модели в разделе 2), $i = \overline{1, N}$. В [82] также рассматривается вопрос оптимизации длительности периодов резервирования.

Беспроводные локальные сети. Авторы [83] анализируют непрерывную систему поллинга без отдельных очередей, представленную общей для всех заявок двумерной областью для ожидания, куда поступление заявок происходит случайным образом. Система моделирует беспроводную локальную сеть FWLAN (ferry based wireless local area network). FWLAN – это концепция беспроводной локальной сети, в которой связь между узлами (пользователями) и внешним миром или между узлами возможна через мобильный ретранслятор – «ферри». В такой сети изолированные узлы разбросаны по определенной области, и прямая глобальная связь между ними невозможна. Движущийся ретранслятор, перемещаясь по заранее заданному циклическому маршруту, связывается со статическими узлами (или пользователями) сети по беспроводному каналу связи. Мобильные базовые станции рассматриваются в контексте мобильных беспроводных сетей (Mobile Ad-Hoc Networks), автомобильных беспроводных сетей (Vanets) и беспроводных (статических) сенсорных сетей.

Сети на кристалле. В [84, 85] рассматривается случай трехфазного шлюзового обслуживания в качестве алгоритма планирования для управления доступом к среде (MAC) в сетевой подсистеме связи на интегральной схеме (Network on Chip) с учетом особенностей двумерной структуры связи.

Энергосбережение. В [86] представлен анализ системы поллинга с двумя очередями и ограниченным по времени обслуживанием очередей с точки зрения энергопотребления. Основное внимание в исследовании уделяется балансу энергопотребления процессоров в персональных компьютерах и их производительности для пользователей. Предполагается, что в модели поллинга скорость обслуживания заявок в очереди не постоянна, а зависит от загрузки очереди. Такой подход позволяет персональным компьютерам регулировать скорость обработки данных и оптимизировать энергопотребление в зависимости от объема выполняемых задач.

Контроль объема памяти. Модель, предложенная в [67, 68], имеет еще одно применение – контроль объема занятой памяти в центрах обработки данных, когда объем данных на отдельных дисках превышает определенный предел (порог), что приводит к неэффективной работе и требует очистки памяти.

Робототехника. В [87] описано использование механизма поллинга для обработки экологической и социальной информации, поступающей от роя роботов. Показано, что способность адаптации роя к нагрузке улучшается при уменьшении количества каналов связи между роботами, что требует решения соответствующей задачи оптимизации.

Спутниковые системы. В [88] исследована модель поллинга с учетом дуплексной связи с временным разделением для разработки единой систе-

мы управления в глобальных навигационных спутниковых системах с целью улучшения взаимодействия данных.

Производственные процессы. В [8] анализируется тандем из двух систем поллинга типа $M/M/1$, каждая из которых имеет две очереди. Данная модель представляет собой процесс производства алюминиевого проката. Каждый производственный этап моделируется системой поллинга, которая описывает процесс последовательной обработки различных типов сплавов.

8. Заключение

В настоящем обзоре систематизированы новые направления практического применения моделей стохастического поллинга на основе анализа результатов многочисленных статей отечественных и зарубежных авторов, опубликованных в последние годы. Рассмотрено применение таких моделей для оценки характеристик производительности и проектирования широкополосных беспроводных городских сетей, сотовых сетей 5G/6G и сетей Интернета вещей, транспортных систем и систем здравоохранения. Кратко отмечена эффективность применения моделей поллинга в других областях, включая робототехнику, спутниковые системы, пассивные оптические сети и производственные процессы. Для отдельных работ, представляющих теоретический и/или практический интерес, в обзоре приведено описание оригинальных математических моделей стохастического поллинга (отличных от основной модели, рассмотренной во втором разделе статьи) и результатов их исследования.

Обсуждены направления дальнейшего развития прикладных работ в области стохастического поллинга. К ним относятся исследования моделей поллинга с входящими коррелированными потоками и обобщенным РН-распределением времени обслуживания, адекватно описывающих функционирование современных вычислительных систем и сетей, применение и развитие методов машинного обучения и матрично-геометрического подхода для прогнозирования и оценки характеристик систем с централизованным механизмом управления. Кроме этого, перспективным направлением дальнейшего развития прикладных исследований является моделирование новых протоколов использования нескольких частотных каналов и обслуживания нескольких пользователей, а также новых механизмов доступа таких, как, например, Triggered Uplink Access (TUA).

Представленный обзор подводит итоги прикладных исследований по системам стохастического поллинга, позволяет сформулировать новые задачи и определить направления дальнейших работ в этой перспективной области теории очередей.

СПИСОК ЛИТЕРАТУРЫ

1. *Takagi H.* Analysis of polling systems. MIT Press, 1986.
2. *Borst S.C.* Polling systems. Amsterdam: Stichting Mathematisch Centrum, 1996.

3. *Borst S.C., Bozma O.J.* Polling: past, present, and perspective // TOP. 2018. V. 26. P. 335–369.
4. *Vishnevskii V.M., Semenova O.V.* Mathematical methods to study the polling systems // Autom. Remote Control. 2006. V. 67. P. 173–220.
5. *Vishnevsky V., Semenova O.* Polling Systems: Theory and Applications for Broadband Wireless Networks. LAMBERT Academic Publishing, 2012.
6. *Vishnevsky V., Semenova O.* Polling systems and their application to telecommunication networks // Mathematics. 2021. V. 9. No. 2. Article No. 117.
7. *Perel E., Yechiali U.* Two-queue polling systems with switching policy based on the queue that is not being served // Stochastic Models. 2017. V. 33. P. 430–450.
8. *Suman R., Krishnamurthy A.* Analysis of tandem polling queues with finite buffers // Annals of Operations Research. 2019. V. 293. P. 343–369.
9. *Boon M.A.A., van der Mei R.D., Winands E.M.M.* Applications of polling systems // Surveys in Operations Research and Management Science. 2011. V. 16. No. 2. P. 67–82.
10. *Dudin A.N., Klimenok V.I., Vishnevsky V.M.* The theory of queuing systems with correlated flows. Heidelberg, Germany: Springer, 2020.
11. *Chakravarthy S.R.* Queueing Models in Services – Analytical and Simulation Approach / Queueing Theory 2: Advanced Trends. John Wiley Sons, 2021. P. 33–80.
12. *Вишневецкий В.М., Ефросинин Д.В.* Теория очередей и машинное обучение. М.: ИНФРА-М, 2024.
13. *Efrosinin D., Vishnevsky V., Stepanova N.* Optimal scheduling in general multi-queue system by combining simulation and neural network techniques // Sensors. 2023. V. 23. No. 12. Article No. 5479.
14. *Winands E.M.M., Adan I.J.B.F., van Houtum G.J.* Mean value analysis for polling systems // Queueing Systems. 2006. V. 54. P. 35–44.
15. *Konheim A.G., Levy H., Srinivasan M.M.* Descendant set: an efficient approach for the analysis of polling systems // IEEE Trans. Commun. 1994. V. 42. No. 2–4. P. 1245–1253.
16. *Рыков В.В.* К анализу поллинг-систем // АИТ. 2008. № 6. С. 90–114.
17. *Yechiali U.* Analysis and control of polling systems // Performance Evaluations of Computer and Communication Systems. 1993. P. 630–650.
18. *Hirayama T.* Analysis of multiclass Markovian polling systems with feedback and composite scheduling algorithms // Annals of Operations Research. 2012. V. 198. No. 1. P. 83–123.
19. *Perel E., Perel N., Yechiali U.* A polling system with “Join the shortest – serve the longest” policy // Computers Operations Research. 2019. V. 114. Article No. 104809.
20. *Boon M.A.A., Winands E.M.M.* Heavy-traffic analysis of k -limited polling systems // Probability in the Engineering and Informational Sciences. 2014. V. 28. P. 451–471.
21. *Bekker R., Vis P., Dorsman J.L., van der Mei R.D., Winands E.M.M.* The impact of scheduling policies on the waiting-time distributions in polling systems // Queueing Systems: Theory and Applications. 2015. V. 79. P. 145–172.
22. *Bozma O.J., Groenendijk W.P.* Pseudo conservation laws in cyclic-service systems // Journal of Applied Probability. 1987. V. 24. P. 949–964.

23. IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: 320MHz Positioning, 2025.
24. *Vishnevsky V.M., Dudin A.N., Semenova O.V., Klimenok V.I.* Performance analysis of the *BMAP/G/1* queue with gated servicing and adaptive vacations // Performance Evaluation. 2011. V. 68. No. 5. P. 446–462.
25. *Vishnevsky V.M., Semenova O.V., Bui D.T.* Investigation of the stochastic polling system and its applications to broadband wireless networks // Autom. Remote Control. 2021. V. 82. No. 9. P. 1607–1613.
26. *Vishnevsky V.M., Semenova O.V., Bui D.T.* Using a machine learning approach for analysis of polling systems with correlated arrivals // Lecture Notes in Computer Science. 2021. V. 13144. P. 336–345.
27. *Vishnevsky V.M., Lyakhov A.I., Guzakov N.N.* An adaptive polling strategy for IEEE 802.11 PCF // Proceedings of 7th International Symposium on Wireless Personal Multimedia Communications, Abano Terme, Italy, September 12–15. 2004. V. 1. P. 87–91.
28. *Vishnevsky V., Lyakhov A.* Analytical study of IEEE 802.11 PCF for regional and metropolitan area networks // Cybernetics and Information Technologies. 2005. V. 5. No. 2. P. 117–136.
29. *Вишневецкий В., Ляхов А., Портной С., Шахнович И.* Широкополосные беспроводные сети передачи информации. М.: Техносфера, 2005. 592 с.
30. *Leonovich A., Ferng H.-W.* Modeling the IEEE 802.11e HCCA mode // Wireless Networks. 2013. V. 19. No. 5. P. 771–783.
31. *Meurfroyt T.M.M., Boon M.A.A., Borst S.C., Boxma O.J.* Performance of large-scale polling systems with branching-type and limited service // Performance Evaluation. 2019. V. 133. P. 1–24.
32. *Resing J.A.C.* Polling systems and multitype branching processes // Queueing Systems. 1993. V. 13. P. 413–426.
33. *Vishnevsky V.M., Semenova O.V., Nguyen V.H., Dang M.C.* Batch service polling system: mathematical analysis and simulation modeling // Lecture Notes in Computer Science. 2024. V. 14123. P. 142–155.
34. *Vishnevsky V.M., Avramenko Y.A., Kalmykov N.S., Nguyen V.H.* Broadband Wireless networks based on tethered high-altitude unmanned platforms // Communications in Computer and Information Science. 2024. V. 2129. P. 101–112.
35. *Belmekki B.E.Y., Alouini M.-S.* Unleashing the potential of networked tethered flying platforms: prospects, challenges, and applications // IEEE Open Journal of Vehicular Technology. 2022. V. 3. P. 278–320.
36. *Marques M.N., Magalhaes S.A., Dos Santos F.N., Mendonca H.S.* Tethered unmanned aerial vehicles – a systematic review // Robotics. 2023. V. 12. No. 4. P. 117.
37. *Zhao W., Zhang J., Li D.* Clustering and beamwidth optimization for UAV-assisted wireless communication // Sensors. 2023. V. 23. No. 23. Article No. 9614.
38. *Назаров А.А., Шульгина К.С., Салимзянов Р.Р. и др.* Исследование математической модели беспроводной сети передачи данных в виде циклической системы случайного множественного доступа // Технологии безопасности жизнедеятельности. 2023. № 2. С. 58–65.

39. *Nazarov A., Phung-Duc T., Shulgina K., et al.* Cyclic retrial queue for building data transmission networks // Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering. 2023. V. 482. P. 167–179.
40. *Dorsman J.-P.L., Borst S.C., Boxma O.J., et al.* Markovian polling systems with an application to wireless random-access networks // Performance Evaluation. 2015. V. 85–86. P. 33–51.
41. *Liu Z., Chu Y., Wu J.* On the three-queue priority polling system with threshold service policy // Journal of Applied Mathematics and Computing. 2017. V. 53. P. 445–470.
42. *Fiems D., Altman E.* Gated polling with stationary ergodic walking times, Markovian routing and random feedback // Annals of Operations Research. 2012. V. 198. No. 1. P. 145–164.
43. *Ding H., Li C., Bao L., et al.* Research on multi-level priority polling MAC protocol in FPGA tactical data chain // IEEE Access. 2019. V. 7. P. 33506–33516.
44. *Chernova N., Foss S., Kim B.* On the stability of a polling system with an adaptive service mechanism // Annals of Operations Research. 2012. V. 198. P. 125–144.
45. *Morais D.H.* Wi-Fi 6/6E and Wi-Fi 7 Overview. In: 5G/5G-Advanced, Wi-Fi 6/7, and Bluetooth 5/6. Springer, 2025.
46. *Karamyshev A., Levitsky I., Bankov D., et al.* A Tutorial on Wi-Fi 8: the journey to ultra high reliability // Problems of Information Transmission. 2025. V. 61. P. 164–210.
47. *Rahaman M.* A review on Internet of Things-IoT- architecture, technologies, future applications & challenges // International Journal of Science and Business. 2022. V. 14. No. 1. P. 80–92.
48. *Nguyen D.C., Ding M., Pathirana P.N., et al.* 6G Internet of Things: a comprehensive survey // IEEE Internet of Things Journal. 2022. P. 359–383.
49. *Wei Z., Jia J., Niu Y., Wang L., et al.* Integrated sensing and communication channel modeling: a survey // IEEE Internet of Things Journal. 2025. P. 18850–18864.
50. *Yang Z., Mao L., Ding H., et al.* Research on two-level priority polling access control protocol based on continuous time // 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 136–140, Chengdu, China.
51. *Yang Z., Mao L., Yan B., Wang J., Gao W.* Performance analysis and prediction of asymmetric two-level priority polling system based on BP neural network // Applied Soft Computing. 2021. V. 99. Article No. 106880.
52. *Ar-Reyouchi E.M., Maslouhi I., Ghoumid K.* A new fast polling algorithm in wireless mesh network for narrowband Internet of Things // Telecommunication Systems. 2020. V. 74. No. 3. P. 405–410.
53. *Cao J., Feng W., Chen Y. et al.* Performance analysis of a polling model with BMAP and across-queue state-dependent service discipline // IEEE Access. 2019. V. 7. P. 127230–127253.
54. *Молчанов Д.А., Безишев В.О., Самуйлов К.Е. и др.* Сети 5G/6G: архитектура, технологии, методы анализа и расчета. М.: Российский университет дружбы народов, 2022.
55. *Cudak M., Ghosh A., Ghosh A., et al.* Integrated access and backhaul: A key enabler for 5G millimeter-wave deployments // IEEE Communications Magazine. 2021. V. 59. No. 4. P. 88–94.

56. *Gupta M., Rao A., Visotsky E., et al.* Learning link schedules in self-backhauled millimeter wave cellular networks // IEEE Transactions on Wireless Communications. 2020. V. 19. No. 12. P. 8024–8038.
57. 3GPP TR 38.901 V14.1.1. Study on channel model for frequencies from 0.5 to 100 GHz (Release 17).
58. *Nikolaev D., Gaidamaka Y.* Leaf node polling model analysis in an integrated access and backhaul network / Information Technologies and Mathematical Modelling. Queueing Theory and Applications. ITMM WRQ 2023. Communications in Computer and Information Science. 2023. V. 2163.
59. *Бесчастный В.А., Мачнев Е.А., Острикова Д.Ю. и др.* Вероятностные характеристики процесса обслуживания пакетов в узле сети IAB // Электросвязь. 2024. № 1. С. 43–46.
60. *Nikolaev D., Gorshenin A., Gaidamaka Y.* Polling model for analysis of round-trip time in the IAB network // Distributed Computer and Communication Networks. DCCN 2024. Lecture Notes in Computer Science. 2025. V. 15460. P. 219–241.
61. *Nikolaev D.I., Beschastny V.A., Gaidamaka Y.V.* Two-queue polling system as a model of an integrated access and backhaul network node in half-duplex mode // Discrete and Continuous Models and Applied Computational Science. 2024. V. 32. No. 4. P. 362–369.
62. *Gaidamaka Yu.V.* Model with threshold control for analysing a server with SIP protocol in the overload mode // Automatic Control and Computer Science. 2013. V. 47. P. 211–218.
63. *Shorgin S., Samouylov K., Gaidamaka Y., et al.* Polling system with threshold control for modeling of SIP server under overload // Advances in Intelligent Systems and Computing. 2014. V. 240. P. 97–107.
64. *Singh V., Kavitha V.* Fair opportunistic schedulers for Lossy Polling systems // 2022. arXiv preprint arXiv:2208.05846.
65. *Czerniak O., Altman E., Yechiali U.* Orchestrating parallel TCP connections: cyclic and probabilistic polling policies // Performance Evaluation. 2012. V. 69(3-4). P. 150–163.
66. *Nimisha M., Manoharan M., Krishnamoorthy A.* Polling models: a short survey and some new results // Queueing Models and Service Management. 2024. V. 7. No. 1. P. 25–42.
67. *Avrachenkov K., Perel E., Yechiali U.* Finite-buffer polling system with threshold-based switching policy // TOP. 2016. V. 24. P. 541–571.
68. *Jolles A., Perel E., Yechiali U.* Alternating server with non-zero switchover times and opposite-queue threshold-based switching policy // Performance Evaluation. 2018. V. 126. P. 22–38.
69. *Choi D.I., Lim D.-E.* An analytical analysis of Markovian queueing model with an alternating server and state-dependent alternating priority policy // ICIC Express Letters. 2024. V. 18. No. 4. P. 385–391.
70. *Miculescu D., Karaman S.* Polling-systems-based autonomous vehicle coordination in traffic intersections with no traffic signals // IEEE Transactions on Automatic Control. 2020. V. 65. P. 680–694.

71. *Dudin A., Sinyugina Y.* Analysis of the polling system with two Markovian arrival flows, finite buffers, gated service and phase-type distribution of service and switching times // ITMM 2021. Communications in Computer and Information Science. V. 1605. P. 1–15.
72. *Dudin A., Dudina O.* Analysis of polling queueing system with two buffers and varying service rate // Distributed Computer and Communication Networks. DCCN 2024. Lecture Notes in Computer Science. V. 15460. P. 129–143.
73. *Семенова О.В., Лыков С.С.* Применение RFID-технологии для управления работой регулируемых перекрестков / Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN-2013). М.: ТЕХНОСФЕРА. 2013. P. 318–324.
74. *Абрамян В.Л., Вишневецкий В.М., Ларионов А.А.* Применение радиочастотной идентификации на транспорте // Проблемы управления. 2024. № 1. С. 3–16.
75. *Ran W., Liu S., Zhang Z.* A polling-based dynamic order-picking system considering priority orders // Complexity. 2020. V. 2020. Article ID 4595316. 15 pages.
76. *Claeys D., Adan I., Bozma O.* Stochastic bounds for order flow times in parts-to-picker warehouses with remotely located order-picking workstations // European Journal of Operational Research. 2016. V. 254. No. 3. P. 895–906.
77. *Kaur M., Bajaj R., Kaur N.* Wireless Body Area Network: performance analysis of polling access MAC protocol // 2021 2nd Global Conference for Advancement in Technology (GCAT). 2021. P. 1–7.
78. *He M., Guan Z., Wu Z., et al.* A polling access control with exhaustive service in wireless body area networks for mobile healthcare using the sleeping schema // Journal of Ambient Intelligence and Humanized Computing. 2019. V. 10. No. 10. P. 3761–3774.
79. *Wang W., Yang Z., Ding H.* Application of polling scheduling in mobile edge computing // Axioms. 2023. V. 17. No. 7. Article No. 709.
80. *Antunes N., Fricker C., Roberts J.* Stability of multi-server polling system with server limits // Queueing Systems. 2011. V. 68. P. 229–235.
81. *van der Mei R.D., Roubos A.* Polling models with multi-phase gated service // Annals of Operations Research. 2012. V. 198. No. 1. P. 25–56.
82. *Abidini M.A., Bozma O., Resing J.* Analysis and optimization of vacation and polling models with retrials // Performance Evaluation. 2016. V. 98. P. 52–69.
83. *Kavitha V., Altman E.* Continuous polling models and application to ferry assisted WLAN // Annals of Operations Research. 2012. V. 198. P. 185–218.
84. *Bao L., Zhao D., Zhao Y.* A priority-based polling scheduling algorithm for arbitration policy in Network on Chip // Journal of Electronics (China). 2012. V. 29. No. 1–2. P. 120–127.
85. *Ling Y., Liu C., Li Y.* Study on queue strategy of gated polling multi-access communication system // Recent Advances in Computer Science and Information Engineering. Lecture Notes in Electrical Engineering. 2012. V. 124. P. 99–105.
86. *Bozma O.J., Saxena M., Janssen A.* Two queues with time-limited polling and workload-dependent service speeds // Stochastic Models. 2021. V. 37. No. 2. P. 265–299.

87. *Talamali M.S., Saha A., Marshall J.A., Reina A.* When less is more: Robot swarms adapt better to changes with constrained communication // *Science Robotics*. 2021. V. 6. eabf1416
88. *Yan Z., Zhao K., Li W., et al.* Topology design for GNSSs under polling mechanism considering both inter-satellite links and ground-satellite links // *IEEE Transactions on Vehicular Technology*. 2022. V. 71. No. 2. P. 2084–2097.

Статья представлена к публикации членом редколлегии А.И. Ляховым.

Поступила в редакцию 02.10.2025

После доработки 27.10.2025

Принята к публикации 28.10.2025