

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Феоктистова Варвара Николаевна

**Математические проблемы управления потоковыми
переключательными сетями**

01.01.09 — Дискретная математика и математическая кибернетика

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата физико-математических наук

Санкт-Петербург
2012

Работа выполнена на кафедре теоретической кибернетики математико-механического факультета Санкт-Петербургского государственного университета.

Научный руководитель: доктор физико-математических наук,
профессор Матвеев Алексей Серафимович

Официальные оппоненты: Петров Николай Николаевич,
доктор физико-математических наук, профессор,
(Санкт-Петербургский государственный
университет, профессор)

Андриевский Борис Ростиславич,
доктор технических наук, доцент,
(Институт проблем машиноведения
Российской академии наук (ИПМаш РАН),
ведущий научный сотрудник)

Ведущая организация: Санкт-Петербургский национальный
исследовательский университет информационных
технологий, механики и оптики.

Защита состоится " __ " _____ 2012 года в __ часов на заседании диссертационного совета Д 212.232.29 при Санкт-Петербургском государственном университете по адресу: 199034, Санкт-Петербург, Университетская наб., 7/9, ауд. 133.

С диссертацией можно ознакомиться в Научной библиотеке им. М. Горького Санкт-Петербургского государственного университета по адресу: 199034, Санкт-Петербург, Университетская наб., 7/9.

Автореферат разослан " __ " _____ 2012 года.

Ученый секретарь
диссертационного совета

В.М. Нежинский

Общая характеристика работы

Актуальность темы. Неуклонно возрастающее распространение и значение сетевых комплексов и технологий и повышение требований к их качеству мотивировали интенсивный интерес к разработке эффективных (в идеале, оптимальных) алгоритмов управления большими сетевыми структурами, состоящими из множества частей и элементов. Развитие этого направления, оцениваемого авторитетными группами международных экспертов как несущее потенциал значительного инновационного продвижения, потребовало междисциплинарных усилий и привлекло внимание многих специалистов по математической теории и практике управления. Это развитие существенно осложняется как типичными для больших сетей особенностями, среди которых высокий порядок системы и связанное с этим 'проклятие размерности', децентрализованный и локальный характер управления и доступных данных, так и разнообразными неопределенностями, спецификой управляющих воздействий и другими факторами. Как следствие, несмотря на активное и широкое внимание к этой тематике, удовлетворительные решения были найдены лишь для ограниченного класса возникающих в этой области задач, который далеко не исчерпывает ее проблематику.

Диссертация посвящена исследованию математических проблем управления потоковыми переключательными сетями, то есть сетями, в которых несколько серверов обслуживает ряд потоков, регулярно переключаясь между ними, и для которых переключение является существенной частью управляющего воздействия. Как математический объект такие сети традиционно используют для моделирования определенных аспектов функционирования гибких производственных систем, коммуникационных, информационных, компьютерных и платежных сетей, транспортных систем, процессов химической кинетики и др. Диссертация сфокусирована на жидкостных (детерминированных) моделях рассматриваемых сетей. Они представляют один из основных используемых в обсуждаемой области типов математических моделей; их изучению посвящены работы многих авторов, в том числе M. Bramson, S. Gershwin, J.G. Dai, P.R. Kumar, T.I. Seidman, H. Chen, R. Herman, S.P. Meyn, B. Hajek, D.N. Nicol, L. Presti, J. Hespanha и др. Жидкостные модели дополняют и аппроксимируют стохастические модели теории систем массового обслуживания. Интерес к жидкостным моделям заметно активизировался в последние два десятилетия. Причины этого включают назревшую потребность в целостном исследовании больших сетей. Для них жидкостная модель в силу своей "локальной" относительной простоты предоставляет практическую возможность содержательного анализа эффектов, связанных с взаимодействием и координацией множества агентов в большой системе.

На уровне математической абстракции некоторые связанные с сетевыми системами задачи управления широко исследовались и ранее. Это в частности касается эффективного распределения сетевых ресурсов и управления буфе-

рами. Ряд возникающих здесь вопросов сводим к проблемам традиционной теории расписаний, например, нахождение оптимального (субоптимального) распределения ресурсов при наличии ограничений. Хотя обширная литература на эту тему предлагает ряд проработанных подходов к проблеме, в целом они ориентированы на статические алгоритмы, реализуемые централизованно, как правило, в виде временной программы и для неизменной и известной обстановки. Это противоречит реалиям многих практических систем, функционирующим децентрализованно в динамичной и непредсказуемой среде. Следствием является констатируемый разрыв между теорией и практикой, квалифицируемый как один из основных тормозов на пути повышения эффективности многих сетевых систем. Разрыв связан с тем, что для парирования неопределенностей и адаптации к изменчивым условиям на практике применяют алгоритмы (протоколы) управления типа обратной связи. Однако в настоящее время большинство из них в той или иной степени опирается на эвристику. При этом неоднократно обнаруживалось, что поведение, вызванное применением основанных на эвристике протоколов, может не соответствовать первоначальным ожиданиям и даже быть неприемлемым.

Системный подход к преодолению упомянутого разрыва был недавно предложен Е. Lefebvre и J.E. Rooda. В качестве цели он ставит развитие общего метода синтеза протокола управления, превращающего априори заданный периодический процесс в глобальный аттрактор замкнутой системы. При этом особый интерес представляет случай, когда этот процесс оптимален или субоптимален. Упомянутыми авторами был предложен специальный метод такого рода, связанный с построением функции Ляпунова. Однако его известные применения ограничены простейшими примерами сетей невысокой сложности. Типичное для рассматриваемых сетей "проклятие размерности" быстро трансформирует необходимое для реализации метода вычисление функции Ляпунова в трудно, а подчас и вовсе невыполнимую задачу по мере увеличения числа элементов сети. Однако даже после вычисления этой функции построение протокола не гарантировано, не подкреплено общей методикой и рассматривается как индивидуальная для каждого приложения задача.

Цель работы. Работа ориентирована на преодоление указанного разрыва в части, касающейся жидкостных моделей потоковых переключательных сетей.

Задачи исследования. Развитие математической теории и конструктивных методов синтеза протоколов управления большими сетями, обеспечивающих сходимость всех процессов в системе к априори заданному периодическому процессу. Обоснование оптимальности широко применяемого на практике класса протоколов управления производственными линиями, в которых реализован принцип обратной связи от текущего запроса (pull-type).

Методы исследований. В работе применялись методы качественной теории динамических систем, математической теории управления, выпуклого анализа и линейной алгебры.

Основные положения и результаты, выносимые на защиту.

Разработан конструктивный метод синтеза протоколов децентрализованного управления потоковыми переключательными сетями, опирающийся на идеи метода сечений Пуанкаре и проиллюстрированный на примере общей жидкостной модели такой сети. В основе метода лежит полученный новый критерий глобальной асимптотической устойчивости положения равновесия стационарной динамической системы с дискретным временем и кусочно-аффинным, монотонным и непрерывным динамическим оператором. Получены необходимые (а в случае сети Кумара-Сейдмана и достаточные) условия оптимальности процесса для потоковой сети с разделением процессора и производственной сети Кумара-Сейдмана. Для указанных сетей, а также для общей односерверной сети решена задача устойчивой генерации заданного периодического процесса. Обоснована оптимальность специального класса протоколов управления производственными линиями.

Научная новизна. Все результаты, представленные в диссертации, являются новыми. Впервые разработан конструктивный подход к проблеме устойчивой генерации требуемых периодических режимов в больших потоковых переключательных сетях.

Теоретическая и практическая ценность. Полученные результаты могут быть использованы для дальнейшего теоретического исследования различных вопросов, связанных с управлением потоковыми переключательными сетями, а также в педагогической практике. Разработанные в диссертации конструктивные методы могут непосредственно применяться при практическом синтезе протоколов управления производственными, транспортными, коммуникационными, компьютерными и другими сетями указанного типа.

Апробация работы. Основные результаты диссертационной работы докладывались и обсуждались на всемирном конгрессе Международной Федерации Автоматического Управления «18th IFAC World Congress» (Milan, Italy, 2011), на международной конференции по физике и управлению «5th International Scientific Conference on Physics and Control: PHYSCON 2011» (Leon, Spain, 2011), на международном симпозиуме по проблемам управления потоками информации в производственных системах «13th IFAC Symposium on Information Control Problems in Manufacturing» (Moscow, Russia, 2009), на международном симпозиуме по интеллектуальным производственным системам «10th IFAC Workshop on Intelligent Manufacturing Systems» (Lisbon, Portugal, 2010), на научном семинаре кафедры исследования операций СПбГУ под руководством зав.каф., д.ф.-м.н., профессора Н.Н. Петрова (в 2012 г.), на научном семинаре лаборатории «Управление сложными системами» Института Проблем Машинovedения РАН под руководством зав. лаб., д.т.н., профессора А.Л. Фрадкова (в 2011 г.), а также на научном семинаре кафедры теоретической кибернетики СПбГУ под руководством зав. каф., чл.корр. РАН, д.ф.-м.н.,

профессора В.А. Якубовича (в 2008—2011 гг.).

Результаты диссертации были частично использованы в исследованиях по грантам РФФИ 11-08-01218-а, 12-01-00808 и ФЦП «Научные и научно-педагогические кадры инновационной России» (проект 1.1-111-128-033).

Публикации. Основные результаты диссертации отражены в публикациях [1]–[7]. В совместных работах [1], [4]–[6] диссертанту принадлежат формулировки и доказательства теорем, а соавторам — постановка задачи, выбор методов решения и — применительно к работам [5] и [6] — частичное вычисление оптимального процесса. В работах [3] и [7] диссертанту принадлежит доказательство оптимальности протокола управления, а соавторами обоснована устойчивость управляемой этим протоколом производственной линии.

Статьи [1–3] опубликованы в изданиях, входящих в перечень рецензируемых научных журналов и изданий; публикации [4,6] имеют признаки соответствия этому перечню.

Структура и объем работы. Диссертация состоит из введения, восьми глав и списка литературы, содержащего 101 наименование. Текст диссертации изложен на 165 страницах и содержит 20 рисунков.

Содержание работы

Во введении изложена история вопроса, приведено обоснование актуальности темы диссертации, обозначена цель исследования, охарактеризован вклад диссертации и её структура и перечислены основные результаты.

Первая глава посвящена математической постановке основной задачи и описанию предложенного метода ее решения. Рассматривается система, состоящая из s серверов и N буферов и получающая извне f потоков, интерпретируемых как потоки жидкости. Поток из i -го источника поступает на постоянной скорости λ_i в буфер b_i , затем перемещается серверами из буфера в буфер и в конечном счете выводится из системы. Пути потоков по буферам неизменны и представлены ориентированным графом с множеством вершин $\{1, \dots, N, \otimes_1, \dots, \otimes_f, \otimes_{\text{out}}\}$. Здесь \otimes_i — источник i -ого потока, \otimes_{out} — внешний сток, вершины $n \in [1 : N]$ соответствуют буферам, а ребра графа указывают пути перемещения содержимого буферов, именуемого *работой*. Источники \otimes_j не имеют входящих ребер, сток — исходящих ребер, каждый буфер имеет входящие ребра и одно исходящее ребро, в графе нет циклов.

Перемещение работы из буферов по ребрам графа осуществляют сервера. Сервер извлекает содержимое буфера n со скоростью $0 \leq u_n(t) \leq \mu_n$, где $\mu_n > 0$ задано, а $u_n(t)$ — управляющая переменная. В зоне обслуживания $I_\sigma \subset [1 : N]$ сервера σ находится несколько буферов, причем $I_{\sigma'} \cap I_{\sigma''} = \emptyset$ при $\sigma' \neq \sigma''$ и $I_1 \cup \dots \cup I_s = [1 : N]$. В каждый момент сервер способен обслуживать не более одного буфера и вынужден переключаться между ними; переключение из буфера n' в n'' занимает фиксированное время $\delta_{n' \rightarrow n''} \geq 0$.

Управление системой состоит в выборе для каждого сервера скорости извлечения содержимого $u_n(t)$ из обслуживаемого буфера n , момента прекращения обслуживания и следующего буфера. Во многих случаях оптимальна максимальная доступная скорость u_n . Тогда задача управления сводится к выбору моментов и порядка переключения, а алгоритм переключения оказывается ключевым инструментом обеспечения требуемого поведения системы.

Текущее состояние (X, Q) системы включает *непрерывную* $X = (x_1, \dots, x_N)$ и *дискретную* $Q = (q_1, \dots, q_s)$ компоненты, где x_n - содержимое буфера n и $q_\sigma \in \widehat{I}_\sigma := I_\sigma \cup \{\ominus\}$ — состояние сервера σ : он обслуживает буфер q_σ , если $q_\sigma \neq \ominus$, и переключается, если $q_\sigma = \ominus$. *Процессом* назовем пару функций $[X(t), Q(t)]$ времени t , описывающую возможный вариант развития событий в системе. Это означает, что для любого сервера σ активные $n \in [1 : N]$ и переключательные \ominus значения кусочно-постоянной функции $q_\sigma(t) \in \widehat{I}_\sigma$ чередуются, соблюдается требуемое время экспозиции переключательного значения, а для любого $n \in [1 : N]$ функция $x_n(\cdot)$ абсолютно непрерывна, $x_n(t) \geq 0$ и $\dot{x}_n(t) = \sum_{j:j \rightarrow n \text{ ветвь графа}} u_j(t) - u_n(t)$, где $u_{\otimes_i}(t) \equiv \lambda_i$ и $0 \leq u_k(t) \leq \mu_k$, $(k \neq q_\sigma(t) \forall \sigma \Rightarrow u_k(t) = 0)$ для $k \in [1 : N]$.

Протоколом (алгоритмом) управления называется система правил, определяющих для каждого сервера скорость $u_n(t)$ извлечения содержимого из обслуживаемого буфера n , время прекращения обслуживания и следующий обслуживаемый буфер. Соответствующие решения принимаются следуя принципу обратной связи на основе данных о времени t и состоянии системы до текущего момента $X|_{[0,t]}, Q|_{[0,t]}$ (где 0 — начальное время).

В рассматриваемых системах простейшим типом аттрактора как правило является периодический процесс (за исключением тривиального случая, когда область обслуживания каждого сервера состоит из одного буфера).

Рассматривается следующая задача:

- Р)** По заданному периодическому процессу π^0 требуется построить протокол управления, который превращает π^0 в глобальный аттрактор.

Другими словами, любой процесс в системе, управляемой этим протоколом, сходится к π^0 . Это особенно ценно, если процесс π^0 обладает какими-либо полезным свойством, например, оптимален или суб-оптимален. Тогда протокол обеспечивает автоматическое выведение сетевой системы в оптимальный режим, его устойчивое поддержание и компенсацию отклонений от оптимума.

В диссертации предложен следующий подход к решению задачи **Р**, во многом мотивированный стремлением к преодолению "проклятия размерности" и созвучный идеям метода сечений Пуанкаре. Вначале процесс π^0 , рассматриваемый на периоде, делится на простые части — фазы

$$\mathfrak{C} = \mathfrak{P}_0, \mathfrak{P}_1, \mathfrak{P}_2, \dots, \mathfrak{P}_{c-1}; \quad (1)$$

каждая фаза \mathfrak{P}_i ассоциируется с пробегаемой дискретным состоянием Q цепочкой значений (единственным значением в простейшем случае). Протокол

состоит в принудительном прохождении через периодически повторяемый цикл фаз (1) с применением внутри каждой из них индивидуального правила управления. Оно определяет скорости работы серверов, время окончания фазы и моменты скачкообразной эволюции состояния Q вдоль требуемой цепочки значений. Выбор такого правила порождает динамический оператор $T^{\mathfrak{F}_i}$ фазы \mathfrak{F}_i , который переводит непрерывное состояние системы X в начале фазы в аналогичное состояние в конце этой фазы. Оператором *монодромии* назовем композицию операторов $T^{\mathfrak{F}_i}$ в соответствии с циклом (1)

$$M = T^{\mathfrak{F}_{c-1}} \circ T^{\mathfrak{F}_{c-2}} \circ \dots \circ T^{\mathfrak{F}_1} \circ T^{\mathfrak{F}_0}. \quad (2)$$

Состояния X , рассматриваемые в моменты начала циклов, эволюционируют посредством итераций этого оператора. Основная задача, решаемая при выборе правил управления фазами, — обеспечить, чтобы все траектории, порождаемые этой рекурсией $X(k+1) = M[X(k)]$, сходились к положению равновесия X^0 , отвечающему требуемому периодическому процессу π^0 . Кроме того каждое правило должно порождать соответствующую часть π^0 .

В типичном случае операторы $T^{\mathfrak{F}_i}$ и M кусочно-аффинны, то есть аффинны на элементах полиэдрального разбиения фазового пространства. Обычно число элементов разбиения невелико для операторов $T^{\mathfrak{F}_i}$, но быстро нарастает при последовательном формировании композиции (2), в итоге достигая астрономических значений, особенно при большом числе серверов, буферов, и фаз. В итоге "явное" описание оператора M трудоёмко и часто практически невозможно, что наряду с узкой номенклатурой управляющих воздействий (в основном, переключения серверов) препятствует применению известных критериев устойчивости и методов стабилизации дискретных динамических систем. Дело осложняется тем, что на этапе синтеза оператор монодромии M еще не существует, а задача состоит в обнаружении и использовании зависимостей между M и правилами управления фазами.

Для преодоления указанной проблемы в диссертации найден и обоснован новый критерий глобальной асимптотической устойчивости положения равновесия стационарной нелинейной дискретной динамической системы $X(k+1) = M[X(k)]$. Его принципиальная особенность — возможность по-фазовой проверки: критерий указывает список свойств, наличие которых у оператора монодромии гарантирует устойчивость и которые наследуются композицией операторов, обладающих этими свойствами. В результате построение правил управления фазами освобождается от проклятия размерности за счет локализации цели построения: обеспечить наличие упомянутых свойств у динамического оператора каждой отдельной фазы.

Вторая глава посвящена проблеме устойчивости положений равновесия стационарных нелинейных дискретных динамических систем

$$x(k+1) = \mathcal{T}[x(k)], \quad k = 0, 1, 2, \dots \quad (3)$$

Исследование сосредоточено на специальных системах, представляющих интерес для охарактеризованных приложений.

В р. 2.1 изложены основные определения. В частности, указано, что неравенства между векторами $x, y \in \mathbb{R}^m$ понимаются покомпонентно, и введены следующие понятия.

Определение 2.1 Оператор $\mathcal{T} : K_+^p := \{x \in \mathbb{R}^p : x \geq 0\} \rightarrow K_+^r$ назовём:

- а) кусочно-аффинным если K_+^p можно покрыть $K_+^p = C_1 \cup \dots \cup C_k$ конечным числом выпуклых многогранных множеств $C_i \subset K_+^p$ (называемых ячейками), на каждом из которых \mathcal{T} является аффинным, т.е. $\mathcal{T}(x) = A_i x + a_i \forall x \in C_i$, где A_i — линейный оператор и $a_i \in \mathbb{R}^r$;
- б) доминантным (строго доминантным) кусочно аффинным, если дополнительно $a_i \geq 0$ (соответственно $a_i > 0$) для любой ячейки;
- в) монотонным, если $\mathcal{T}(x) \leq \mathcal{T}(y) \forall x \leq y$.

Далее КАМН = "кусочно-аффинный монотонный непрерывный".

В р. 2.2 представлен следующий основной результат главы.

Теорема 2.1 Предположим, что КАМН оператор $\mathcal{T}[\cdot] : K_+^p \rightarrow K_+^p$ имеет неподвижную точку $\mathcal{T}[x_*] = x_* \in K_+^p$ и некоторая его итерация \mathcal{T}^m строго доминантна. Тогда неподвижная точка единственна и притягивает $x(k) \xrightarrow{k \rightarrow \infty} x_*$ любую траекторию системы (3) (где $x(0) \in K_+^p$).

Возможность по-фазовой проверки условий Теоремой 2.1 обоснована в р. 2.3.

Оставшаяся часть главы 2 посвящена общей теории, из которой вытекает Теорема 2.1 и которая может быть использована для нахождения других аналогичных критериев устойчивости. Так как эта теория нацелена на результаты в духе теоремы Ляпунова об устойчивости по первому приближению, в ней центральное значение приобретает исследование спектральных свойств нелинейной производной негладкого динамического оператора \mathcal{T} из (3), которая для КАМН операторов \mathcal{T} является (положительно) однородным КАМН оператором T , т.е. $T[\theta x] = \theta T[x] \forall \theta \in [0, +\infty), x \geq 0$.

Р. 2.4 посвящен изучению собственные чисел и векторов однородных кусочно-аффинных монотонных непрерывных (ОКАМН) операторов \mathcal{T} .

Определение 2.2 Если $\mathcal{T}[e] = \rho e$ для $\rho \in \mathbb{R}$ и $e \geq 0, e \neq 0$, число ρ и вектор e назовем собственным числом и собственным вектором ОКАМН оператора \mathcal{T} , соответственно. Собственный вектор назовем неустойчивым, если существует такое $\varepsilon > 0$, что любой ОКАМН-оператор \mathcal{T}_* из доминантной ε -окрестности $\mathfrak{V}_\varepsilon := \{\mathcal{T}_* : \mathcal{T}(x) < \mathcal{T}_*(x) \forall x \geq 0, x \neq 0; |\mathcal{T}_*(x) - \mathcal{T}(x)| < \varepsilon \forall x \geq 0, |x| = 1\}$ оператора \mathcal{T} не имеет собственных векторов в ε -окрестности e ; и устойчивым в противном случае. Если собственное число имеет хотя бы один устойчивый собственный вектор, оно называется устойчивым.

Теорема 2.2 *ОКАМН-оператор \mathcal{T} имеет одно и только одно устойчивое собственное число $\rho = \rho(\mathcal{T})$. Оно неотрицательно, доминирует $\rho \geq \varrho$ прочие собственные числа ϱ и зависит от оператора монотонно и непрерывно: $\mathcal{T}(x) \leq \mathcal{T}_*(x) \forall x \in K_+^p \Rightarrow \rho[\mathcal{T}] \leq \rho[\mathcal{T}_*]$ и $\mathcal{T}_* \in \mathfrak{B}_\varepsilon \wedge \varepsilon \rightarrow 0 \Rightarrow \rho[\mathcal{T}] \rightarrow \rho[\mathcal{T}_*]$.*

В р. 2.5 вводится и изучается однородное ядро КАМН оператора \mathcal{T} , представляющее собой аппроксимацию \mathcal{T} однородным оператором вблизи бесконечно удаленной точки. Направление, заданное вектором e , назовем *рецессивным* для выпуклого множества $C \subset \mathbb{R}^p$, если $x + te \in C \forall t \geq 0$ для некоторого $x \in C$; совокупность всех таких векторов обозначим через $Ex[C]$.

Теорема 2.3 *Пусть C_1, \dots, C_k - ячейки КАМН-оператора \mathcal{T} . Тогда преобразование $x \geq 0 \mapsto \mathcal{T}^\infty(x) := A_i x$ для $x \in Ex(C_i)$ является корректно определенным ОКАМН-оператором и $K_+^p = Ex[C_1] \cup \dots \cup Ex[C_k]$.*

Это преобразование называется *однородным ядром* оператора \mathcal{T} . Следующая лемма демонстрирует его аппроксимирующие свойства.

Лемма 2.1 $|x|^{-1}\mathcal{T}(x) - |x|^{-1}\mathcal{T}^\infty(x) \rightarrow 0$ при $|x| \rightarrow \infty, x \geq 0$.

Теорема 2.4 *Если $\rho[\mathcal{T}^\infty] < 1$ для КАМН-оператора \mathcal{T} , то он имеет такие две неподвижные точки $x_*^- \leq x_*^+$, что любая траектория системы (3) притягивается к \mathcal{T} -инвариантному множеству $\Xi := \{x \in K_+ : x_*^- \leq x \leq x_*^+\}$.*

В р. 2.6 введены основные понятия, связанные с дифференцированием КАМН-операторов. Вектор $e \in \mathbb{R}^p$ назовем *допустимым* для выпуклого множества $C \subset \mathbb{R}^p$ в точке $x \in C$, если $x + \theta e \in C$ для всех $\theta \geq 0, \theta \approx 0$; совокупность всех таких e обозначим через $T_x(C)$. В диссертации показано, что для КАМН-оператора $\mathcal{T} : K_+^p \rightarrow K_+^p$ и $x_* \in K_+^p$, преобразование

$$\mathcal{T}'_{x_*} : e \in T_{x_*}(K_+^p) \mapsto \lim_{\theta \rightarrow +0} \theta^{-1} \{\mathcal{T}[x_* + \theta e] - \mathcal{T}[x_*]\}$$

является корректно определенным ОКАМН-оператором. Его сужение $\mathcal{T}'_{x_*}{}^+ := \mathcal{T}'_{x_*}|_{K_+}$ назовем *правым дифференциалом* оператора \mathcal{T} в точке x_* , а оператор $\mathcal{T}'_{x_*}{}^- : e \in K_+^p(x_*) := \{e \geq 0 : e_i = 0 \text{ если } x_{*,i} = 0\} \mapsto -\mathcal{T}'_{x_*}[-e]$ — *левым дифференциалом*. Неподвижную точку x_* назовем *устойчивой по первому приближению*, если устойчивое собственное число как правого, так и левого дифференциала в точке x_* меньше 1.

В р. 2.7 обоснована следующая теорема, которая в качестве источника Теоремы 2.1 демонстрирует ее родство с теоремой Ляпунова об устойчивости по первому приближению и может служить базой для других критериев глобальной устойчивости систем с КАМН динамическим оператором.

Теорема 2.5 *Если устойчивое собственное число однородного ядра КАМН-оператора $\mathcal{T} : K_+^p \rightarrow K_+^p$ меньше 1, то этот оператор имеет неподвижную точку. Если дополнительно известно, что любая неподвижная точка оператора \mathcal{T} устойчива по первому приближению, то неподвижная точка единственна и притягивает все траектории системы (3).*

Главы 3—7 иллюстрируют дееспособность развитой теории и предложенного метода исследованием потоковых переключательных сетей, представляющих самостоятельный интерес.

Третья глава изучает сеть, состоящую из нескольких непрерывно растущих очередей и одного сервера и являющуюся частным случаем сети из гл. 1. В каждый момент сервер способен обслуживать не более одной очереди; переключение между очередями занимает фиксированное время. Для такой сети известны простые необходимые условия оптимальности процесса: буфер обслуживается на максимальной доступной скорости, что при пустом буфере означает на скорости входящего потока (W.M. Lan, T.L. Olsen, J. van Eekelen и др.). Имеется ряд работ, в которых найдено оптимальное или субоптимальное решение (J.B. Kruskal, 1969; O.J. Voxma, H. Levy, J.A. Weststrate, 1991; B. Gaujal, A. Hordijk, D. van der Laan, 2007 и др.) при специальных предположениях либо о системе, либо о допустимых алгоритмах управления.

В главе 3 решена задача синтеза протокола управления общей системой указанного типа. Протокол обеспечивает сходимость всех процессов к априори заданному периодическому процессу π^0 , который удовлетворяет упомянутому необходимому условию оптимальности и в остальном произволен. Использовано разбиение (1) процесса π^0 на простейшие фазы, в течение которых дискретное состояние Q постоянно. Для активных фаз предложенное правило управления предписывает обслуживать буфер n на максимальной скорости μ_n пока отношение конечного и начального (для данной фазы) объема буфера не сравняется с отношением, наблюдаемым для π^0 . Затем буфер обслуживается на скорости входящего потока столько времени, сколько такое обслуживание производилось для π^0 .

Теорема 3.1 *Предложенный протокол управления порождает единственный периодический процесс, равный априори заданному π^0 , и обеспечивает притяжение к нему всех остальных процессов при $t \rightarrow \infty$.*

В главе 4 дано решение задачи синтеза протокола управления потоковой переключательной сетью с одним сервером и разделением его ресурсов. Рассматриваемая система состоит из N буферов и одного сервера; работа поступает в n -ый буфер с постоянной скоростью $\lambda_n > 0$ и после обработки сервером покидает систему. Сервер имеет конечное число рабочих режимов (мод) $t \in \mathcal{M}$. В моде t он одновременно обслуживает фиксированный набор $A_m \subset [1 : N]$ буферов, извлекая содержимое $x_n(t)$ буфера $n \in A_m$ со

скоростью $u_n(t)$, варьируемой в заданных границах $u_n(t) \in [0, \mu_{n|m}]$. Переключение между модами $m' \rightarrow m'' \in \mathcal{M}, m' \neq m''$ требует $\delta_{m' \rightarrow m''} > 0$ единиц времени. Множества A_m необязательно дизъюнкты; $\bigcup_{m \in \mathcal{M}} A_m = [1 : N]$. Скорости $u_n(t)$ обслуживания, а также момент отключения текущей моды и следующая за ней мода определяются протоколом управления.

Определение 4.1 *Процесс назовем интенсивным маргинальным, если в любой моде $q(t) \equiv m, t \in [t_-, t_+]$ каждый буфер $n \in A_m$ обслуживается на некотором начальном отрезке с максимальной скоростью $u_n(t) = \mu_{n|m}, t \in [t_-, t_*^n]$, а на некотором заключительном — с нулевой $u_n(t) = 0, t \in (t_\diamond^n, t_+]$ с возможным средним отрезком, где буфер пусть $x_n(t) = 0, t \in (t_*^n, t_\diamond^n]$, и каждый буфер $n \in A_m^\uparrow := \{n \in A_m : \mu_{n|m} \leq \lambda_n\}$ постоянно обслуживается на максимальной скорости $u_n(t) \equiv \mu_{n|m}$.*

Лемма 4.1 *Для любого периодического процесса $\pi = [X(\cdot), q(\cdot)]$ существует интенсивный маргинальный периодический процесс $\pi^\dagger = [X^\dagger(\cdot), q^\dagger(\cdot)]$ с лучшим качеством обслуживания $x_n^\dagger(t) \leq x_n(t) \forall t, n$, для которого любой буфер опустошается хотя бы раз за период.*

Лемма 4.1 дает основание ограничиться рассмотрением интенсивного маргинального периодического процесса $\pi^0 = [X^0(\cdot), q^0(\cdot)]$. В главе 4 предложен протокол управления, предписывающий серверу пробегать циклически повторяемую последовательность мод m_1, \dots, m_k , наблюдаемую в течение периода π^0 . В моде m_i буферы $n \in A_{m_i}^\uparrow$ обслуживаются на скорости $u_n \equiv \mu_{n|m_i}$, причем при $A_{m_i}^\downarrow := A_m \setminus A_{m_i}^\uparrow = \emptyset$ время обслуживания в моде m_i то же, что и для π^0 . Если $A_{m_i}^\downarrow \neq \emptyset$, каждый буфер $n \in A_{m_i}^\downarrow$ сначала обслуживается на скорости $\mu_{n|m_i}$ до тех пор, пока отношение его текущего и начального (для данной моды) объема не достигнет того же значения, что и для π^0 . Затем он последовательно обслуживается на скорости $u_n = \lambda_n$ и $u_n = 0$, причем длительности соответствующих интервалов времени те же, что и для π^0 . В момент завершения предыдущей задачи сервер записывает $\mathbf{end}(n|i) = 1$ и переходит к обслуживанию буфера на скорости $u_n = \lambda_n$, поддерживая его уровень неизменным. Мода завершается, как только $\mathbf{end}(n|i) = 1 \forall n \in A_{m_i}^\downarrow$, после чего $\mathbf{end}(n|i) := 0$ и сервер начинает переключение в следующую моду. Таким образом, обслуживание буфера опирается на данные только о нем самом, а "межбуферная" координация — на однобитовые сообщения $\mathbf{end}(n|i)$.

Теорема 4.1 *Предложенный протокол порождает единственный периодический процесс, этот процесс равен априори заданному π^0 и к нему сходятся все процессы в рассматриваемой системе.*

В пятой главе рассматривается популярная в качестве тестового примера производственная сеть Кумара-Сейдмана. Этот частный случай сети из главы 1 состоит из четырех буферов и двух серверов и обслуживает один

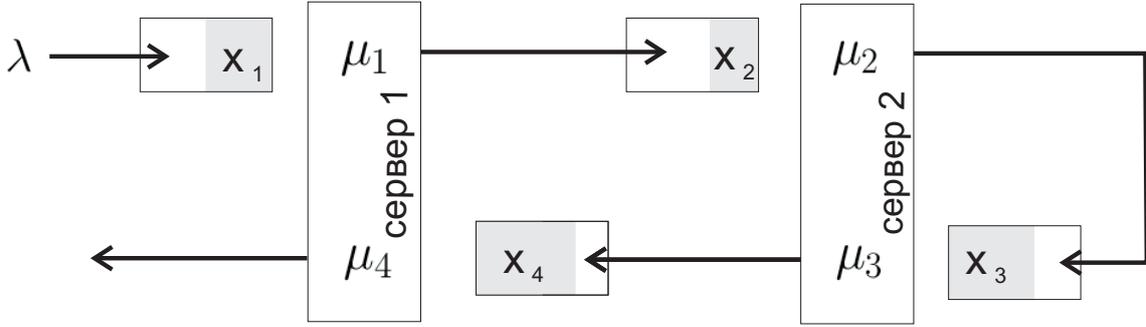


Рис. 1: Поточковая переключательная сеть Кумара-Сейдмана

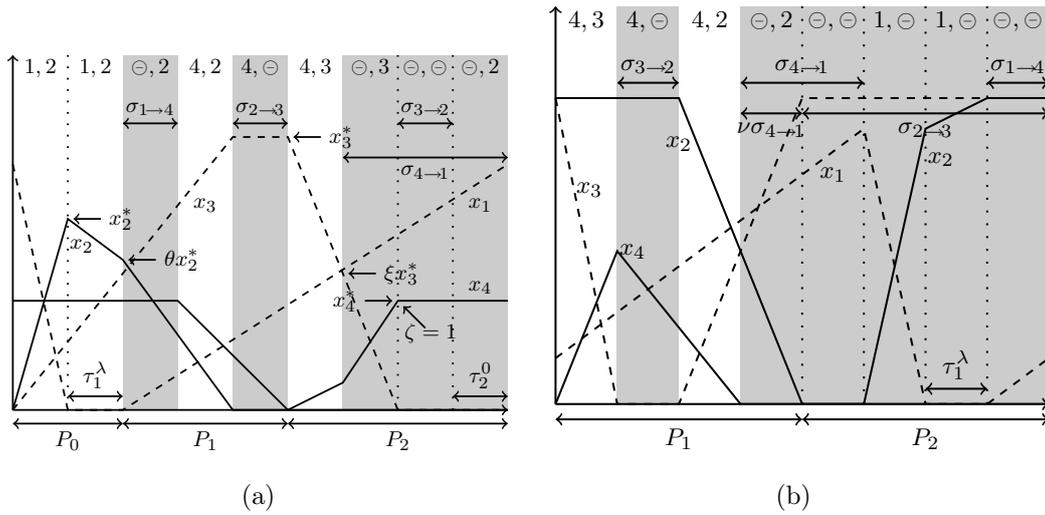


Рис. 2: Оптимальный периодический процесс.

поток, см. Рис. 1. Периодический процесс в системе назовем *простым*, если каждый сервер обслуживает каждый ассоциированный с ним буфер только один раз за минимальный период. Рассмотрим задачу минимизации среднего по времени количества работы в системе:

$$W := \overline{\lim}_{T \rightarrow \infty} \frac{1}{T} \int_0^T w(t) dt, \quad \text{где} \quad w(t) := \sum_{n=1}^4 c_n x_n(t). \quad (4)$$

Здесь $c_n > 0$ — заданный коэффициент, причем $c_1 \geq c_2 \geq c_3 \geq c_4 > 0$.

Теорема 5.1 *Оптимальный простой периодический процесс существует и имеет структуру, изображенную на одном из Рис. 2 (a), (b), причем буферы обслуживаются на максимальной возможной скорости. Вариант структуры зависит от параметров системы и коэффициентов c_n из (4).*

В диссертации найден критерий, позволяющий определить, какой из изображенных случаев имеет место для данных параметров и коэффициентов.

Решена задача синтеза протокола, асимптотически выводящего систему из любого начального состояния в оптимальный режим π^0 из Теоремы 5.1. Охарактеризуем протокол и соответствующий результат в случае (b).

Процесс делится на две фазы P_1 и P_2 в соответствии с Рис. 2(b); буфера обслуживаются на максимальной скорости. В течение фазы P_1 сервер 2 обслуживает буфер 3 до опустошения, затем переключается в буфер 2 и обслуживает его до опустошения, и, наконец, простаивает, ожидая, когда сервер 1 завершит свою миссию на этой фазе. Тем временем сервер 1 опустошает буфер 4, затем проходит первую часть переключения в буфер 1 в течение времени, которое эта часть длилась для π^0 , а затем возможно простаивает, ожидая пока сервер 2 опустошит буфер 2. Фаза завершается, как только буфер 2 опустошен, а сервер 1 завершил требуемую часть переключения $4 \rightarrow 1$. В фазе P_2 сервер 1 завершает переключение $4 \rightarrow 1$, затем опустошает буфер 1 и обслуживает его на скорости входящего потока в течение времени, которое такое обслуживание наблюдалось для π^0 , и возможно дольше для того, чтобы покинуть буфер 1 не раньше, чем пройдет $\sigma_{2 \rightarrow 3} - \sigma_{1 \rightarrow 4}$ единиц времени с начала фазы, и наконец переключается в буфер 4 ($\sigma_{i \rightarrow j}$ — время переключения из буфера i в j). Тем временем сервер 2 переключается из буфера 2 в 3, а затем простаивает, ожидая пока сервер 1 завершит свою миссию. Фаза завершается в момент окончания переключения $1 \rightarrow 4$.

Теорема 5.2 *В сети Кумара-Сейдмана, управляемой предложенным протоколом, все процессы сходятся к оптимальному периодическому процессу.*

Для случая (a) с Рис. 2 получен аналогичный результат.

В главе 6 метод и результаты, изложенные в главах 1 и 2, проиллюстрированы решением задачи синтеза протокола управления общей жидкостной моделью потоковой переключательной сети с произвольным числом буферов, серверов и внешних потоков, описанной в главе 1. Топология сети ограничена минимальными предположениями, приведенными на с. 6 автореферата.

Доказана теорема, показывающая, что при определенных условиях оптимальна максимально возможная скорость обслуживания буферов. При построении протокола рассматривается процесс π^0 с указанным свойством и применяется его деление (1) на простые фазы \mathfrak{P}_i , в течение которых дискретное состояние постоянно $Q \equiv Q_i$. Переключательная фаза $Q_i = (\ominus, \dots, \ominus)$ поддерживается столько времени, сколько она поддерживалась для π^0 . Для прочих фаз $Q_i = (n_1, \dots, n_s) \neq (\ominus, \dots, \ominus)$ каждый обслуживаемый на этой фазе буфер $n = n_\sigma \neq \ominus$ вначале обслуживается со скоростью μ_n до момента τ , когда его содержимое начинает уменьшаться, и далее до момента, когда отношение текущего объема буфера к его объему в момент "перегиба" τ не сравняется с отношением, наблюдавшимся на π^0 . После этого объем буфера поддерживается постоянным до конца фазы, но не менее времени, в течение которого он поддерживался постоянным для процесса π^0 . Фаза прекращается как только задание выполнено для всех обслуживаемых на фазе буферов.

Буфер, непосредственно принимающий поток из некоторого внешнего источника, назовем *входным*, а их множество обозначим через \mathbf{IN} .

Теорема 6.1 Пусть у периодического процесса π^0 с вышеуказанным свойством существует чисто переключательная фаза $Q_i = (\ominus, \dots, \ominus)$ и более того, такая фаза разделяет любую фазу, завершающуюся опустошением буфера $n \in \mathbf{IN}$, и любую последующую фазу, на которой этот буфер обслуживается. Тогда предложенный протокол управления порождает процесс π^0 и обеспечивает сходимость к нему всех процессов при $t \rightarrow \infty$.

Седьмая глава диссертации содержит доказательство Теоремы 6.1.

Восьмая глава диссертации посвящена анализу специального, но широко применяемого на практике класса протоколов управления производственными линиями, в которых реализован принцип обратной связи от текущего запроса (pull-type, pull-controlled). Единая математическая модель этого класса протоколов предложена К.К. Старковым и А.Ю. Погромским. В главе 8 впервые обоснована оптимальность этих протоколов.

Модель производственного процесса в системе с одним производственным сервером задана уравнением

$$y(k+1) = y(k) + u(k) + f(k), \quad k = 0, 1, \dots,$$

где $y(k) \in \mathbb{R}$ - суммарный выход сервера в момент k , $u(k) \in \mathbb{R}$ - управляющий сигнал и $f(k) \in \mathbb{R}$ - неизвестное внешнее возмущение, влияющее на скорость работы сервера. Целью управления является отслеживание неубывающего запроса $y_d(k) \in \mathbb{R}$ на продукцию, удовлетворяющего соотношению

$$y_d(k) = y_{d0} + v_d k + \varphi(k),$$

где $y_{d0} > 0$ — начальный запрос, v_d — средняя скорость приращения запроса и $\varphi(k) \in \mathbb{R}$ — ее ограниченное колебание. В каждый момент времени k управление u может принимать только два значения: 1 или 0 ("вкл" или "выкл"). Требуется минимизировать ошибку отслеживания запроса $\varepsilon(k) = y_d(k) - y(k)$ в классе всех стратегий управления, основанных на доступных данных:

$$u(k) = U_k[y(0), \dots, y(k), y_d(0), \dots, y_d(k)] \in \{0, 1\}.$$

Предполагаем, что $0 < \xi(k) := v_d + \Delta\varphi(k) - f(k) < 1 \forall k$, а также отсутствие данных, позволяющих уточнить значение $\xi(k) \in (0, 1)$.

В главе 8 исследованы следующие две оптимизационные задачи:

$$J_T = \sup_{\xi(0), \dots, \xi(T-1)} \sum_{k=0}^T |\varepsilon(k)|^p \rightarrow \min_U, \quad (5)$$

$$J_\infty = \overline{\lim}_{k \rightarrow \infty} \sup_{\xi(\cdot)} |\varepsilon(k)| \rightarrow \min_U. \quad (6)$$

Здесь $p \geq 1$ - заданная константа, \sup берется по всем $\xi(k) \in (0, 1)$ и $U = \{U_k(\cdot)\}$ - множество всех допустимых стратегий управления.

Теорема 8.2 *Стратегия управления $u(k) = \mathfrak{S}[\varepsilon(k)]$ оптимальна как для критерия (5) при любых T и $p \in [1, +\infty)$, так и для критерия (6). Здесь $\mathfrak{S}(\varepsilon) := 1$ при $\varepsilon > 0$, $\mathfrak{S}(\varepsilon) := 0$ при $\varepsilon < 0$, и $\mathfrak{S}(\varepsilon) := 0, 1$ при $\varepsilon = 0$.*

Публикации автора по теме диссертации.

Статьи в рецензируемых журналах и изданиях:

- [1] Феоктистова В.Н., Матвеев А.С. *Динамическая интерактивная стабилизация переключательной системы Кумара-Сейдмана.* // Вестник СПбГУ, 2009. Серия 1, вып. 3, стр. 86–97.
- [2] Феоктистова В.Н. *Динамическая интерактивная стабилизация потоковых систем с разделением процессора.* // Вестник СПбГУ, 2011. Серия 1, вып. 3, стр. 75–80.
- [3] K.K. Starkov, V. Feoktistova, A.Y. Pogromsky, A. Matveev, J.E. Rooda. *Performance analysis of a manufacturing line operated under optimal surplus-based production control.* // Mathematical Problems in Engineering, 1-26, 2012, doi: 10.1155/2012/602094, online.

Другие публикации:

- [4] V. Feoktistova, A. Matveev. *Generation of production cycles in multiple server systems with setup times: The case study.* / Proceedings of the 13th IFAC Symposium on Information Control Problems in Manufacturing, Moscow, Russia, 2009, pp. 568-573.
- [5] V. Feoktistova, A. Matveev, E. Lefebvre, J.E. Rooda. *Designs of optimal switching feedback decentralized control policies for re-entrant queueing networks: A case study.* / Proceedings of the 10th IFAC Workshop on Intelligent Manufacturing Systems, 2010, pp. 187-193.
- [6] V. Feoktistova, A. Matveev, E. Lefebvre, J.E. Rooda. *On Optimal Switching Interactive Decentralized Control of Networked Manufacturing Systems.* / Proceedings of the 18th IFAC World Congress, Milan, Italy, 2011, pp. 14048-14054.
- [7] K.K. Starkov, V. Feoktistova, A.Y. Pogromsky, A. Matveev, J.E. Rooda. *Optimal production control method for tandem manufacturing lines.* / Proceedings of the PHYSCON 2011, Leon, Spain, 2011.

Публикации [4,6] имеют признаки соответствия перечню рецензируемых журналов и изданий: они опубликованы в изданиях, которые охвачены системой цитирования Scopus и содержание которых включено в журнал *IFAC papers online* издательства Elsevier.