

УДК 519.248:656.71
ББК 22.18 39.513

ОПТИМИЗАЦИЯ УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ УЗЛОВОГО АЭРОПОРТА КАК СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С НЕСТАЦИОНАРНЫМИ ПОТОКАМИ И ЧАСТИЧНОЙ ВЗАИМОПОМОЩЬЮ КАНАЛОВ

Романенко В. А.¹

*(Самарский государственный аэрокосмический
университет им. академика С.П.Королёва
(национальный исследовательский университет), Самара)*

Сформулированы и решены задачи формирования оптимальной программы управления параметрами системы обслуживания перевозок узлового аэропорта как системы массового обслуживания (СМО) с нестационарными потоками и частичной взаимопомощью между обслуживающими каналами. Рассмотрены результаты решения задач, на базе которых сформулированы рекомендации в адрес производственных подразделений узловых аэропортов.

Ключевые слова: оптимальное управления, система массового обслуживания, узловой аэропорт.

1. Введение

С позиции теории систем узловой аэропорт (хаб) представляет собой сложную динамическую стохастическую систему, в которую поступают существенно нестационарные потоки воздушных судов (ВС), пассажиров и др. Выраженные циклические колебания интенсивности поступающих потоков обусловлены

¹ Владимир Алексеевич Романенко, докторант, кандидат технических наук, доцент (vla_rom@mail.ru).

волновым характером расписания хаба, предполагающим организацию т.н. «волн» массовых прибытий-отправлений рейсов, в пределах которых выделяют три временных интервала, первый из которых соответствует «волне» (комплексу) прибытия рейсов, второй – периоду трансфера пассажиров и их багажа на стыковочные рейсы, третий – «волне» (комплексу) отправления рейсов.

Важной задачей хаба является сокращение времени пересадки с рейса на рейс трансферных пассажиров. Для решения этой задачи расписание хаба строится таким образом, чтобы комплекс прибытия рейсов инициировали прилёты ВС большой пассажировместимости, обслуживание которых требует больших затрат времени, а завершали прилёты малых ВС. Поскольку среднее время обслуживания ВС в пределах интервала комплекса плавно уменьшается, нестационарным является не только входящий поток ВС хаба, но и поток их обслуживания.

Задачи сокращения времени трансфера пассажиров и длительности наземного простоя ВС в хабе определяют повышенные требования к численности и производительности средств обслуживания перевозок аэропортовых комплексов. Высокая насыщенность хаба обслуживающими средствами даёт возможность привлекать к выполнению технологических операций обслуживания отдельных ВС и их загрузки большее число перронных и терминальных средств по сравнению с «обычным», не узловым аэропортом. Так, например, использование двух средств вместо одного позволяет сократить время выполнения технологической операции, уменьшив тем самым как наземный простой ВС, так и время трансфера пассажиров. Недостаточное число обслуживающих средств приводит к вынужденным задержкам в обслуживании, образованию очередей из ожидающих обслуживания ВС, пассажиров и т.д. Длины таких очередей ограничены числом мест стоянки ВС, вместимостью пассажирского или грузового комплексов и другими параметрами.

Таким образом в терминах теории массового обслуживания система обслуживания перевозок хаба в рамках выбранной технологической операции представляет собой СМО с ожиданием в очереди ограниченной длины, нестационарными потоками

заявок и обслуживания, частичной взаимопомощью между обслуживающими каналами, выражающейся в возможности одновременного использования двух каналов для обслуживания одной заявки. Под заявками в СМО понимаются требующие обслуживания ВС, пассажиры, багаж и т.п.; под обслуживающими приборами – персонал аэропорта, перронные и терминальные технические средства.

Прежде, чем формулировать задачи оптимизации и формировать математическую модель рассматриваемой СМО, дадим её краткое описание, остановившись на особенностях потоков заявок и обслуживания, а также на обосновании возможных критериев оптимальности функционирования системы.

2. Описание системы наземного обслуживания перевозок как СМО

2.1. ОБЩЕЕ ОПИСАНИЕ СМО

Ниже в качестве заявок рассматриваются только прилетающие или вылетающие ВС, поскольку характеристики потоков пассажиров, багажа, грузов в значительной степени определяются характеристиками потоков прилетающих и вылетающих ВС.

Волновая система прилётов-вылетов хаба имеет, как правило, жесткую структуру с постоянным числом комплексов в течение суток, установленными интервалами между комплексами и фиксированной максимальной интенсивностью движения ВС. Примеры волновых систем трех аэропортов различного масштаба, работающих (или работавших ранее) по системе хаба, в виде графиков временных зависимостей интенсивностей потоков ВС $\lambda(t)$ приведены на рис.1. Поскольку структура комплексов рассмотренных хабов постоянна для каждого дня недельного расписания, и характер графиков интенсивности потоков ВС слабо зависит от дня недели, то на рис.1 отображены значения интенсивностей, полученные усреднением по дням недели.

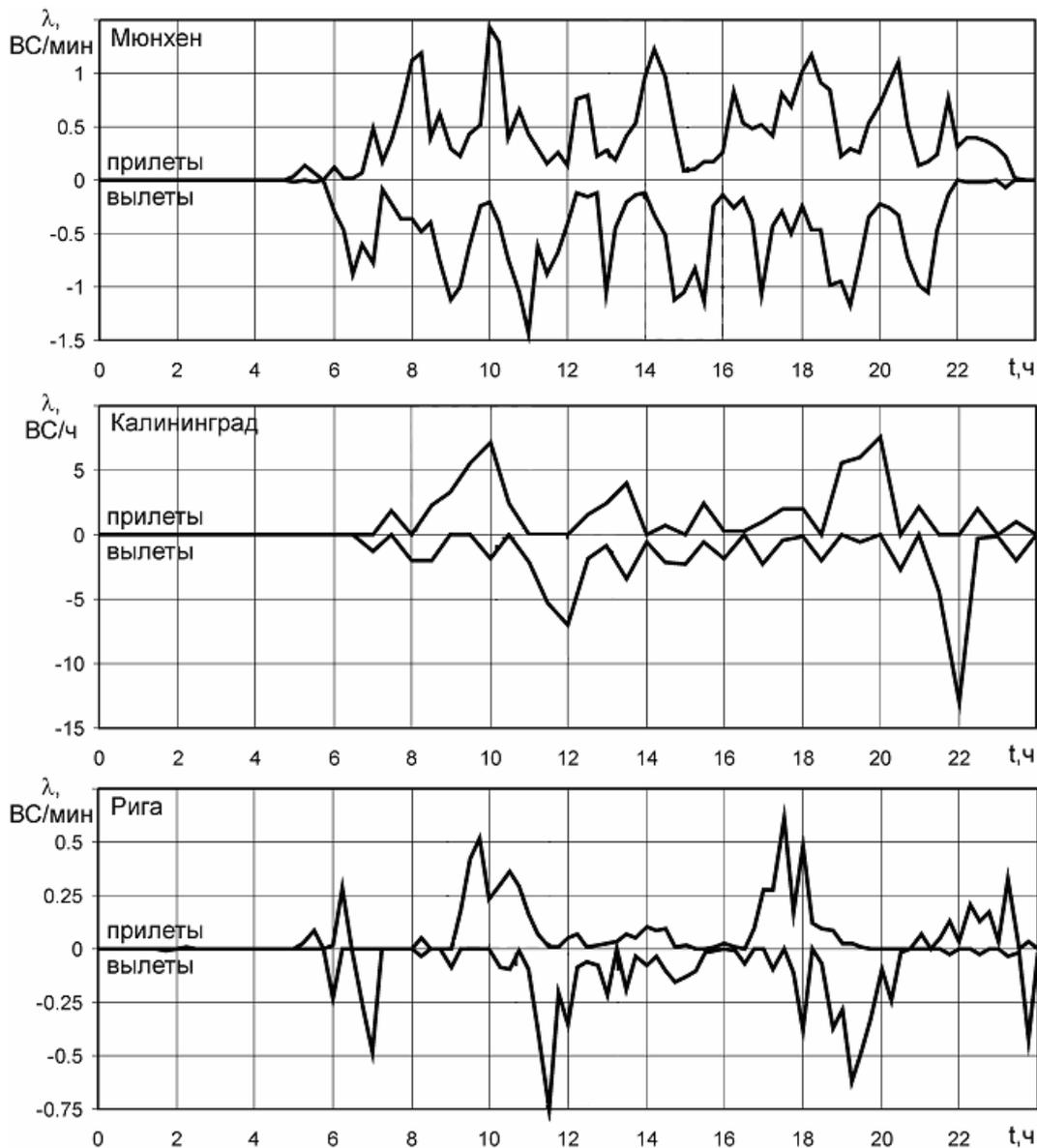


Рис. 1. Примеры волновых систем узловых аэропортов

Анализ волновых систем хабов позволяет сделать следующие выводы

– потоки ВС удовлетворяют требованиям ординарности и отсутствия последствия и могут рассматриваться в качестве нестационарных пуассоновских потоков [3];

– потоки ВС хаба характеризуется высокой неравномерностью. Пиковые значения интенсивности потока в зависимости от масштаба хаба колеблются от 5 ВС/ч. и менее до 1-1,5 ВС/мин. и более, при существенно меньших значениях между комплексами;

– длительность комплексов прилётов и вылетов ВС не превышает двух часов, в противном случае время ожидания пересадки может возрасти до неприемлемых для пассажиров величин;

– с уменьшением масштаба хаба сокращается число комплексов с одновременным ростом интервалов между ними до 10-12 ч.;

– графики интенсивности потоков ВС в течение комплексов прибытий и отправок имеют сложную форму, в целом близкую к треугольной.

Учитывая, что волна прилетов-вылетов ВС хаба включает два комплекса, то далее под термином «комплекс» понимается либо комплекс прилётов ВС (если обслуживание выполняется по прилёту), либо комплекс вылетов ВС (если обслуживание выполняется по отправлению). Предполагается, что комплексы циклически повторяются через равные интервалы времени продолжительностью T_B . Мгновенная интенсивность потока $\lambda(t)$ является кусочно-непрерывной функцией времени с периодом T_B , в течение которого выделяются три следующих друг за другом временных отрезка:

1) отрезок длительностью T_1 с интенсивностью потока заявок, возрастающей линейно от минимального значения λ_{min} до максимального значения λ_{max} ;

2) отрезок длительностью T_2 с интенсивностью, линейно убывающей от λ_{max} до λ_{min} ;

3) отрезок длительностью $T_3 = T_B - (T_1 + T_2)$ с постоянной интенсивностью потока заявок λ_{min} .

2.2. ОПИСАНИЕ ПРОЦЕССА ОБСЛУЖИВАНИЯ

Предполагается, что технологические нормы аэропорта допускают возможность использования как одного, так и одновременно двух средств (стоек регистрации, авиатопливозаправщиков, спецмашин и т.п.) для выполнения определённой технологической операции технического или коммерческого обслуживания ВС. На практике число используемых средств зависит от типа ВС, компоновки его салона, категории рейса или перевозки, загруженности аэропортовых комплексов и других

факторов. Виртуальные (мгновенные) продолжительности обслуживания заявки как одним каналом $t_{обсл\ 1}(t)$, так и двумя – $t_{обсл\ 2}(t)$ считаются показательно распределёнными случайными функциями времени. Мгновенные интенсивности обслуживания заявки одним каналом $\mu_1(t)$ и одновременно двумя каналами $\mu_2(t)$ определяются как

$$\mu_1(t) = [T_{обсл\ 1}(t)]^{-1}, \quad \mu_2(t) = [T_{обсл\ 2}(t)]^{-1},$$

где $T_{обсл\ 1}(t) = M[t_{обсл\ 1}(t)]$, $T_{обсл\ 2}(t) = M[t_{обсл\ 2}(t)]$ – среднее виртуальное время обслуживания заявки одним каналом и двумя каналами, соответственно.

Связь между величинами $\mu_1(t)$ и $\mu_2(t)$ задаётся соотношением

$$\mu_2(t) = \theta \mu_1(t),$$

где θ – коэффициент, учитывающий относительное увеличение интенсивности обслуживания при использовании двух каналов.

В условиях реального аэропорта связь между числом привлекаемых средств и интенсивностью обслуживания имеет довольно сложный характер, определяемый особенностями рассматриваемой операции обслуживания. Для операций, продолжительность которых связана с объёмом выполняемых работ (например, заправка ВС авиатопливом посредством авиатопливозаправщиков, посадка в ВС или высадка из ВС пассажиров и др.), зависимость интенсивности обслуживания от числа каналов приближается к прямо пропорциональной, не являясь, однако, строго таковой из-за наличия затрат времени на подготовительно-заключительные операции, на которые число средств не влияет. Для таких операций $\theta \leq 2$. Для ряда операций зависимость продолжительности выполнения от числа средств или исполнителей менее выражена (например, регистрация или предполётный досмотр пассажиров). В этом случае $\theta \approx 1$.

Поведение $T_{обсл\ 1}(t)$ и $T_{обсл\ 2}(t)$ как функций времени должно отражать наблюдаемые в течение комплекса изменения соотношения между численностью ВС, требующих более и менее длительного обслуживания. В связи с этим предполагается, что величина $T_{обсл\ 1}(t)$ (как и $T_{обсл\ 2}(t)$) является кусочно-непрерывной

функцией времени с периодом T_B , в течение которого последовательно выделяются четыре отрезка времени:

1) отрезок длительностью T_1' со значениями $T_{обсл\ 1}(t)$, линейно убывающими от некоторого среднесуточного значения $T_{обсл\ 1}^{mid}$ до минимального $T_{обсл\ 1}^{min}$.

2) отрезок длительностью T_2' с функцией $T_{обсл\ 1}(t)$, линейно возрастающей от $T_{обсл\ 1}^{min}$ до $T_{обсл\ 1}^{max}$;

3) отрезок длительностью T_3' с функцией $T_{обсл\ 1}(t)$, линейно убывающей от $T_{обсл\ 1}^{max}$ до исходного $T_{обсл\ 1}^{mid}$;

4) отрезок длительностью $T_4' = T_B - (T_1' + T_2' + T_3')$ с постоянным $T_{обсл\ 1}^{mid}$.

2.3. ФОРМИРОВАНИЕ СИСТЕМЫ КРИТЕРИЕВ ОПТИМАЛЬНОСТИ

Формирование совокупности критериев оптимальности рассматриваемой СМО, необходимых для решения задач оптимизации её параметров, производится с использованием несколько модифицированной модели эффективности сложной авиатранспортной системы, предложенной в работе [2]. Эффективность рассматривается как свойство системы выполнять заданные функции с заданными качеством и производительностью при минимальных затратах ресурсов. В формализованном виде модель записывается как

$$W(t) = (W_{\Pi}(t), W_{\mathcal{E}}(t), W_K(t)),$$

где $W(t)$ – обобщённый показатель эффективности системы; $W_{\Pi}(t)$ – комплексный показатель производительности системы; $W_{\mathcal{E}}(t)$ – комплексный показатель экономичности системы; $W_K(t)$ – комплексный показатель качества работы системы. Для рассматриваемой системы, потоки заявок и обслуживания в которой нестационарны, все перечисленные показатели зависят от времени.

Основное преимущество воздушной перевозки с точки зрения пользователей воздушного транспорта состоит в малых затратах времени на её выполнение. Важная роль в сокращении этих затрат принадлежит аэропорту. Аэропортовое предприятие,

способное обеспечить наземное обслуживание пассажиров, грузов и ВС за сравнительно короткое время, обладает серьезным преимуществом в конкурентной борьбе за привлечение клиентуры. Особое значение фактор затрат наземного времени имеет в случае хабов, для которых малая величина времени трансфера является одним из главных критериев их привлекательности с точки зрения пассажиров, грузоотправителей и авиаперевозчиков. Сокращение затрат наземного времени в хабе служит непосредственной причиной повышения числа отправок рейсов, уровня пассажиро- и грузопотоков, обеспечивающего рост доходов аэропорта от авиационной и неавиационной деятельности. Длительность пребывания в хабе должна быть небольшой для пассажира, ВС или груза, поступившего в хаб в любой, в том числе наиболее загруженный, момент времени.

Таким образом, в рамках целей исследования СМО предполагается, что производительность системы на выбранном этапе обслуживания непосредственно связана с затратами времени на пребывание заявки в СМО, возрастая с уменьшением этих затрат. В качестве показателя, выражающего уровень временных затрат, принято среднее виртуальное время пребывания заявки в СМО

$$T_{\text{преб}}(t) = M[t_{\text{преб}}(t)],$$

где $t_{\text{преб}}(t)$ – виртуальное время пребывания заявки в СМО, т.е. то время, в течение которого находилась бы в СМО гипотетическая заявка, вошедшая в СМО в момент времени t . Величина $t_{\text{преб}}(t)$ складывается из виртуального времени ожидания $t_{\text{ож}}(t)$ и виртуального времени обслуживания $t_{\text{обсл}}(t)$

$$t_{\text{преб}}(t) = t_{\text{обсл}}(t) + t_{\text{ож}}(t),$$

откуда

$$(1) \quad T_{\text{преб}}(t) = T_{\text{обсл}}(t) + T_{\text{ож}}(t),$$

где $T_{\text{обсл}}(t) = M[t_{\text{обсл}}(t)]$, $T_{\text{ож}}(t) = M[t_{\text{ож}}(t)]$ – средние виртуальные значения времени обслуживания и времени ожидания, соответственно.

С учётом введенных обозначений, связь между производительностью системы и затратами времени на пребывание в ней в общем виде выразится как

$$W_{\Pi}(t) = f_{\Pi} \left(\frac{1}{T_{\text{преб}}(t)} \right),$$

Следует ожидать, что на производительности рассматриваемой системы отразится перераспределение каналов между индивидуальным и групповым обслуживанием, приводящее к изменению соотношения между временем обслуживания и временем ожидания. Так, увеличение доли заявок, обслуживаемых парами каналов, вызовет в общем случае сокращение затрат времени на обслуживание и рост времени ожидания. Напротив, увеличение доли заявок, обслуживаемых индивидуальными каналами, приведёт к противоположному эффекту.

Под показателем экономичности системы понимается числовая характеристика расхода ресурсов на получение заданного конечного результата функционирования системы. Чтобы упростить задачу с целью обеспечения возможности её решения методами теории массового обслуживания, далее предполагается, что экономичность производственных комплексов хаба зависит только от их насыщенности техническими средствами обслуживания перевозок. Действительно, численность аэропортовых технических средств влияет на уровень расходов аэропорта по целому ряду основных статей [4], таких как расходы на эксплуатацию и техническое обслуживание, в том числе на персонал и материалы; капитальные затраты, в том числе износ и/или амортизацию и др. Таким образом, правомерна следующая запись, означающая, что увеличение числа обслуживающих каналов на выбранном этапе обслуживания отрицательно влияет на степень экономичности СМО

$$W_{\text{э}}(t) = f_{\text{э}} \left(\frac{1}{N(t)} \right),$$

где $N(t)$ – численность средств обслуживания на выбранном этапе, имеющаяся в распоряжении СМО в момент времени t . Зависимость от времени численности средств $N(t)$ объясняется спецификой расписания хаба, приводящей к временному сдвигу комплекса вылетов ВС относительно комплекса прилётов, и тем самым дающей возможность аэропортовым службам оперативно

маневрировать своими ресурсами, перераспределяя их между отправляемыми и прибывающими рейсами.

Показатель качества работы СМО должен учитывать два аспекта. Первый связан с необходимостью обеспечения заданного качества обслуживания клиентуры хаба – авиакомпаний, пассажиров, грузоотправителей и т.д.; второй – с качеством использования трудовых, материальных и финансовых ресурсов аэропортового предприятия. Поскольку очевидно, что качество обслуживания тем выше, чем меньше время тратится клиентом аэропорта на непроизводительное ожидание, то количественный показатель качества обслуживания заявки в модели СМО должен быть определённым образом связан с временем ожидания обслуживания. Учитывая стохастичность как потока заявок, так и процесса обслуживания, невозможно полностью исключить вероятность образования очереди ни при каком уровне производительности системы, однако длительность ожидания начала обслуживания для большинства заявок должна быть достаточно мала. Таким образом, для описания качества обслуживания заявки введены два параметра, а именно $t_{ожр}$ – расчётное время ожидания заявкой обслуживания в очереди и $P_{ожр}$ – вероятность превышения расчётного времени фактическим временем ожидания.

В рамках рассматриваемой модели СМО качество использования ресурсов выражает мгновенный коэффициент загрузки системы

$$(2) \quad K_3(t) = \frac{\bar{N}_3(t)}{N(t)},$$

где $\bar{N}_3(t) = M[N_3(t)]$ – среднее мгновенное число занятых обслуживанием каналов.

Перечисленные показатели качества обслуживания представляется целесообразным использовать в качестве ограничений при решении формулируемой задачи оптимизации. В этом случае, как следует из изложенного выше, «претендентами» на роль критериев оптимальности выступают минимальное число обслуживающих каналов и минимальные затраты времени на пребывание в системе. Одновременный учёт обоих противоре-

тивных критериев приводит к необходимости решения оптимизационной задачи в многокритериальной постановке. Выбор же единственного глобального критерия, включающего оба названных критерия, или их свёртка весьма затруднительны. В качестве такого единственного критерия могут выступать экономические результаты работы аэропорта (например, прибыль), влияние на которые оказывают как численность аэропортовых средств, так и затраты времени на пребывание в аэропорту. Однако, оценить такое влияние, особенно со стороны второго фактора, без существенного усложнения модели невозможно. Чтобы исключить указанные трудности, здесь использован подход, основанный на выделении из общей многокритериальной задачи двух однокритериальных задач и выборе для каждой из них главного критерия.

2.4. ОБЩЕЕ ОПИСАНИЕ СМО

С учетом изложенного выше, уточним описание оптимизируемой системы. Рассматривается СМО с ожиданием в очереди ограниченной длины, содержащая однотипные обслуживающие каналы, число которых $N(t)$ с течением времени может меняться. Число заявок, находящихся в СМО, не должно превышать постоянной величины K , что может быть обусловлено техническими ограничениями по числу обустроенных в аэропорту мест стоянки ВС, вместимости аэровокзального или грузового комплекса и т.п. Частичная взаимопомощь каналов выражается в том, что обслуживание заявки может производиться каналом как индивидуально, так и в составе группы из двух каналов. В случае наличия не менее двух свободных каналов поступившая заявка с вероятностью v_1 занимает для обслуживания один из них и – с вероятностью $v_2 = 1 - v_1$ – оба канала. Если же в момент поступления на обслуживание заявки СМО располагает только одним свободным каналом, то эта заявка в любом случае занимает имеющийся единственный канал. В случае отсутствия незанятых каналов вновь поступившая заявка «становится в очередь» и ожидает обслуживания. Если число заявок, находящихся в очереди, составляет $K - N(t)$, то вновь прибывшая заявка покидает СМО необслуженной. Вероятность такого события

должна быть малой. Поток заявок и обслуживания относятся к категории нестационарных пуассоновских.

3. Постановка задач параметрической оптимизации СМО

Для решения задач оптимизации параметров исследуемой СМО должна быть получена математическая модель, связывающая её входные переменные через переменные состояния с выходными переменными. Вектор входных переменных $\mathbf{x}(t)$ в модели рассматриваемой СМО включает характеристики входящего потока заявок ($\lambda(t)$) и процесса обслуживания ($T_{обсл\ 1}(t), T_{обсл\ 2}(t), v_1(t)$), а также численность каналов $N(t)$ и мест K в СМО:

$$\mathbf{x} = (\lambda(t), T_{обсл\ 1}(t), T_{обсл\ 2}(t), v_1(t), N(t), K).$$

Выделим из состава компонентов вектора $\mathbf{x}(t)$ переменные, значения которых можно задавать по своему усмотрению – управляемые переменные. К их числу отнесём $v_1(t)$ и $N(t)$. Вектор управляемых переменных обозначим как $\mathbf{u}(t) = (v_1(t), N(t))$, $\mathbf{u}(t) \in U$, где U – область допустимых управлений, определяемая естественными ограничениями, накладываемыми на $v_1(t)$ и $N(t)$. Величина $v_1(t)$ может задаваться в пределах $0 \leq v_1(t) \leq 1$. Значения $N(t)$ должны быть целыми. Кроме того, логично предположить, что, несмотря на возможность манёвра ресурсами, аэропорт не может неограниченно наращивать $N(t)$ для выполнения рассматриваемой операции. Пусть заданы минимальное N_{min} и максимальное N_{max} значения числа средств обслуживания, которое может быть задействовано аэропортом для данной операции, тогда $N(t)$ должно удовлетворять условию $N_{min} \leq N(t) \leq N_{max}$.

Остальные компоненты $\mathbf{x}(t)$ будем считать неуправляемыми независимыми переменными и объединим их в вектор $\mathbf{x}'(t) = (\lambda(t), T_{обсл\ 1}(t), T_{обсл\ 2}(t), K)$.

Компонентами вектора переменных состояния \mathbf{P} являются вероятности состояний СМО $P_i, i = 0, 1, \dots, M$, зависящие в общем случае от начального состояния СМО, её входных управляемых и неуправляемых переменных и времени. Таким обра-

зом, сам вектор переменных состояния \mathbf{P} является вектор-функцией и может быть определён как

$$\mathbf{P}(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t) = (P_0(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t), P_1(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t), \dots, P_M(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t)),$$

где $M+1$ – максимально возможное число состояний СМО; $\mathbf{P}^0 = (P_0^0, P_1^0, \dots, P_M^0)$ – вектор вероятностей состояний СМО в начальный момент времени $t = 0$.

Вероятности, составляющие вектор-функцию $\mathbf{P}(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t)$, определяются путём численного интегрирования системы дифференциальных уравнений Колмогорова, описанной ниже.

В состав компонентов вектора выходных переменных $\mathbf{z}(t)$ должны быть включены основные вероятностно-временные характеристики СМО, необходимые для решения формулируемых оптимизационных задач:

$$\mathbf{z} = (T_{обсл}(t), T_{ож}(t), P\{t_{ож}(t) < \tau\}, P\{t_{преб}(t) < \tau\}, \bar{N}_3(t), P_{отк}(t)),$$

где $P\{t_{ож}(t) < \tau\}$ – функция распределения времени ожидания (вероятность того, что виртуальное время ожидания заявкой обслуживания $t_{ож}(t)$ не превысит величины τ); $P\{t_{преб}(t) < \tau\}$ – функция распределения времени пребывания в СМО (вероятность того, что виртуальное время пребывания заявки в СМО $t_{преб}(t)$ не превысит величины τ); $P_{отк}(t)$ – вероятность отказа очередной заявке в обслуживании.

Связи перечисленных выходных переменных с переменными состояния представим в виде

$$\begin{aligned} T_{обсл} &= Z_{T_{обсл}}(\mathbf{P}(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t)), \\ T_{ож} &= Z_{T_{ож}}(\mathbf{P}(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t)), \\ (3) \quad P\{t_{ож}(t) < \tau\} &= Z_{P_{ож}}(\mathbf{P}(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t), \tau), \\ P\{t_{преб}(t) < \tau\} &= Z_{P_{преб}}(\mathbf{P}(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t), \tau), \\ \bar{N}_3 &= Z_{\bar{N}_3}(\mathbf{P}(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t)), \\ P_{отк} &= Z_{P_{отк}}(\mathbf{P}(\mathbf{P}^0, \mathbf{x}'(t), \mathbf{u}(t), t)), \end{aligned}$$

где $Z_{T_{обсл}}(\cdot)$, $Z_{T_{ож}}(\cdot)$, $Z_{P_{ож}}(\cdot)$, $Z_{P_{преб}}(\cdot)$, $Z_{\bar{N}_3}(\cdot)$, $Z_{P_{отк}}(\cdot)$ – функциональные зависимости соответствующих выходных переменных

от переменных состояния, которые определяются по приведённым в следующем пункте формулам. Формулы (1) и (3) позволяют определить виртуальное среднее время пребывания заявки в системе как

$$T_{\text{преб}}(t) = Z_{\text{Тобсл}} \left(P(P^0, x'(t), u(t), t) \right) + Z_{\text{Тож}} \left(P(P^0, x'(t), u(t), t) \right).$$

Сформулируем теперь две упомянутые выше задачи оптимизации с различными критериями, решения которых можно рассматривать как некоторые предельные случаи решения многокритериальной задачи, когда значимость одного из критериев полностью превалирует над значимостью другого. Первая задача, в которой в роли критерия оптимальности выступает минимум численности аэропортовых средств обслуживания, соответствует ситуации, когда аэропортовое предприятие своей приоритетной целью ставит обеспечение экономичности перевозок. Вторая задача, где критерием является минимальное время пребывания в аэропорту, соответствует ситуации, когда аэропортом в качестве более важной выбрана цель сокращения временных затрат на перевозку.

Задача минимизации числа каналов (Задача А).

Для выбранной операции обслуживания на заданном временном интервале определить управление $u(t)$, обеспечивающее минимум мгновенному потребному числу обслуживающих каналов $N(t)$ в СМО с заданными средними виртуальными продолжительностями индивидуального $T_{\text{обсл}1}(t)$ и группового $T_{\text{обсл}2}(t)$ обслуживания, в которую поступает нестационарный пуассоновский поток заявок заданной мгновенной интенсивности $\lambda(t)$:

$$u^{opt}(t) = \underset{u(t) \in U}{\operatorname{argmin}} N(t).$$

Временной интервал оптимизации определим, используя следующие соображения. Чтобы состояния СМО в течение некоторого комплекса не оказывали влияния на её состояния в течение последующего комплекса необходимо, чтобы интервал времени T_B между двумя соседними комплексами был достаточно продолжительным для формирования в СМО близкого к установившемуся режима. При выполнении этого условия оптимальная программа управления $u^{opt}(t)$, определённая для

интервала одного комплекса, будет в неизменном виде воспроизводиться в течение интервалов всех других комплексов. Таким образом, при введении допущения о значительной продолжительности интервалов между комплексами, задачу поиска оптимального управления достаточно будет решить только для одного интервала продолжительностью T_B , момент начала которого совпадает с моментом начала комплекса. К указанному моменту СМО уже должна пребывать в установившемся режиме, соответствующем постоянным λ_{min} , N_{min} , $T_{обсл}^{mid}$ и некоторому v_1 , например $v_1 = 0$.

Задача решается при следующих ограничениях.

1. Ограничение на длительность пребывания заявки в СМО, для задания которого вводятся два параметра: $t_{пребP}$ – заданное расчётное время пребывания заявки в системе, $P_{пребP}$ – заданная расчётная вероятность того, что фактическое виртуальное время пребывания заявки в системе $t_{преб}(t)$ не превысит расчётного времени пребывания $t_{пребP}$. С учётом введенных обозначений ограничение принимает вид:

$$(4) \quad P\{t_{преб}(t) < t_{пребP}\} = Z_{Преб} \left(P(P^0, x'(t), u(t), t), t_{пребP} \right) > P_{пребP}.$$

2. Ограничение на нижнюю границу уровня качества обслуживания заявок. При заданных расчётном времени ожидания заявкой обслуживания в очереди $t_{ожP}$ и вероятности превышения расчётного времени фактическим временем ожидания $P_{ожP}$ это ограничение формулируется как

$$(5) \quad P\{t_{ож}(t) < t_{ожP}\} = Z_{Ож} \left(P(P^0, x'(t), u(t), t), t_{ожP} \right) > P_{ожP}.$$

3. Ограничение на верхнюю границу вероятности отказов в обслуживании заявки.

$$(6) \quad P_{отк}(t) = Z_{Отк} \left(P(P^0, x'(t), u(t), t) \right) < P_{откP},$$

где $P_{откP}$ – расчётная вероятность отказа, которая с учётом специфики объекта оптимизации должна задаваться весьма малой.

4. Ограничения на значения управляемых переменных:

$$(7) \quad N(t) - \text{целое}, \quad N_{min} < N(t) \leq N_{max}, \quad 0 \leq v_1 \leq 1.$$

Отметим важную особенность задачи А. Целочисленность $N(t)$ может приводить к ситуации, когда оптимальному значе-

нию этой величины соответствует не единственное оптимальное значение v_1 , а множество значений, принадлежащее некоторому диапазону, все точки которого удовлетворяют ограничениям (4)-(7). В этом случае оптимальным решением целесообразно считать то значение v_1 из указанного диапазона, которое обеспечивает минимум величине среднего виртуального времени пребывания в системе $T_{преб}(t)$.

Задача минимизации времени пребывания в системе (Задача Б).

Для выбранной операции обслуживания на описанном выше временном интервале продолжительностью T_B определить управление $u(t)$ обеспечивающее минимум среднему виртуальному времени пребывания заявки $T_{преб}(t)$ в СМО с заданными средними продолжительностями индивидуального $T_{обсл1}(t)$ и группового $T_{обсл2}(t)$ обслуживания, в которую поступает нестационарный пуассоновский поток заявок заданной интенсивности $\lambda(t)$:

$$u^{opt}(t) = \operatorname{argmin}_{u(t) \in U} \left[Z_{\bar{T}_{обсл}} \left(P(P^0, x'(t), u(t), t) \right) + Z_{\bar{T}_{ож}} \left(P(P^0, x'(t), u(t), t) \right) \right].$$

Помимо используемых при решении задачи А ограничений (5)-(7), здесь учитывается ограничение на нижнюю границу уровня качества использования ресурсов, которое с учетом (2) записывается в виде

$$(8) \quad K_3(t) = \frac{\bar{N}_3(t)}{N(t)} = \frac{Z_{\bar{N}_3} \left(P(P^0, x'(t), u(t), t) \right)}{N(t)} > K_{3P},$$

где K_{3P} – расчётный коэффициент занятости системы, задаваемый таким образом, чтобы непроизводительные простои каналов обслуживания были достаточно малыми.

Поскольку решение обеих сформулированных задач при наличии динамической модели СМО, позволяющей определять её вероятностно-временные характеристики, не предполагает особых затруднений, нет смысла подробно описывать здесь оптимизационный алгоритм. Отметим лишь, что оптимальные параметры управления определяются на каждом шаге численного интегрирования системы уравнений Колмогорова, входящей в модель СМО. На каждом шаге с последовательным увеличе-

нием перебираются значения $N(t)$, для каждого из которых делается попытка определить значение $v_1(t)$, минимизирующее критерий и удовлетворяющее ограничениям. В случае успеха найденные таким образом параметры $N(t)$, $v_1(t)$ и считаются оптимальными. Если для очередного $N(t)$ не удаётся найти удовлетворяющее ограничениям значение $v_1(t)$, то число каналов $N(t)$ увеличивается на единицу и расчёт повторяется. При этом используется ряд простых эвристических приёмов, учитывающих близкий к линейному характер зависимости вероятностно-временных параметров СМО от $v_1(t)$.

4. Модель СМО с нестационарными потоками и частичной взаимопомощью каналов

4.1. ПОСТРОЕНИЕ МОДЕЛИ СМО

Рассматриваемая СМО в произвольный момент времени t может находиться в одном из $b + 1$ дискретных состояний – S_0, S_1, \dots, S_b . Переход из состояния в состояние осуществляется в произвольный момент времени. В любой момент времени t для вероятностей $P_i(t)$ ($i=0, 1, \dots, b$) нахождения СМО в состоянии S_i ($i=0, 1, \dots, b$) должно выполняться условие нормировки $\sum_{i=0}^b P_i(t) = 1$. Знание вероятностей $P_0(t), P_1(t), \dots, P_b(t)$ позволяет определять виртуальные вероятностно-временные характеристики СМО, необходимые для решения задач её параметрической оптимизации.

Вероятности состояний $P_i(t)$ находятся путём решения системы дифференциальных уравнений Колмогорова, в общем виде записываемой как [1]

$$(9) \quad \frac{dP_i(t)}{dt} = \sum_{j=1}^b \varphi_{ji}(t)P_j(t) - P_i(t) \sum_{j=1}^b \varphi_{ij}(t), \quad i = 0, 1, \dots, b,$$

где $\varphi_{ij}(t)$ – плотности (интенсивности) вероятностей перехода из состояния с порядковым номером i в состояние с порядковым номером j . Величины $\varphi_{ij}(t)$ определяются по формуле

$$\varphi_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(t; \Delta t)}{\Delta t},$$

где $P_{ij}(t; \Delta t)$ – вероятность того, что СМО, пребывавшая в момент t в состоянии S_i , за время Δt перейдет из него в состояние S_j ($i \neq j$).

Чтобы упростить составление уравнений Колмогорова, используют размеченный граф состояний СМО, в котором над стрелками, ведущими из S_i в S_j , проставляют соответствующие интенсивности φ_{ij} . Производная вероятности каждого состояния определяется как сумма всех потоков вероятности, идущих из других состояний в данное состояние, за вычетом суммы всех потоков вероятности, идущих из данного состояния в другие.

Начальные условия интегрирования системы уравнений Колмогорова, если в начальный момент времени $t = 0$ в СМО нет заявок, имеют вид:

$$(10) P_0(0) = 1, \quad P_1(0) = P_2(0) = \dots = P_b(0) = 0.$$

Запись в общем виде (9) уравнений Колмогорова с граничными условиями (10) для рассматриваемой СМО, отличающейся переменным $N(t)$, программно труднореализуема. В соответствии с алгоритмическим подходом [1] рассмотрим методику формирования системы уравнений Колмогорова в виде, пригодном для компьютерных вычислений. Используем двойную систему обозначений состояний СМО, в которой i – число занятых обслуживанием каналов плюс длина очереди, j – число заявок в СМО.

Реализовано может быть не любое состояние из формальной совокупности S_{ij} ($i = 0, 1, \dots, K; j = 0, 1, \dots, K$). В частности, в рамках описываемой модели невозможны состояния, при которых две или более заявок одновременно обслуживаются одним каналом, т.е. $P_{ij}(t) = 0$, если $j > i$. Обозначим символом \mathcal{S} множество допустимых состояний СМО. Состояние S_{ij} существует и соответствующая ему вероятность $P_{ij}(t)$ может быть ненулевой, если выполняется одно из условий:

$$\left. \begin{array}{l} 1) j \leq i \leq 2j, \\ 2) j \leq i \leq j+q-1, \\ 3) j \leq i \leq K, \end{array} \right\} \begin{array}{l} \text{если } 2j \leq N, \\ \text{если } \begin{cases} 2j > N, \\ j+q-1 < K, \end{cases} \\ \text{если } j+q-1 \geq K, \end{array} \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \begin{array}{l} i = 0, 1, \dots, K, \\ j = 0, 1, \dots, K, \end{array}$$

где q – максимальное число состояний с различным количеством обслуживающих каналов для заданного числа заявок, определяемое по формуле

$$q = N - \lceil N/2 \rceil.$$

Здесь скобки обозначают операцию отбрасывания дробной части. В формулах данного пункта обозначение функциональной зависимости числа каналов от времени опущено с целью экономии места.

Для реализации компьютерных вычислений с использованием системы уравнений Колмогорова с переменным N её уравнения должны быть приведены к некоторой универсальной форме, допускающей запись любого уравнения.

С целью выработки такой формы рассмотрим фрагмент графа состояний, отображающий одно произвольное состояние S_{ij} с ведущими из него стрелками интенсивностей. Обозначим римскими цифрами соседние состояния, непосредственно связанные с S_{ij} , так, как это показано на рис.2.

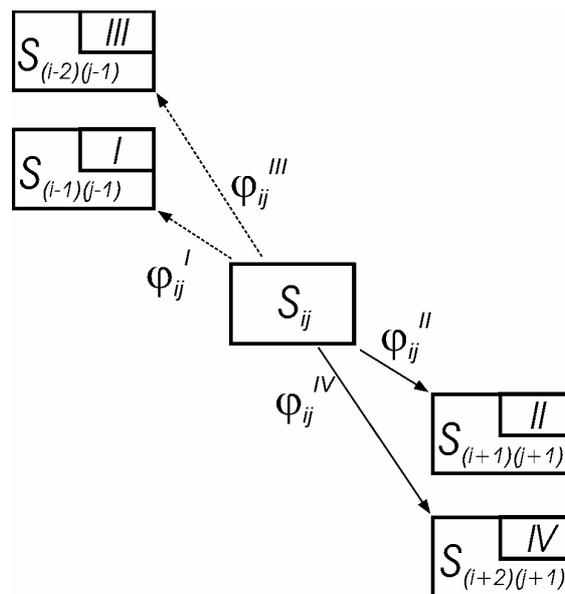


Рис. 2. Фрагмент графа состояний СМО

С учётом наличия соседних по отношению к S_{ij} состояний, уравнение, описывающее S_{ij} , запишется следующим образом

$$(11) \quad \frac{dP_{ij}(t)}{dt} = -[\varphi_{ij}^I(t) + \varphi_{ij}^{II}(t) + \varphi_{ij}^{III}(t) + \varphi_{ij}^{IV}(t)]P_{ij}(t) + \\ + \varphi_{(i+1)(j+1)}^I(t)P_{(i+1)(j+1)}(t) + \varphi_{(i-1)(j-1)}^{II}(t)P_{(i-1)(j-1)}(t) + \\ + \varphi_{(i+2)(j+1)}^{III}(t)P_{(i+2)(j+1)}(t) + \varphi_{(i-2)(j-1)}^{IV}(t)P_{(i-2)(j-1)}(t).$$

Для каждого состояния S_{ij} ($i = 0, 1, \dots, K; j = 0, 1, \dots, K$), такого, что $S_{ij} \in \mathcal{S}$, в момент времени t величины $\varphi_{ij}^I(t), \varphi_{ij}^{II}(t), \varphi_{ij}^{III}(t), \varphi_{ij}^{IV}(t)$ принимают различные значения (в том числе равные нулю). Однако при этом структура уравнения (11) сохраняется неизменной, что позволяет использовать его для компьютерной реализации системы уравнений Колмогорова произвольной размерности.

Интенсивности $\varphi_{ij}^{II}(t), \varphi_{ij}^{IV}(t)$, стремящиеся перевести СМО в состояния с большими значениями i и j , если наличие таких состояний является возможным, определяются исходя из ряда условий следующим образом:

$$\varphi_{ij}^{II}(t) = \left\{ \begin{array}{l} 0, \quad \text{если } S_{ij} \notin \mathcal{S} \text{ или } S_{(i+1)(j+1)} \notin \mathcal{S}, \\ v_1 \lambda(t), \text{ если } \left\{ \begin{array}{l} S_{ij} \in \mathcal{S}, \\ S_{(i+1)(j+1)} \in \mathcal{S}, \\ S_{(i+2)(j+1)} \in \mathcal{S}, \\ i \leq N-2, \end{array} \right. \\ 1, \text{ если } \left\{ \begin{array}{l} S_{ij} \in \mathcal{S}, \\ S_{(i+1)(j+1)} \in \mathcal{S}, \text{ или } \\ S_{(i+2)(j+1)} \notin \mathcal{S}, \end{array} \right. \left\{ \begin{array}{l} S_{ij} \in \mathcal{S}, \\ S_{(i+1)(j+1)} \in \mathcal{S}, \\ i > N-A, \end{array} \right. \end{array} \right. \left. \begin{array}{l} i = 0, 1, \dots, K, \\ j = 0, 1, \dots, K, \end{array} \right.$$

$$\varphi_{ij}^{IV}(t) = \left\{ \begin{array}{l} 0, \quad \text{если } S_{ij} \notin \mathcal{S} \text{ или} \\ \quad \quad \quad S_{(i+1)(j+1)} \notin \mathcal{S} \text{ или } i > N - 2, \\ v_2 \lambda(t), \quad \text{если } \left\{ \begin{array}{l} S_{ij} \in \mathcal{S}, \\ S_{(i+1)(j+1)} \in \mathcal{S}, \\ S_{(i+2)(j+1)} \in \mathcal{S}, \\ i \leq N - 2, \end{array} \right. \\ 1, \quad \text{если } \left\{ \begin{array}{l} S_{ij} \in \mathcal{S}, \\ S_{(i+1)(j+1)} \notin \mathcal{S}, \\ S_{(i+2)(j+1)} \in \mathcal{S}, \\ i \leq N - 2, \end{array} \right. \end{array} \right. \left. \begin{array}{l} i = 0, 1, \dots, K, \\ j = 0, 1, \dots, K. \end{array} \right.$$

Интенсивности $\varphi_{ij}^I(t)$, $\varphi_{ij}^{III}(t)$, переводящие СМО из состояния S_{ij} в состояния с меньшими значениями i и j (если наличие таких состояний возможно), прямо пропорциональны задействованному числу каналов, обслуживающих находящиеся в СМО заявки различного типа (занимающие для обслуживания один или два канала). Группу из двух каналов, занятых обслуживанием одной заявки соответствующего типа, можно рассматривать в качестве одного канала. Поэтому в общем случае:

$$\varphi_{ij}^I(t) = k_{ij1} \mu_1(t), \quad \varphi_{ij}^{III}(t) = k_{ij2} \mu_2(t),$$

где k_{ij1} – число заявок, занимающих один канал, обслуживаемых СМО в состоянии S_{ij} ; k_{ij2} – число заявок, занимающих по два канала, обслуживаемых СМО в состоянии S_{ij} .

Через i и j указанные величины определяются следующим образом:

$$k_{ij1} = \begin{cases} 2j - i, & \text{если } i \leq N, \\ N - 2(i - j), & \text{если } i > N, \end{cases}$$

$$k_{ij2} = i - j.$$

С учетом ограничений по возможности существования состояний выражения для $\varphi_{ij}^I(t)$, $\varphi_{ij}^{III}(t)$ имеют вид

$$\varphi_{ij}^I(t) = \begin{cases} 0, & \text{если } S_{ij} \notin \mathcal{S} \text{ или } S_{(i-1)(j-1)} \notin \mathcal{S}, \\ k_{ij1}\mu_1(t), & \text{если } \begin{cases} S_{ij} \in \mathcal{S}, \\ S_{(i-1)(j-1)} \in \mathcal{S}, \end{cases} \end{cases}$$

$$\varphi_{ij}^{III}(t) = \begin{cases} 0, & \text{если } S_{ij} \notin \mathcal{S} \text{ или } S_{(i-1)(j-1)} \notin \mathcal{S}, \\ k_{ij2}\mu_2(t), & \text{если } \begin{cases} S_{ij} \in \mathcal{S}, \\ S_{(i-1)(j-1)} \in \mathcal{S}, \end{cases} \end{cases}$$

$$i = 0, 1, \dots, K, \quad j = 0, 1, \dots, K.$$

4.2. ПОКАЗАТЕЛИ ЭФФЕКТИВНОСТИ ФУНКЦИОНИРОВАНИЯ СМО

Описанная модель позволяет определить временные зависимости следующих показателей эффективности функционирования рассматриваемой СМО.

По очевидным формулам определяются **среднее мгновенное число занятых каналов** как

$$\bar{N}_z(t) = \sum_{i=0}^{N(t)} \sum_{j=0}^K iP_{ij}(t),$$

и виртуальная вероятность отказа в обслуживании

$$P_{отк}(t) = \sum_{j=K-q+1}^K P_{Kj}(t).$$

Может быть получено **распределение виртуального времени ожидания заявкой обслуживания** $W(\tau, t) = P\{t_{ож}(t) < \tau\}$, позволяющее характеризовать качество обслуживания рассматриваемой СМО. Поступившая в систему заявка вынуждена ожидать обслуживания в случае, если все каналы заняты обслуживанием заявок, поступивших ранее. Существует вероятность $P_{w=0}(t)$ немедленного обслуживания поступившей заявки при наличии свободного канала (или нескольких свободных каналов)

$$P_{w=0}(t) = \sum_{i=0}^{N-1} \sum_{j=0}^{K-1} P_{ij}(t), \quad P_{ij}(t) \equiv 0, \text{ если } S_{ij} \notin \mathcal{S}.$$

С учётом возможности отказа искомая величина функции распределения $W(\tau, t)$ определится как

$$W(\tau, t) = \frac{1}{1 - P_{отк}(t)} \left(P_{w=0}(t) + \sum_{i=N}^{K-1} \sum_{j=0}^{K-1} W(\tau, t|(i, j)) P_{ij}(t) \right),$$

$$P_{ij}(t) \equiv 0, \text{ если } S_{ij} \notin \mathcal{S}.$$

Здесь $W(\tau, t|(i, j))$ – условная функция распределения времени ожидания некоторой заявки при условии, что в момент своего поступления t она застала СМО в состоянии ij .

В рассматриваемой СМО длительность ожидания обслуживания входящей заявкой зависит не только от числа заявок, уже находящихся в СМО, но и от распределения каналов между групповым и индивидуальным обслуживанием имеющихся заявок. Если бы взаимопомощи между каналами не существовало, то рассматриваемая СМО представляла бы собой традиционную СМО с ожиданием в очереди ограниченной длины, для которой общее время ожидания начала обслуживания заявкой, заставшей в момент поступления t других заявок в очереди, имело бы распределение Эрланга $E_{m+1}^{N\mu(t)}(x)$ [1]. Здесь верхний индекс содержит интенсивность обслуживания заявок всеми N каналами, действующими при наличии очереди; нижний индекс – порядок распределения по закону Эрланга. В рассматриваемой здесь СМО описанный закон справедлив только в отношении заявок, вошедших в СМО в состояниях, когда заняты все каналы, причём все они обслуживают по одной заявке. Для этих состояний можно записать

$$(12) \quad W(\tau, t|(N + m, N + m)) = E_{m+1}^{N\mu(t)}(\tau).$$

Обозначим как $E_{\alpha_1, \alpha_2, \dots, \alpha_n}^{y_1, y_2, \dots, y_n}(x)$ функцию распределения обобщённого закона Эрланга, имеющего порядок $\sum_{i=1}^n \alpha_i - 1$, где α_i – число случайных величин, распределённых по показательному закону с параметром y_i . Запишем выражения для функции распределения времени ожидания в других состояниях. С этой целью введём, вместо двух-, трёхиндексную систему обозначений, в которой состояние СМО в произвольный момент времени характеризуется тремя параметрами: числом занятых каналов n ($n = 0, 1, \dots, N$), числом заявок, обслуживаемых k ($k = 0, 1, \dots, K$)

и ожидающих обслуживания m ($m = 0, 1, \dots, K - N$). Связи между системами обозначений выражают зависимости:

$$i = n + m, \quad j = k + m;$$

$$\begin{cases} n = i, m = 0, k = j, & \text{если } i \leq N \\ n = N, m = i - j, k = j - i + N, & \text{если } i > N. \end{cases}$$

Описываемую формулой (12) условную функцию распределения времени ожидания в состоянии, когда заняты все каналы, обслуживающие по одной заявке, и при этом m требований находятся в очереди, с использованием трёх символов можно обозначить как $W(\tau, t|(n, k, m)) = W(\tau, t|(N, N, m)) = E_{m+1}^{N\mu_1(t)}(\tau)$.

Далее, в качестве примера, приведём выражения для функции распределения времени ожидания в трёх первых состояниях полной занятости каналов:

– к моменту входа заявки очереди нет, однако все каналы заняты:

$$W(\tau, t|(n, k, m)) = W(\tau, t|(N, N - g, 0)) =$$

$$= E_1^{H(N), (N-g)}(\tau), \quad g = 0, 1, \dots, q,$$

где $H_{ij} = k_{ij1}\mu_1(t) + k_{ij2}\mu_2(t)$;

– к моменту входа заявки в очереди находится одна заявка:

$$W(\tau, t|(n, k, m)) = W(\tau, t|(N, N - g, 1)) =$$

$$= \frac{k_{ij2}\mu_2(t)}{H(N), (N-g)} E_1^{H(N), (N-g)}(\tau) + \frac{k_{ij1}\mu_1(t)}{H(N), (N-g)} E_2^{H(N), (N-g)}(\tau);$$

– к моменту входа заявки в очереди находятся две заявки:

$$W(\tau, t|(n, k, m)) = W(\tau, t|(N, N - g, 2)) =$$

$$= \frac{k_{ij2}\mu_2(t)}{H(N), (N-g)} E_{1,1}^{H(N), (N-g), H(N), (N-g+1)}(\tau) +$$

$$+ \frac{k_{ij1}\mu_1(t)}{H(N), (N-g)} \left[\frac{k_{ij2}\mu_2(t)}{H(N), (N-g)} E_2^{H(N), (N-g)}(\tau) + \right.$$

$$\left. + \frac{k_{ij1}\mu_1(t)}{H(N), (N-g)} E_3^{H(N), (N-g)}(\tau) \right].$$

По аналогии записываются выражения и для последующих состояний. С ростом значения параметра t вид выражений усложняется, что не мешает их программной реализации.

Среднее виртуальное время ожидания заявки $T_{ож}(t)$ определяется численно как

$$T_{ож}(t) = \int_0^{\infty} \tau dW(\tau, t).$$

Распределение виртуального времени обслуживания произвольно выбранной заявки $t_{обсл}(t)$ определяется следующим образом. Поскольку изменение $t_{обсл}(t)$ в рассматриваемой СМО является случайным процессом, представляющим собой смесь двух показательно распределённых случайных процессов $t_{обсл1}(t)$ и $t_{обсл2}(t)$, то распределение $V(\tau, t) = P\{t_{обсл}(t) < \tau\}$ не будет показательным. С учётом возможности отказа выражение для функции распределения $V(\tau, t)$ запишется в виде

$$V(\tau, t) = \frac{1}{1 - P_{отк}(t)} \left(\sum_{i=N}^{K-1} \sum_{j=0}^{K-1} V(\tau, t|(i, j)) P_{ij}(t) \right),$$

$$P_{ij}(t) \equiv 0, \text{ если } S_{ij} \notin \mathcal{S}.$$

Здесь $V(\tau, t|(i, j))$ – условная функция распределения времени обслуживания некоторой заявки при условии, что в момент своего поступления она застала СМО в состоянии ij .

Если в момент начала обслуживания заявки СМО находится в состоянии, при котором возможно как групповое, так и индивидуальное обслуживание, то время обслуживания представляет собой смесь двух продолжительностей $t_{обсл1}(t)$ и $t_{обсл2}(t)$, распределённых показательно с параметрами $\mu_1(t)$ и $\mu_2(t)$, соответственно. Если же в этот момент нет возможности выделения двух каналов, то время обслуживания заявки распределено показательно с параметром $\mu_1(t)$. При подходе заявки к обслуживающим каналам в состоянии S_{ij} , переход к индивидуальному обслуживанию допустим при наличии возможности состояния $S_{(i+1)(j+1)}$, переход к групповому обслуживанию – при наличии возможности состояния $S_{(i+2)(j+1)}$ (рис.2). Таким образом, имеем:

$$V(\tau, t|(i, j)) = \begin{cases} 1 - e^{-\mu_1(t)\tau}, & \text{если } \begin{cases} S_{(i+1)(j+1)} \in \mathcal{S}, \\ S_{(i+2)(j+1)} \notin \mathcal{S}, \end{cases} \\ v_1(1 - e^{-\mu_1(t)\tau}) + \\ + v_2(1 - e^{-\mu_2(t)\tau}), & \text{если } \begin{cases} S_{(i+1)(j+1)} \in \mathcal{S}, \\ S_{(i+2)(j+1)} \in \mathcal{S}, \end{cases} \end{cases} \begin{cases} i = 0, 1, \dots, N-1, \\ j = 0, 1, \dots, N-1. \end{cases}$$

Поскольку, при отсутствии двух свободных каналов, любая заявка обслуживается одним каналом, то фактическая вероятность $v_1^\Phi(t)$ выделения одного канала будет больше заданной $v_1(t)$. Функция $v_1^\Phi(t)$ определяется как

$$v_1^\Phi(t) = \frac{1}{1 - P_{омк}(t)} \left(\sum_{i=N}^{K-1} \sum_{j=0}^{K-1} v_1(i, j) P_{ij}(t) \right),$$

$$P_{ij}(t) \equiv 0, \text{ если } S_{ij} \notin \mathcal{S}.$$

Здесь $v_1(i, j)$ – вероятность выделения одного аппарата для обслуживания заявки, поступившей в СМО в состоянии ij :

$$v_1(i, j) = \begin{cases} 1, & S_{(i+1)(j+1)} \in \mathcal{S}, S_{(i+2)(j+1)} \notin \mathcal{S}, \\ v_1, & S_{(i+1)(j+1)} \in \mathcal{S}, S_{(i+2)(j+1)} \in \mathcal{S}, \end{cases} \begin{cases} i = 0, 1, \dots, K-1, \\ j = 0, 1, \dots, K-1. \end{cases}$$

Средняя виртуальная продолжительность обслуживания заявки, вошедшей в СМО в момент t , может быть через $v_1^\Phi(t)$ определена как

$$T_{обсл}(t) = v_1^\Phi(t) T_{обсл1}(t) + [1 - v_1^\Phi(t)] T_{обсл2}(t).$$

Распределение виртуального времени пребывания заявки в СМО $U(\tau, t) = P\{t_{преб}(t) < \tau\}$ определяется с использованием полученных ранее выражений для функций распределения времени ожидания и времени обслуживания как

$$U(\tau, t) = \frac{1}{1 - P_{омк}(t)} \left(\sum_{i=N}^{K-1} \sum_{j=0}^{K-1} U(\tau, t|(i, j)) P_{ij}(t) \right),$$

$$P_{ij}(t) \equiv 0, \text{ если } S_{ij} \notin \mathcal{S}.$$

Здесь $U(\tau, t|(i, j))$ – условная функция распределения времени пребывания в СМО некоторой заявки при условии, что в момент своего поступления t она застало систему в состоянии ij .

Для состояний со свободными каналами время пребывания в СМО совпадает со временем обслуживания:

$$U(\tau, t|(i, j)) = \begin{cases} 1 - e^{-\mu_1(t)\tau}, & \text{если } \begin{cases} S_{(i+1)(j+1)} \in \mathcal{S}, \\ S_{(i+2)(j+1)} \notin \mathcal{S}, \end{cases} \\ \nu_1(1 - e^{-\mu_1(t)\tau}) + \\ + \nu_2(1 - e^{-\mu_2(t)\tau}), & \text{если } \begin{cases} S_{(i+1)(j+1)} \in \mathcal{S}, \\ S_{(i+2)(j+1)} \in \mathcal{S}, \end{cases} \end{cases} \quad \begin{cases} i = 0, 1, \dots, N-1, \\ j = 0, 1, \dots, N-1. \end{cases}$$

Для других состояний формулы условной функции распределения записываются по аналогии с формулами для $W(\tau, t|(n, k, m))$ с использованием трёхсимвольной индексации. Ниже они приведены для трёх первых состояний полной занятости каналов:

– к моменту входа заявки очереди нет, однако все каналы заняты:

$$U(\tau, t|(n, k, m)) = U(\tau, t|(N, N-g, 0)) = E_{1,1}^{\mu_1(t), H_{(N), (N-g)}}(\tau), \quad g = 0, 1, \dots, q;$$

– к моменту входа заявки в очереди находится одна заявка:

$$U(\tau, t|(n, k, m)) = U(\tau, t|(N, N-g, 1)) = \frac{k_{ij2}\mu_2(t)}{H_{(N), (N-g)}} E_{1,1}^{\mu_1(t), H_{(N), (N-g)}}(\tau) + \frac{k_{ij1}\mu_1(t)}{H_{(N), (N-g)}} E_{1,2}^{\mu_1(t), H_{(N), (N-g)}}(\tau);$$

– к моменту входа заявки в очереди находятся две заявки:

$$U(\tau, t|(n, k, m)) = U(\tau, t|(N, N-g, 2)) = \frac{k_{ij2}\mu_2(t)}{H_{(N), (N-g)}} E_{1,1,1}^{\mu_1(t), H_{(N), (N-g)}, H_{(N), (N-g+1)}}(\tau) + \frac{k_{ij1}\mu_1(t)}{H_{(N), (N-g)}} \left[\frac{k_{ij2}\mu_2(t)}{H_{(N), (N-g)}} E_{1,2}^{\mu_1(t), H_{(N), (N-g)}}(\tau) + \frac{k_{ij1}\mu_1(t)}{H_{(N), (N-g)}} E_{1,3}^{\mu_1(t), H_{(N), (N-g)}}(\tau) \right].$$

Таким образом, получены все вероятностно-временные характеристики, необходимые для решения сформулированных выше оптимизационных задач.

5. Примеры решения задач параметрической оптимизации СМО

Рассматриваются результаты решения задач параметрической оптимизации системы обслуживания перевозок узлового аэропорта при выполнении одной технологической операции. Интервал между комплексами принят достаточно продолжительным для того, чтобы комплексы могли считаться независимыми, $T_B = 12$ ч. Остальные исходные данные, сгруппированные в три совокупности, приняты следующими:

– характеристики потока заявок в течение комплекса:

$$T_1 = T_2 = 60 \text{ мин.}, \lambda^{\max} = 20 \text{ ВС/ч.}, \lambda^{\min} = 5 \text{ ВС/ч.};$$

– характеристики обслуживания: $T_1' = T_3' = 30$ мин.,

$$T_2' = 30 \text{ мин.}, T_{\text{обсл}1}^{\min} = 13,5 \text{ мин.}, T_{\text{обсл}1}^{\max} = 16,5 \text{ мин.}, T_{\text{обсл}1}^{\text{mid}} = 15 \text{ мин.},$$

$$\theta = 1,9, K = 30, N_{\min} = 4, N_{\max} = 7;$$

– характеристики качества обслуживания и использования

$$\text{ресурсов: } t_{\text{ож}P} = 5 \text{ мин.}, P_{\text{ож}P} = 0,95, t_{\text{преб}P} = 2T_{\text{обсл}1}^{\text{mid}} = 30 \text{ мин.},$$

$$P_{\text{преб}P} = 0,85, K_{3P} = 0,3, P_{\text{отк}P} = 0,001.$$

Все принятые значения соответствуют расписанию регионального хаба и аэропортовым нормам обслуживания перевозок. Расчётный коэффициент занятости каналов принят сравнительно малым, поскольку невысокая загрузка каналов позволяет обеспечить более высокое качество обслуживания, сократив время ожидания в очереди. Вероятность отказа в обслуживании с учётом специфики исследуемой системы задана весьма малой. Графики зависимостей $\lambda(t)$ и $T_{\text{обсл}1}(t)$, построенные на интервале времени продолжительностью 3 ч., отмеряемом от момента начала комплекса, в качестве иллюстрации приведены на рис.3.

Результаты решения обеих задач оптимизации в виде графиков зависимостей $N(t)$, $T_{\text{преб}}(t)$, $v_1(t)$, $P\{t_{\text{ож}}(t) < t_{\text{ож}P}\}$, $K_3(t)$ и $P\{t_{\text{преб}}(t) < t_{\text{преб}P}\}$, построенных на указанном интервале, приведены на рис.4а – 4е. Штриховыми линиями на рисунках отобра-

жён уровень ограничений. Как видно из рисунков, выбранная продолжительность интервала достаточна для того, чтобы в рассматриваемой СМО установился режим, характеристики которого мало не отличаются от стационарных.

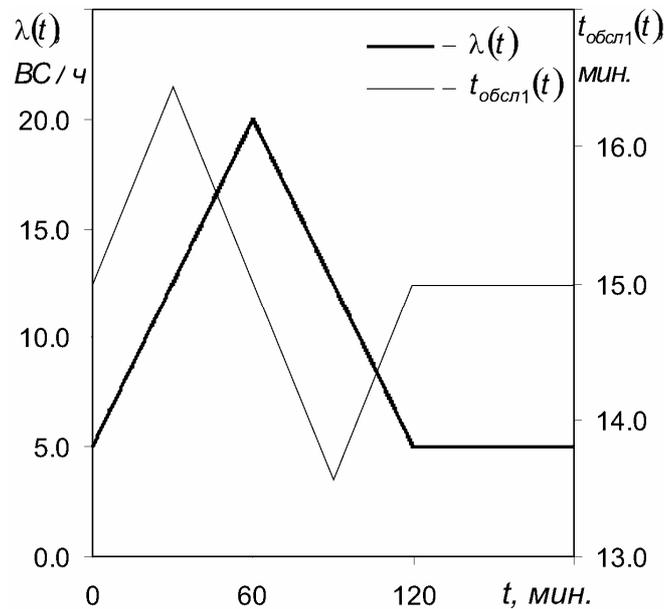


Рис. 3. Графики зависимостей $\lambda(t)$ и $T_{обсл1}(t)$

Сравнение графиков на рис.4а позволяет сделать вывод о существенной экономии потребного числа каналов, достигаемой в результате решения задачи А. Так, время, в течение которого предполагается использование всех семи доступных системе каналов для задачи А составляет около 57 мин. Аналогичный показатель для задачи Б значительно выше – около 100 мин. Таким образом, оптимизация численности обслуживающих каналов позволяет почти в два раза сократить время, в течение которого требуется задействовать максимальное число каналов.

Сокращение числа каналов достигается ценой роста времени пребывания заявки в системе. Сравнение результатов решения задач А и Б, приводимых на рис.4б, говорит о том, что в первом случае наблюдаются семь временных интервалов, а во втором – только один интервал, в течение которых среднее виртуальное время пребывания заявки в системе превышает 11 мин. Суммарная продолжительность таких интервалов в задаче А составляет 54 мин., в задаче Б – лишь 19 мин.

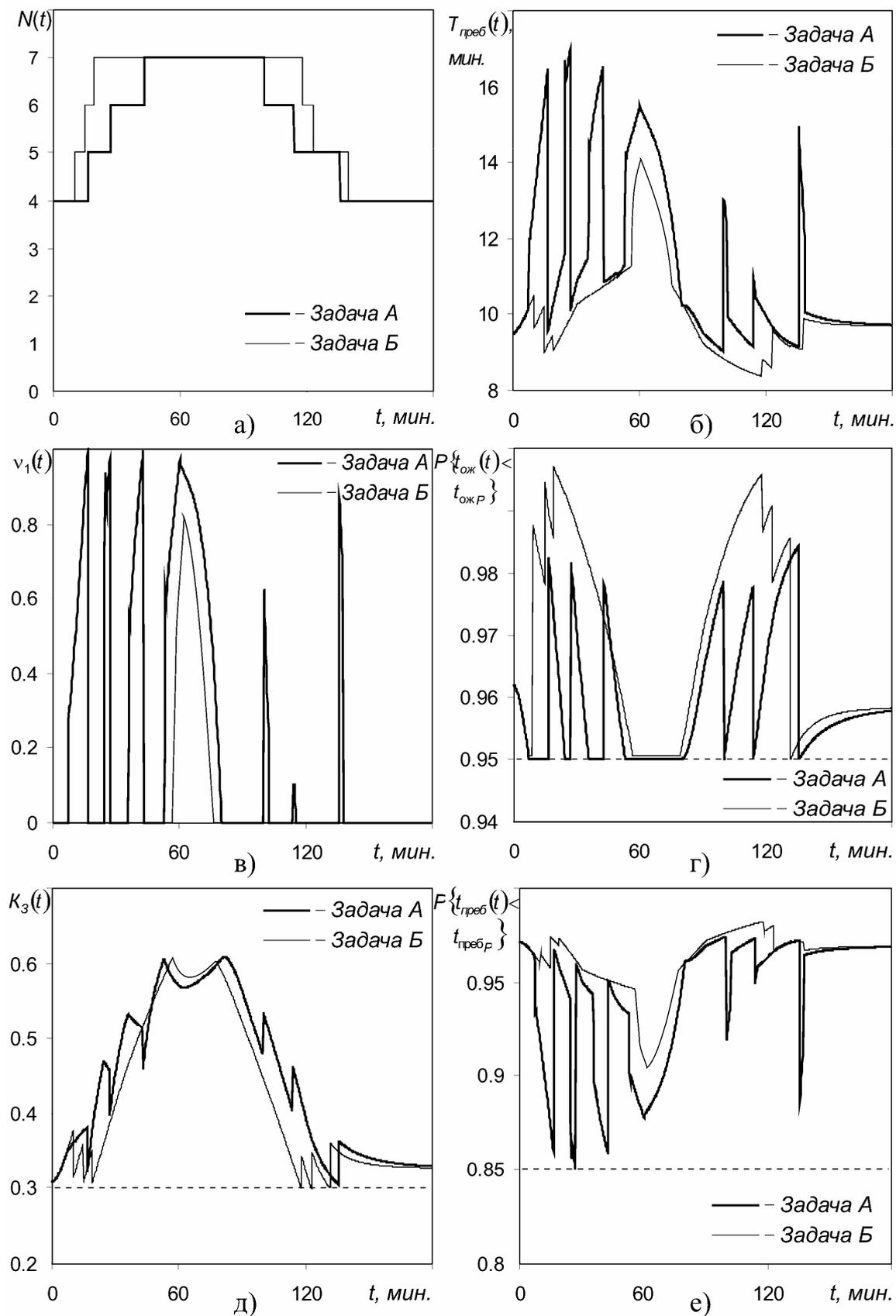


Рис. 4. Результаты решения задач оптимизации А и Б

Объяснить всплески среднего виртуального времени пребывания в системе в течение указанных интервалов позволяет сопоставление рис.4б, 4в, а также рис.4г и 4е. Как видно из перечисленных рисунков, рассматриваемым всплескам величины $T_{преб}(t)$ соответствуют резкие колебания всех остальных параметров СМО. Рассмотрим природу всплесков на примере первого из них в результатах задачи А. На начальном этапе комплекса, когда минимального числа каналов $N_{min} = 4$ более чем достаточно для выполнения всех ограничений задачи А, оптимизирующий алгоритм, «стремясь» уменьшить значения $T_{преб}(t)$, «определяет» всем поступающим заявкам обслуживание с использованием двух каналов. На рис.4в этому этапу соответствует краткий начальный интервал времени длительностью около 10 мин. с $v_1 = 0$.

Затем, с ростом интенсивности потока заявок и снижением интенсивности их обслуживания, минимального числа каналов, выделяемых попарно, становится недостаточно для обеспечения ограничения по времени ожидания (5). Алгоритм реагирует на это увеличением доли заявок, обслуживаемых индивидуально. Этот этап отражён на рисунках следующим образом. На рис.4в отчётливо различим первый «скачок» величины $v_1(t)$ до значения $v_1 \approx 1$. На рис.4г этапу соответствует участок постоянства величины $P\{t_{ож}(t) < t_{ожP}\}$ на уровне ограничения $P\{t_{ож}(t) < t_{ожP}\} > 0,95$. На рис.4б и 4е отображены колебания, соответствующие резкому росту $T_{преб}(t)$. Когда имеющегося числа каналов при любом $v_1(t)$ становится недостаточно для выполнения одного из ограничений, алгоритм вводит в СМО ещё один канал. На рисунках это проявляется в виде резкого падения значений $v_1(t)$, $T_{преб}(t)$ и такого же резкого повышения вероятностей $P\{t_{ож}(t) < t_{ожP}\}$, $P\{t_{преб}(t) < t_{пребP}\}$. Описанный процесс предваряет каждое увеличение числа $N(t)$. На завершающем этапе комплекса при снижении $\lambda(t)$ похожий процесс в обратном порядке воспроизводится перед каждым уменьшением числа $N(t)$.

Отметим, что интервалы всплесков на этапе роста интенсивности потока заявок имеют значительно большую продолжительность (в рассматриваемом примере от 3 до 10 мин.), чем на

этапе её снижения (не более 2 мин.). Наибольшей продолжительностью (27 мин.) характеризуется всплеск во время пикового уровня интенсивности потока

Единственный всплеск $T_{преб}(t)$ в результатах задачи Б объясняется выходом процесса на ограничение по времени ожидания в очереди. Поскольку при $N(t) = N_{max}$ и значениях интенсивности потока заявок, близких к λ_{max} , обслуживание каждой заявки парой каналов не обеспечивает его требуемого качества, а введение дополнительных каналов невозможно, то оптимизирующий алгоритм реагирует на это увеличением $v_1(t)$ с соответствующим ростом $T_{преб}(t)$. Если бы возможность управления величиной $v_1(t)$ отсутствовала, то имеющимся числом средств обеспечить обслуживание заданного потока заявок при заданных требованиях к качеству обслуживания было бы нельзя.

Как и следовало ожидать, решение задачи А приводит к несколько большей загрузке каналов, по сравнению с задачей Б (см.рис.4д). В целом, уровень загрузки может считаться удовлетворительным, а различия в загрузке для обеих задач несущественными.

Полученные результаты позволяют сформулировать следующие рекомендации в адрес производственных подразделений аэропорта:

– в случае устойчиво повторяющихся колебаний интенсивности прилётов и вылетов представляется целесообразным выработать определённую программу управления численностью обслуживающих средств и персонала и их перераспределением как между различными операциями, так и между индивидуальным и групповым обслуживанием в пределах одной операции;

– управление перераспределением средств между индивидуальным и групповым обслуживанием позволит эффективно влиять как на потребную численность средств, так и на временные характеристики процесса технического и коммерческого обслуживания ВС в аэропорту;

– если приоритетной целью аэропорта является сокращение времени простоя ВС на техническом и коммерческом обслуживании, то представляется целесообразным максимально возможное число ВС обслуживать группами из двух средств. Выбор

указанной цели требует от аэропорта повышения общей численности средств, что неизбежно влечёт снижение качества их использования между максимумами интенсивности комплексов. Ограничить увеличение численности средств можно применив во время пиков интенсивности управление, предполагающее повышение доли индивидуально обслуживаемых ВС;

– если приоритетной целью аэропорта является сокращение мгновенной численности используемых средств и персонала с ограничением времени простоя ВС, то программа управления должна быть более сложной. Между комплексами для обслуживания ВС средства желательно выделять попарно. На начальном этапе комплекса, предполагающем рост интенсивности потока ВС, введению «в дело» каждого дополнительного средства должен предшествовать сравнительно короткий (5-15 мин.) период, когда для обслуживания вновь прибывших ВС средства выделяются не парами, а индивидуально. Для конечного этапа комплекса, с наблюдающимся снижением интенсивности, зафиксировать краткосрочные интервалы с высокой долей индивидуального обслуживания на практике вряд ли возможно, поэтому здесь целесообразно использовать только групповое обслуживание. Во время пиков интенсивности рационально применять управление, предполагающее повышение доли индивидуально обслуживаемых ВС.

Приведённые результаты не только подтверждают правомерность постановки сформулированных оптимизационных задач и работоспособность алгоритма их решения, но и позволяют сформировать практические рекомендации по повышению эффективности управления ресурсами для такой перспективной группы аэропортов, как хабы с их нестационарными потоками рейсов и колеблющейся интенсивностью обслуживания.

Литература

1. БОЧАРОВ, П. П. Теория массового обслуживания / П. П. Бочаров, А. В. Печинкин. – М.: Изд-во РУДН, 1995. – 529 с.
2. ГОЛУБЕВ, И. С. Эффективность воздушного транспорта / И. С. Голубев. – М.: Транспорт, 1982. – 230 с.

3. РОМАНЕНКО, В. А. Математические модели функционирования узловых аэропортов в условиях современного авиа-транспортного рынка: монография / В. А. Романенко. – Самара: Изд-во «Ас Гард», 2010. – 224 с.
4. Руководство по экономике аэропортов. Документ ИКАО 9562. Второе издание. – Монреаль, 2006.

CONTROL OPTIMIZATION OF THE TECHNOLOGICAL PROCESSES HUB AIRPORT AS QUEUEING SYSTEM WITH NON-STATIONARY AND PARTIAL MUTUAL ASSISTANCE BETWEEN CHANNELS

Vladimir Romanenko, Samara State Aerospace University named after academician S.P.Korolyov (National Research University), Samara, Cand.Sci., associate professor, the person working for doctor's degree (vla_rom@mail.ru).

Abstract: Are formulated and solved problems of formation of an optimum program control in parameters of the hub airport transportation service system as queueing system with non-stationary flows and partial mutual assistance between channels. Solution problems results are considered. Recommendations for industrial subdividing of the hub airports are formulated.

Keywords: optimal control, queueing system, hub airport.