Математическая Теория Игр и её Приложения, т.8, в.3, с. 67–99

УДК 519.711.7
ББК 22.1

ОПТИМАЛЬНЫЕ ОБРАЩЕНИЯ К 2-СЕРВЕРНОЙ СИСТЕМЕ С ПОТЕРЯМИ И РАЦИОНАЛЬНЫМ СЛУЧАЙНЫМ ДОСТУПОМ

Юлия В. Чиркова*
Институт прикладных математических исследований
Карельского научного центра РАН
185910, Петрозаводск, ул. Пушкинская, 11
e-mail: julia@krc.karelia.ru

В работе исследуется 2-серверная система обслуживания с потерями, которая принимает запросы на интервале времени $[0, T]$. Пользователи отправляют свои запросы в систему, которая либо случайным образом с известной пользователям вероятностью перенаправляет их на один из двух свободных серверов, либо на единственный свободный сервер, либо отказывает в обслуживании. Для данной системы рассматривается некооперативная игра, в которой стратегией игрока является момент времени обращения к системе обслуживания и выигрышем является вероятность, что его запрос получит обслуживание. В качестве критерия оптимальности используется симметричное равновесие по Нэшу. Для данной игры рассматриваются две модели. В первой число игроков фиксировано, во второй является случайной величиной, имеющей распределение Пуассона. Для обоих случаев доказано существование

* Работа выполняется в рамках проекта «Задачи оптимального распределения ресурсов и защиты информации в высокопроизводительных вычислительных системах и сетях» по программе фундаментальных исследований ОМН РАН, а также поддержана грантом РГНФ №15-02-00352_а.

единственного симметричного равновесия и проведены численные эксперименты по сравнению равновесий при различных значениях параметров модели, а также сравнению ее эффективности с односерверной моделью и моделью со случайным доступом, в которой система, перенаправляющая запросы на сервера, не имеет информации об их занятости.

Ключевые слова: система обслуживания, оптимальные поступления, равновесие по Нэшу.

1. Введение

С развитием информационных технологий активно разрабатывается теория очередей (queueing theory) в применении к анализу работы вычислительных сетей. Один из классов задач данного направления связан с управлением загруженностью множества серверов, которые рассматриваются как система массового обслуживания, обрабатывающая поток пользовательских заявок. В зависимости от назначения и условий работы система может обрабатывать одновременно один или несколько запросов на обслуживание, может иметь одну или несколько очередей, или же терять запросы в случае занятости системы. Теоретико-игровой подход позволяет рассматривать такую задачу как игру, в которой участники действуют эгоистично и могут достигать некоторого равновесного состояния, когда никому не выгодно отклоняться от выбранной стратегии. В работах [1, 2, 9, 10, 15, 16] исследуются модели, в которых дисциплина поступления запросов в систему определяется стратегиями игроков, при этом в [1, 2, 15, 16] игроки стремятся максимизировать вероятность получения обслуживания, а в [9, 10] критерием оптимальности является минимизация времени ожидания запросов в очереди на обслуживание. В моделях [3, 4, 6, 7] дисциплина поступления запросов задается сверху, а в качестве стратегии рассматривается схема выбора одной из очередей. В работах [5, 8] рассматриваются модели, в которых пользователь, зная длину очереди на обслуживание на мощном сервере общего доступа, решает, отправлять ли свой запрос в очередь или выполнить его на своей рабочей станции, стремясь минимизировать временные затраты. Работа [2] представляет вероятностное расширение безбуферного варианта модели, исследуемой в работах [15, 16].

На заданном временном интервале в систему поступают запросы и принимаются на обслуживание при наличии свободных мест. В системе имеется два обслуживающих сервера с идентичной функциональностью и, возможно, различной производительностью. Когда пользователь обращается к системе, он случайным образом перенаправляется на один из серверов, на котором запрос либо принимается на обслуживание, либо теряется.

Примером организации такого случайного доступа в Internet является использование циклического алгоритма (round robin [12, 13]) в DNS-системе для распределения нагрузки между несколькими серверами, которые предоставляют некоторый Web-сервис. В этом случае разные пользователи получают разные IP адреса при обращении к домену. В простейшем случае IP адреса выдаются по очереди (сначала первый, потом — второй и т.п.), в более общем каждый адрес выдается с определенной вероятностью. Кроме задач распределения нагрузки между Web-ресурсами модели со случайным доступом могут найти приложение в задачах оптимизации облачных вычислений [14], работы call-центров и т.п.

Данная работа представляет модель, аналогичную [2], в которой система действует рационально при перенаправлении заявок на серверы. В модели [2] система не имеет информации о состоянии серверов и заявка может быть перенаправлена на занятый сервер при имеющемся свободном. Так на практике система DNS, выдающая клиенту очередной IP-адрес запрашиваемого домена, не имеет информации о его доступности. Здесь же доступ к серверам будет случайным только в случае, когда оба сервера свободны. Кроме того, как частный случай данной модели рассматривается задача с различными по производительности серверами, доступ к которым определяется согласно установленным системой приоритетам.

Для данной системы рассматривается некооперативная игра, в которой чистой стратегией игрока является момент времени обращения к системе обслуживания и смешанной – распределение моментов таких обращений. Выигрышем является вероятность, что его запрос получит обслуживание. В качестве критерия оптимальности используется симметричное равновесие по Нэшу. Для данной игры рассматриваются две модели. В первой число игроков фиксировано, во второй является случайной величиной, имеющей распределение Пуассона. Для обоих случаев доказано существование единственного симметричного равновесия и предложены алгоритмы для численного нахождения равновесий. Данные алгоритмы положены в основу реализации программной системы, позволяющей графически представлять равновесные стратегии при различных значениях параметров модели для их визуального сравнения.

2. Модель

2.1. Система обслуживания

Рассмотрим систему массового обслуживания с 2 серверами, которая принимает запросы от пользователей на интервале времени $[0, T]$. Каждый из серверов системы способен обслуживать не более одного запроса одновременно. Времена обслуживания запросов – независимые и экспоненциально распределенные случайные величины с интенсивностями μ_1 и μ_2 на первом и втором серверах соответственно.

При обращении пользователя к системе в случае, если оба сервера свободны, его запрос с вероятностью r направляется на первый сервер и с вероятностью $\overline{r} = 1 - r$ на второй. Если свободен только один из серверов, то заявка перенаправляется на него. Если оба сервера заняты, заявка получает отказ в обслуживании. В системе отсутствует какая-либо очередь, заявка, получившая отказ в обслуживании, теряется. Заметим, что при $r \in \{0, 1\}$ рассматриваемая модель является системой обслуживания с приоритетами.

Возможна ситуация, когда несколько пользователей обращаются в систему в одно и то же время. В этом случае получают обслуживание столько из них, сколько серверов в данный момент свободно. Один или два запроса, получающие обслуживание, выбираются из числа всех поступивших одновременно запросов равновероятным образом.

Дисциплина поступления запросов в систему не задана. Она определяется пользователями системы, стремящимися максимизировать вероятность получения получения обслуживания для своих запросов.

2.2. Игра

Расмотрим задачу определения оптимальной дисциплины поступления запросов в систему как некооперативную игру. Игроками являются пользователи системы, отправляющие запросы на обслуживание. Множество игроков обозначим $\mathbb{N}$. Каждый игрок выбирает момент времени для обращения к системе, стремясь максимизировать вероятность получения обслуживания для своего запроса. Чистой стратегией игрока i является выбранный им момент времени t_i обращения к системе. Смешанная стратегия игрока i – функция распределения $F_i(t)$ моментов обращения к системе на интервале $[0, T]$. Профиль стратегий обозначим $F = \{F_i(t), i \in \mathbb{N}\}$. Выигрышем игрока в момент t является вероятность получения обслуживания при обращении к системе в момент t.

Все игроки одинаковы, независимы, и действуют эгоистично, без кооперации, поэтому в качестве критерия оптимальности рассматривается симметричное равновесие по Нэшу. В этом случае все стратегии игроков одинаковы: $F_i(t) = F(t)$ для всех i.

Определение 2.1. *Функция распределения $F(t)$ моментов обращения к системе t является симметричным равновесием по Нэшу, если существует константа C, такая что в любой момент $t \in [0, T]$ вероятность получения обслуживания не более C и равна C на носителе $F(t)$.*

Задача рассматривается для двух случаев. В первом количество игроков фиксировано и известно каждому игроку. Во втором каждый игрок знает, что количество противников имеет распределение Пуассона с известным параметром. В обоих случаях, как и в работе [16], симметричное равновесие по Нэшу существует и единственно, а его структура обладает следующими свойствами.

Рассмотрим некоторого игрока, который пытается обратиться к системе в момент t, зная, что остальные игроки обращаются в этот момент с вероятностью p. Обозначим X_p случайную величину противников данного игрока, которые обращаются в систему и могут помешать ему получить обслуживание в данный момент. Количество игроков может быть фиксировано, и тогда у каждого игрока N противников, либо N является случайной величиной. Тогда для каждого значения случайной или фиксированной величины N случайная величина $X_p = X_{N,p}$ имеет распределение $Bin(N, p)$. Для удобства обозначения мы опускаем индекс N, указывая в контексте, является число игроков фиксированным или случайным.

Если p – вероятность обращения в систему в нулевой момент времени, когда система изначально свободна, то вероятность получения обслуживания в этот момент равна

$$C(p) = E\left(\min\{1, \frac{2}{X_p+1}\}\right). \tag{2.1}$$

В случае, если N фиксирована, данная вероятность равна

$$P(X_p < 2) + \sum_{i=2}^{n}\left(\frac{2}{i+1}P(X_p = i)\right),$$

и для случайной N равна

$$P(X_p < 2) + \sum_{n=2}^{\infty} P(N = n) \sum_{i=2}^{N}\left(\frac{2}{i+1}P(X_p = i)\right).$$

Пусть некоторое время после нулевого момента на интервале $(0, t)$ нет поступлений. Тогда вероятность получить обслуживание в момент $t > 0$ равна

$$1 - P(X_p \geq 2)e^{-(\mu_1+\mu_2)t} = P(X_p < 2) + P(X_p \geq 2)(1 - e^{-(\mu_1+\mu_2)t}).$$

Заметим, что данная вероятность возрастает по t.

Пусть теперь игрок отправляет в систему запрос в момент $0+$, бесконечно близкий к моменту 0, но сразу после него такой, что каждый из поступивших в нулевой момент запросов либо начал обслуживаться, либо покинул систему, но ни один новый запрос не поступил в систему и ни один не завершил свое обслуживание. Эта вероятность равна

$$\lim_{t\to 0+}\left(P(X_p < 2) + P(X_p \geq 2)(1 - e^{-(\mu_1+\mu_2)t})\right) = P(X_p < 2),$$

что меньше, чем $C(p)$. Это означает, что лучше обратиться в систему в нулевой момент времени, чем сразу после него. Тогда имеет место следующая лемма.

Лемма 2.1. *Выигрыш игрока в нулевой момент времени строго больше, чем выигрыш в момент $0+$, бесконечно близкий к нулевому моменту, сразу после него. Выигрыш игрока в момент t возрастает по времени t, если на интервале $(0, t)$ нет обращений в систему.*

Лемма 2.2. *Носитель стратегии в равновесии содержит атом в точке $t = 0$, то есть равновесная вероятность $p_e = F(0)$ обращения в начальный нулевой момент времени строго положительна. Далее имеется интервал времени $(0, t_e)$, на котором нет обращений к системе.*

Доказательство. Положительность вероятности обращения в систему в нулевой момент времени обусловлена тем, что если бы никто не обращался в систему в этот момент, то при обращении к системе в нулевой момент времени любой отклонившийся от равновесия игрок получал бы обслуживание с вероятностью 1.

Лемма 2.1 показывает, что некоторое время после нулевого момента выигрыш игрока, возрастая по времени, остается меньше, чем в нулевой момент, даже если известно, что в это время никто не обращается в систему. Этим объясняется наличие интервала времени без обращений после нулевого момента времени. На этом интервале значение выигрыша возрастает до уровня равновесного в момент t_e, до которого обращений в систему нет. □

Таким образом, в носителе равновесной стратегии существует область разрыва $(0, t_e)$. Эта область возникает по причине того, что в нулевой момент времени, когда система изначально свободна, возможно поступление в систему нескольких обращений одновременно и вероятность выиграть в розыгрыше мест обслуживания в этот момент выше, чем вероятность получить обслуживание сразу после него, когда система с достаточно большой вероятностью оказывается занята.

Предположим теперь, что равновесная вероятность обращения игрока к системе в нулевой момент времени известна и равна $0 < p_e \leq 1$. Вероятность получить обслуживание в момент $t > 0$, такой что на интервале $(0, t)$ нет поступлений запросов в систему, возрастает по t и стремится к 1. В момент $t = 0+$ вероятность получения обслуживания меньше, чем в нулевой момент. Поэтому, существует момент t_e (возможно, после момента T), где данные вероятности равны. Тогда решение уравнения

$$E\left(\min\{1, \frac{2}{X_{p_e}+1}\}\right) = 1 - P(X_{p_e} \geq 2)e^{-(\mu_1+\mu_2)t_e} \qquad (2.2)$$

дает момент $t_e > 0$ для данной вероятности обращения игрока в нулевой момент времени. При этом после нулевого момента и до момента t_e обращений в систему нет.

Для доказательства следующей леммы будем использовать аппарат теории стохастических порядков [17]. Приведем формулировки определений из [17] в терминах нашего случая, а также теоремы, являющиеся следствиями теорем из [17], описывающих более общий случай, на которые мы будем ссылаться в данном доказательстве. При этом, как и в данном источнике, под возрастанием будем иметь ввиду неубывание. Иначе, будем писать "строгое возрастание". Кроме того, будем использовать следующее обозначение из данного источника. $[Z|A]$ обозначает случайную величину, имеющую в качестве своего распределения условное распределение величины Z при условии A.

Определение 2.2. *Случайная величина X_p стохастически возрастает по p согласно обычному стохастическому порядку тогда и только тогда, когда вероятность $P(X_p \geq i)$ возрастает по p для всех ее возможных значений $i \in (-\infty, +\infty)$.*

Определение 2.3. *Случайная величина X_p стохастически возрастает по p согласно порядку отношения правдоподобия тогда и только тогда, когда $\frac{P(X_p=i)}{P(X_q=i)}$ возрастает по $i \in (-\infty, +\infty)$ для любых $0 \leq p \leq q \leq 1$.*

Теорема 2.1. Из теоремы 1.C.1 [17] *Если случайная величина X_p стохастически возрастает согласно порядку отношения правдоподобия, то она стохастически возрастает согласно обычному стохастическому порядку.*

Теорема 2.2. Из теоремы 1.C.6 [17] *Если случайная величина X_p стохастически возрастает согласно порядку отношения правдоподобия, то случайная величина $[X_p|X_p \geq i]$ также стохастически возрастает согласно порядку отношения правдоподобия.*

Теорема 2.3. Из теоремы 1.C.8 [17] *Если случайная величина X_p стохастически возрастает согласно порядку отношения правдоподобия, а ψ – любая возрастающая функция, то случайная величина $\psi(X_p)$ также стохастически возрастает согласно порядку отношения правдоподобия.*

Теорема 2.4. Из теоремы 1.C.9 [17] *Если случайные величины $X_{n,p}$ для всех $n = 1, 2, \ldots$ стохастически возрастают по p согласно порядку отношения правдоподобия, а их плотности распределения обладают свойством логарифмической вогнутости, то случайная величина $\sum\limits_n X_{n,p}$ также стохастически возрастает по p согласно порядку отношения правдоподобия.*

Лемма 2.3. *Для любых вещественных $0 < p_e \leq 1$ и $\mu_1, \mu_2 > 0$ уравнение (2.2) определяет функцию $t_e(p_e)$, строго убывающую по p_e.*

Доказательство. Данная лемма, фактически, является частным случаем леммы 3 в [15]. Идея доказательства такая же, как в работах [15] и [16], но мы приведем доказательство более детально.

Преобразуем уравнение (2.2) следующим образом:

$$\frac{1 - E\left(\min\{1, \frac{2}{X_{p_e}+1}\}\right)}{P(X_{p_e}) \geq 2} = e^{-(\mu_1+\mu_2)t_e},$$

$$\frac{E\left(\max\{0, \frac{X_{p_e}-1}{X_{p_e}+1}\}\right)}{P(X_{p_e}) \geq 2} = e^{-(\mu_1+\mu_2)t_e},$$

и далее получим

$$E\left(\frac{X_{p_e}-1}{X_{p_e}+1} | X_{p_e} \geq 2\right) = e^{-(\mu_1+\mu_2)t_e}. \tag{2.3}$$

Правая часть уравнения (2.3) не зависит от p_e и строго убывает по t_e. Левая часть зависит только от p_e, нужно показать, что она возрастает по p_e.

Пусть сначала величина N фиксирована. Проверим, что для любых вещественных $0 \leq q < p \leq 1$ отношение $\frac{P(X_p=i)}{P(X_q=i)}$ возрастает относительно натурального индекса i. Отношение $\frac{P(X_p=k+1)}{P(X_p=k)} = \frac{N-k}{k+1}\left(\frac{1}{1-p} - 1\right)$ возрастает по p. Тогда справедливо $\frac{P(X_q=k+1)}{P(X_q=k)} < \frac{P(X_p=k+1)}{P(X_p=k)}$ и $\frac{P(X_p=k)}{P(X_q=k)} < \frac{P(X_p=k+1)}{P(X_q=k+1)}$. Это означает по определению 2.3, что случайная величина X_{p_e} стохастически возрастает по p_e, согласно порядку отношения правдоподобия.

Пусть теперь N случайна. По теореме 2.4, случайная величина $\sum\limits_{n=1}^{\infty} P(N=n)X_{n,p_e}$ также стохастически возрастает по p_e, согласно порядку отношения правдоподобия, где X_{n,p_e} обозначает X_{p_e} для каждого фиксированного значения $N = n$, так как плотность биномиального распределения обладает свойством логарифмической вогнутости [11]. Тогда случайная величина X_{p_e} стохастически возрастает по p_e, согласно порядку отношения правдоподобия, для случайной N.

По теореме 2.2 случайная величина $[X_{p_e}|X_{p_e} \geq 2]$ также стохастически возрастает по p_e согласно порядку отношения правдоподобия. Тогда по теореме 2.3 и случайная величина $\left[\frac{X_{p_e}-1}{X_{p_e}+1}|X_{p_e} \geq 2\right]$ стохастически возрастает по p_e согласно порядку отношения правдоподобия.

По теореме 2.1 из стохастического возрастания случайной величины согласно порядку отношения правдоподобия следует ее возрастание согласно обычному стохастическому порядку, и, следовательно, возрастание ее математического ожидания. Таким образом, $E\left(\frac{X_{p_e}-1}{X_{p_e}+1}|X_{p_e} \geq 2\right)$ возрастает по p_e. □

Следствие 2.1. *Значение математического ожидания* $E\left(\frac{1}{X_{p_e}+1}\right)$ *убывает по* p_e.

Доказательство. Из доказательства леммы 2.3 имеем, что случайная величина $\frac{1}{X_{p_e}+1}$ стохастически убывает по p_e согласно обычному стохастическому порядку как убывающая функция от случайной величины, стохастически возрастающей по p_e. Тогда ее математическое ожидание также убывает по p_e. □

В соответствии с леммой 2.3, чем больше вероятность p_e обращения к системе в нулевой момент, тем раньше момент t_e, когда возобновляются поступления запросов в систему после интервала без поступления запросов, начавшегося сразу за нулевым моментом. Заметим, что значение t_e для данного p_e может получиться больше, чем T. В этом случае необходимо увеличивать значение p_e. Если же $t_e(1) \geq T$, тогда равновесной является чистая стратегия, согласно которой все запросы приходят в систему в нулевой момент времени с вероятностью 1. Далее в данном разделе будем предполагать, что $t_e(1) < T$.

Лемма 2.4. *Если* $t_e < T$, *то с момента* t_e *до конца* T *существует строго положительная плотность распределения моментов обра-*

щения к системе $f(t) > 0$. Данный интервал не содержит атомов и разрывов.

Доказательство. Рассмотрим интервал $[t_e, T]$, на котором возобновляются обращения в систему. Покажем, что равновесная плотность распределения моментов обращения в систему на всем интервале строго положительна. Пусть, наоборот, на интервале $[t_e, T]$ существует некоторая область (s_1, s_2), на которой обращений в систему нет. Пусть $p_{ij}(t)$ – вероятность состояния (i, j) системы в момент t, где $i, j \in \{0, 1\}$ обозначают состояния занятости (0 – свободен, 1 – занят) первого и второго серверов соответственно. Тогда вероятность получения обслуживания в момент s_1 равна

$$p_{00}(s_1) + p_{01}(s_1) + p_{10}(s_1).$$

С учетом того, что на (s_1, s_2) нет поступлений, выигрыш в момент s_2 равен

$$\begin{aligned}
&p_{00}(s_2) + p_{01}(s_2) + p_{10}(s_2) = \\
&p_{00}(s_1) + p_{10}(s_1)(1 - e^{-\mu_1(s_2 - s_1)}) + p_{01}(s_1)(1 - e^{-\mu_2(s_2 - s_1)}) + \\
&\qquad p_{11}(s_1)(1 - e^{-\mu_1(s_2 - s_1)})(1 - e^{-\mu_2(s_2 - s_1)}) \\
&\qquad p_{01}(s_1) + p_{11}(s_1)(1 - e^{-\mu_1(s_2 - s_1)}) + \\
&\qquad p_{10}(s_1) + p_{11}(s_1)(1 - e^{-\mu_2(s_2 - s_1)}),
\end{aligned}$$

что больше выигрыша в момент s_1. Следовательно, в равновесии таких разрывов в носителе стратегии после момента t_e не может быть.

Покажем, что после момента t_e носитель стратегии не содержит атомов. Пусть такой атом есть в точке $t \in [t_e, T]$ и вероятность обращения игрока в момент t строго положительна и равна p. Пусть момент $s = t-$ перед моментом t и бесконечно близкий к нему, такой, что вероятности ухода обслуженных и поступления новых заявок между этими моментами равны нулю. Рассмотрим некоторого игрока, который пытается обратиться в систему в момент t, зная, что остальные обращаются в этот же момент с вероятностью p. Пусть случайная величина X_p обозначает число его противников, обратившихся в систему в момент времени t. Так как вероятность p строго положительна, то и математическое ожидание X_p должно быть положительно. Тогда вероятность получения обслуживания в момент времени t равна

$$p_{00}(t)E\left(\min\{1,\tfrac{2}{X_p+1}\}\right)+(p_{01}(t)+p_{10}(t))E\left(\tfrac{1}{X_p+1}\right)=$$
$$p_{00}(s)E\left(\min\{1,\tfrac{2}{X_p+1}\}\right)+(p_{01}(s)+p_{10}(s))E\left(\tfrac{1}{X_p+1}\right),$$

что меньше выигрыша в момент s, равного

$$p_{00}(s)+p_{01}(s)+p_{10}(s).$$

То есть в случае, если в распределении обращений в систему после момента t_e есть атом, всегда лучше обратиться в систему непосредственно до него. В отличие от нулевого момента времени, когда система изначально свободна и игроку нужно только выиграть розыгрыш мест для получения обслуживания, здесь сервера может быть уже заняты, и розыгрыш мест на обслуживание дополнительно снижает вероятность получения обслуживания. □

3. Фиксированное число игроков

Обозначим $N+1$ количество игроков, которые отправляют запросы в систему. У каждого из них N противников, которые могут помешать ему получить обслуживание своего запроса. Полагаем $N>1$, так как иначе обслуживание будет получено всеми игроками при обращении в любой момент времени. Пусть с вероятностью p_e каждый из N противников обращается к системе в момент времени $t=0$. Случайная величина X_{p_e} – число игроков, обратившихся к серверу в нулевой момент времени, ее распределение $Bin(N,p_e)$. В начальный момент оба сервера свободны, поэтому обслуживание получат два запроса, если $X_{p_e}\geq 2$, либо один. Тогда вероятность получения обслуживания в нулевой момент времени равна, согласно (2.1),

$$C(p_e)=(1-p_e)^N+C_N^1p_e(1-p_e)^{N-1}+\sum_{i=2}^{N}C_N^ip_e^i(1-p_e)^{N-i}\tfrac{2}{i+1}=$$
$$(1-p_e)^N+\sum_{i=1}^{N}C_N^ip_e^i(1-p_e)^{N-i}\tfrac{2}{i+1}=$$
$$(1-p_e)^N+2\tfrac{1-(1-p_e)^{N+1}}{p_e(N+1)}-2(1-p_e)^N=2\tfrac{1-(1-p_e)^{N+1}}{p_e(N+1)}-(1-p_e)^N. \tag{3.1}$$

Найдем вероятность получения обслуживания в момент $t_e>0$ при условии, что вероятности обращений к системе на интервале

$(0, t_e)$ равны нулю. Эта вероятность равна

$$1 - P(X_{p_e} \geq 2)e^{-(\mu_1+\mu_2)t_e} = \\ 1 - (1 - P(X_{p_e} = 0) - P(X_{p_e} = 1))e^{-(\mu_1+\mu_2)t_e} = \\ 1 - (1 - (1 - p_e)^N - Np_e(1 - p_e)^{N-1})e^{-(\mu_1+\mu_2)t_e}.$$

Тогда уравнение (2.2) для случая фиксированного числа игроков имеет вид

$$\frac{2(1 - (1 - p_e)^{N+1})}{p_e(N + 1)} - (1 - p_e)^N = \\ 1 - (1 - (1 - p_e)^N - Np_e(1 - p_e)^{N-1})e^{-(\mu_1+\mu_2)t_e}. \quad (3.2)$$

Согласно лемме 2.3, если $t_e(1) \geq T$, то равновесной является чистая стратегия – обращение в систему в момент $t = 0$ с вероятностью 1. Иначе на интервале $[t_e, T]$ существует строго положительная плотность распределения моментов обращения в систему. Далее будем считать, что $t_e(1) < T$.

Необходимо найти равновесную плотность $f(t)$ распределения моментов обращения к системе на интервале $[t_e, T]$. Определим Марковский процесс с состояниями системы (i, j, k) в каждый момент времени $t \in [t_e, T]$, где $i, j \in \{0, 1\}$ – состояния занятости первого и второго сервера соответственно (0 – свободен, 1 – занят), $k \in \{0, \ldots, N\}$ – число игроков, обратившихся в систему до момента t. Данный процесс неоднороден по времени, так как интенсивность поступлений запросов в систему скачком уменьшается с каждым обращением очередного игрока в систему и равна $\lambda_k(t) = (N - k)\frac{f(t)}{1-F(t)}$. Система Колмогорова для вероятностей состояний системы p_{ijk} выглядит следующим образом.

$$\begin{aligned}
&p'_{000}(t) = -\lambda_0(t)p_{000}(t)\\
&p'_{101}(t) = r\lambda_0(t)p_{000}(t) - (\lambda_1(t) + \mu_1)p_{101}(t)\\
&p'_{011}(t) = \overline{r}\lambda_0(t)p_{000}(t) - (\lambda_1(t) + \mu_2)p_{011}(t)\\
&p'_{00i}(t) = -\lambda_i(t)p_{00i}(t) + \mu_1 p_{10i}(t) + \mu_2 p_{01i}(t)\\
&p'_{10i}(t) = r\lambda_{i-1}(t)p_{00i-1}(t) - (\lambda_i(t) + \mu_1)p_{10i}(t) + \mu_2 p_{11i}(t)\\
&p'_{01i}(t) = \overline{r}\lambda_{i-1}(t)p_{00i-1}(t) - (\lambda_i(t) + \mu_2)p_{01i}(t) + \mu_1 p_{11i}(t)\\
&p'_{11i}(t) = \lambda_{i-1}(t)(p_{01i-1}(t) + p_{10i-1}(t) + p_{11i-1}(t)) - (\lambda_i(t) + \mu_1 + \mu_2)p_{11i}(t)\\
&\qquad i = 2, \ldots, N.
\end{aligned} \quad (3.3)$$

Вероятность получить обслуживание в момент $t \in [t_e, T]$ в равновесии постоянна для всех $t \in [t_e, T]$:

$$\sum_{i=0}^{N} p_{00i}(t) + \sum_{i=1}^{N} p_{01i}(t) + \sum_{i=1}^{N} p_{10i}(t) = 1 - \sum_{i=2}^{N} p_{11i}(t) = C(p_e).$$

Тогда сумма соответствующих производных должна быть равна нулю. Отсюда, подставив в суммы производные вероятностей состояний из системы Колмогорова (3.3), получаем дифференциальное уравнение для определения равновесной плотности

$$\frac{f(t)}{1-F(t)} = \frac{(\mu_1+\mu_2)(1-C(p_e))}{\sum\limits_{i=0}^{N-1}(N-i)(p_{01i}(t)+p_{10i}(t))}. \tag{3.4}$$

для $t \in [t_e, T]$.

Носитель равновесного распределения моментов обращений в систему находится на интервале $[0, T]$, поэтому должно выполняться $F(T) = 1$. Это создает неопределенность в уравнении (3.4) в точке $t = T$, принадлежащей интервалу, на котором обращения в систему определяются положительной плотностью распределения. Преобразуем уравнение (3.4) для получения равновесного распределения, чтобы исключить из него множитель $1 - F(t)$.

Обозначим T_i моменты обращений игроков $i \in \{1, \dots, N\}$, независимые и одинаково распределенные согласно функции F. Пусть $A(t)$ – число обращений в систему до момента t и $B_N(t) \in \{0, 1, 2\}$ – количество занятых серверов в момент t.

Знаменатель правой части в уравнении (3.4) можно преобразовать как

$$N\sum_{i=0}^{N}(p_{01i}(t)+p_{10i}(t)) - \sum_{i=0}^{N} i(p_{01i}(t)+p_{10i}(t)) =$$
$$NP(B_N(t)=1) - E(A(t)\mathbb{1}_{B_N(t)=1}).$$

Преобразуем вычитаемое

$$\phi(t) = E(A(t)\mathbb{1}_{B_N(t)=1}) = E\sum_{i=1}^{N} \mathbb{1}_{B_N(t)=1, T_i \le t} =$$
$$E\sum_{i=1}^{N} \mathbb{1}_{B_N(t)=1, T_1 \le t} = NE\mathbb{1}_{B_N(t)=1, T_1 \le t} =$$
$$NP(B_N(t)=1, T_1 \le t).$$

При $t \to T$ вероятность $P(B_N(t) = 1, T_1 \leq t)$ стремится к $P(B_N(t) = 1)$ и знаменатель стремится к нулю, что создает неопределенность для плотности в уравнении (3.4). Далее, предполагая, что $t < T$, получим выражение для плотности и доопределим значение плотности в точке T как предел.

$$\phi(t) = N(P(B_N(t) = 1) - P(B_N(t) = 1, T_1 > t)).$$

Тогда знаменатель правой части в уравнении (3.4) равен

$$NP(B_N(t) = 0, T_1 > t) = NP(B_N(t) = 0 | T_1 > t)(1 - F(t) = \\ NP(B_{N-1}(t) = 0)(1 - F(t)),$$

где $B_{N-1}(t)$ – число занятых серверов в момент t в модели с числом игроков меньшим на 1, где моменты обращений $N-1$ игроков независимые и одинаково распределенные согласно той же функции F. Заметим, что для этой модели интенсивности поступлений обращений в систему $\lambda_i(t)$ остаются теми же, что и для N игроков, с точностью до постоянных множителей $\frac{N-1}{N}$, определяемых числом игроков, так как распределение остается неизменным, однако вероятности состояний здесь свои.

Тогда для $t \in [t_e, T)$ плотность распределения моментов обращения в систему можно представить выражением, не зависящим от $1 - F(t)$

$$f(t) = \frac{(\mu_1 + \mu_2)(1 - C(p_e))}{NP(B_{N-1}(t) = 1)}. \tag{3.5}$$

Правая часть данного выражения определена для $t = T$ и тогда по непрерывности можно доопределить плотность в точке $t = T$

$$f(T) = \frac{(\mu_1 + \mu_2)(1 - C(p_e))}{NP(B_{N-1}(T) = 1)}.$$

То есть выражение (3.5) определяет плотность распределения обращений в систему на всем интервале $[t_e, T]$.

Данное выражение позволяет, преобразовав систему (3.3), исключить из нее неизвестную функцию плотности $f(t)$ и получить систему дифференциальных уравнений относительно вероятностей состояний.

Определим для системы начальные условия – вероятности состояний в момент t_e, при условии, что на $(0, t_e)$ нет поступлений.

$$\begin{aligned}
&p_{00i}(t_e) = C_N^i p_e^i (1-p_e)^{N-i} \left[\mathbb{1}_{i=0} + \mathbb{1}_{i=1} \left(r(1-e^{-\mu_1 t_e}) + \overline{r}(1-e^{-\mu_2 t_e})\right) + \right.\\
&\qquad \left. \mathbb{1}_{i>1}(1-e^{-\mu_1 t_e})(1-e^{-\mu_2 t_e})\right]\\
&\qquad i = 0, \ldots, N\\
&p_{10i}(t_e) = C_N^i p_e^i (1-p_e)^{N-i} \left[\mathbb{1}_{i=1} r e^{-\mu_1 t_e} + \mathbb{1}_{i>1} e^{-\mu_1 t_e}(1-e^{-\mu_2 t_e})\right]\\
&\qquad i = 1, \ldots, N\\
&p_{01i}(t_e) = C_N^i p_e^i (1-p_e)^{N-i} \left[\mathbb{1}_{i=1} \overline{r} e^{-\mu_2 t_e} + \mathbb{1}_{i>1}(1-e^{-\mu_1 t_e}) e^{-\mu_2 t_e}\right]\\
&\qquad i = 1, \ldots, N\\
&p_{11i}(t_e) = C_N^i p_e^i (1-p_e)^{N-i} \mathbb{1}_{i>1} e^{-\mu_1 t_e} e^{-\mu_2 t_e}\\
&\qquad i = 2, \ldots, N
\end{aligned} \tag{3.6}$$

Таким образом, имеется задача Коши с системой диференциальных уравнений (3.3) и начальными условиями (3.6). Ее решение дает вероятности состояний $p_{ijk}(t)$, зависящие от параметра p_e. С известными вероятностями состояний выражение (3.5) определяет равновесную плотность распределения моментов обращения $f(t)$ на интервале $[t_e, T]$, также зависящую от параметра p_e. Необходимо, чтобы положительная плотность распределения моментов обращения заканчивалась точно в момент T, поэтому значение параметра p_e выбирается так, чтобы

$$p_e + \int\limits_{t_e}^{T} f(t)dt = 1. \tag{3.7}$$

В итоге вышеизложенного анализа получаем следующую теорему.

Теорема 3.1. *Любое симметричное равновесное по Нэшу распределение моментов обращения к системе обслуживания с двумя серверами со случайным доступом и потерями с функцией распределения $F(t)$ на интервале $[0, T]$ обладает следующими свойствами.*

1. *Существует строго положительная вероятность обращения в нулевой момент времени $p_e = F(0) > 0$.*

2. *На интервале $(0, t_e)$ игроки обращаются в систему с нулевой*

вероятностью, где

$$t_e = \left(\ln \frac{1-(1-p_e)^N - Np_e(1-p_e)^{N-1}}{1-\frac{2(1-(1-p_e)^{N+1})}{p_e(N+1)}+(1-p_e)^N}\right)/(\mu_1+\mu_2).$$

3. *Если для $p_e = 1$ решение уравнения (3.2) дает $t_e > T$, то равновесная стратегия – чистая стратегия, когда все игроки обращаются в систему в нулевой момент времени.*

4. *Иначе, $p_e < 1$ и на интервале $[t_e, T]$ существует непрерывная положительная плотность распределения моментов обращения в систему $f(t)$, определяемая выражением (3.5).*

5. *Равновесная вероятность обращения в нулевой момент времени находится из уравнения (3.7).*

6. $C(p_e) = 2\frac{1-(1-p_e)^{N+1}}{p_e(N+1)} - (1-p_e)^N$ *– вероятность получить обслуживание на всем носителе стратегии.*

Лемма 3.1. *Функция распределения $F(t)$, являющаяся решением (3.2) и (3.4) для заданного значения начального условия $F(0) = p$ в любой точке интервала $[0, T]$ возрастает по значению p.*

Доказательство. Пусть заданы две вероятности обращений в нулевой момент времени $0 < p < q \leq 1$, определяющие начальные условия для получения двух соответствующих функций распределения $F_p(t)$ и $F_q(t)$, являющихся решениями (3.2) и (3.4). Соответствующие вероятности обслуживания $C(p)$ и $C(q)$ постоянны на всем носителе распределения. По следствию 2.1 функция $C(\cdot)$ убывающая. Тогда вероятность потери для p должна быть меньше, чем для q, на всем носителе распределения.

По лемме 2.3 для соответствующих начал интервалов, на которых возобновляются обращения в систему, справедливо $t_q = t(q) < t_p = t(p)$. То есть функция $F_q(t)$ начинает возрастать от значения q в момент, когда $F_p(t)$ еще остается константой $p < q$. Для $t \in [0, t_p]$ утверждение леммы справедливо, так как здесь $F_p(t) = p < q \leq F_q(t)$.

Пусть есть некоторый момент $s > t_p$, такой, что $F_p(t) < F_q(t)$ и $F_p(s) = F_q(s)$. Тогда $f_p(s) > f_q(s)$, так как обе функции неубывающие по t, в точке s функция $F_p(t)$ должна пересечь $F_q(t)$ снизу вверх и, следовательно, угол наклона $F_p(t)$ больше угла наклона $F_q(t)$. Тогда $\frac{f_p(s)}{1-F_p(s)} > \frac{f_q(s)}{1-F_q(s)}$, что означает что интенсивность обращений в момент s больше для вероятности обращений в начальный момент p, чем для q при том, что интенсивности обслуживания в обоих случаях одинаковы. Тогда вероятность потери в момент s для p должна быть больше, чем для q, что противоречит тому, что на всем носителе распределения вероятность потери для p должна быть меньше, чем для q. □

Теорема 3.2. *Симметричное равновесное распределение поступлений F, определеное в теореме 3.1, существует и единственно.*

Доказательство. Единственность равновесия следует из леммы 3.1. Условие равновесия (3.7) представляет собой уравнение, левая часть которого монотонно возрастает по p_e. При p_e, близких к нулю, она равна вероятности обращения на интервале $[t_e, T]$, которая не превышает 1. При $p_e = 1$ значение левой части не меньше 1. Поэтому существует единственное решение p_e, которому соответствует единственное значение t_e и плотность распределения $f(t)$ на $[t_e, T]$. □

3.1. Вычислительные эксперименты

Для нахождения равновесий в рассматриваемой задаче при $N > 1$ используется численный алгоритм, являющийся комбинацией методов дихотомии решения уравнений с одним неизвестным и метода Эйлера решения системы обыкновенных дифференциальных уравнений первого порядка. Сначала проверяется, что $t_e(1) < T$, иначе равновесной является чистая стратегия, когда все игроки обращаются в систему в нулевой момент времени. Далее выбирается некоторое начальное значение p_e, для него находится t_e как решение уравнения (3.2). В точке t_e вычисляются начальные значения для систем (3.3) для N и $N-1$ игроков, а также $f(t_e)$. Для каждой следующей точки разбиения отрезка $[t_e, T]$ строится решение обеих систем по методу Эйлера: $p_{ijk}(t+\delta) \approx p_{ijk}(t)+\delta p'_{ijk}(t)$, используя соотношение (3.4) для вычисления $\lambda_i(t) = \frac{(N-i)f(t)}{1-F(t)}$, и по формуле (3.5) находится очередное $f(t)$. Затем вычисляется $F(T)$ и сравнивается с 1. Если получено равенство (с точностью до некоторого ϵ, близкого к нулю), алгоритм завершает работу. Если это значение больше 1, то p_e уменьшается, а если меньше, то увеличивается, и алгоритм повторяется для нового значения p_e.

Таблица 1. Равновесные характеристики

	$r = 0.1$			$r = 0.9$		
N	p_e	t_e	$C(p_e)$	p_e	t_e	$C(p_e)$
2	0.236	0.999	0.981	0.397	0.099	0.947
5	0.275	0.085	0.836	0.319	0.082	0.794
10	0.255	0.067	0.632	0.267	0.065	0.613
100	0.169	0.011	0.117	0.169	0.011	0.117
200	0.161	0.006	0.061	0.162	0.006	0.061
300	0.159	0.004	0.042	0.159	0.004	0.042

В табл. 1 приведены результаты вычисления равновесий для конкретных значений параметров системы. Рассматривается временной интервал работы системы $[0, 1]$. В системе первый сервер значительно уступает в производительности второму: $\mu_1 = 1, \mu_2 = 10$. Сравниваются равновесия для различного числа игроков в случаях высокой $(r = 0.1)$ и низкой $(r = 0.9)$ вероятностей попадания запроса на быстрый сервер. На рис. 1 и 2 приведены графики соответствующих равновесных плотностей распределения моментов обращения в систему. На примерах видно, что при увеличении числа игроков распределение моментов обращения в систему стремится к равномерному, причем одному и тому же для различных вероятностей попадания на быстрый сервер. В последних трех случаях для $N = 100, 200, 300$ равновесная плотность распределения моментов обращения в систему примерно равна 0.84 на всем интервале $[t_e, T]$. Рис. 3 показывает, как изменяется вид равновесных плотностей при изменении параметра r – вероятности, что обратившийся в систему игрок пренаправляется на первый сервер.

Сравним эффективность системы с рациональным случайным доступом, рассмотренной в данной работе, с двумя рассмотренными ранее. Первая – это система с одним сервером [16]. Вторая – модель со случайным доступом, где система, перенаправляющая запросы на сервера не имеет информации о занятости серверов [2]. Рассмотрим на примере проблему: что лучше – использовать два сервера, или один с производительностью, равной сумме производительностей

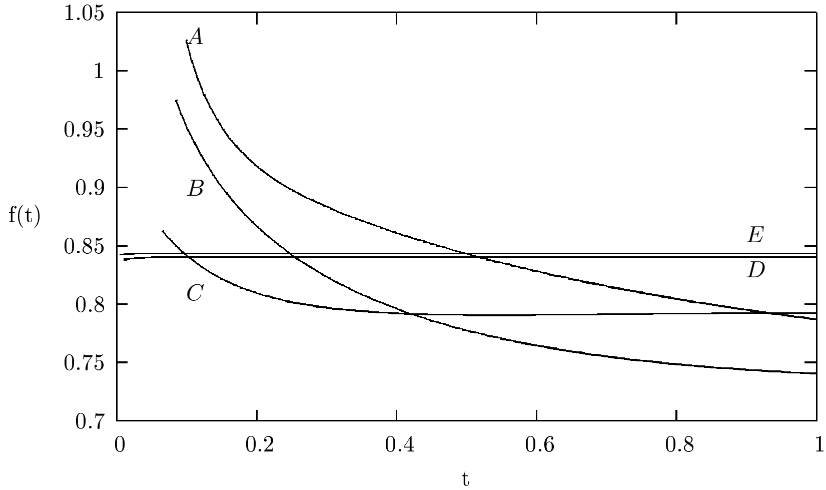

Рисунок 1. Равновесные плотности $f(t)$ при $T = 1, \mu_1 = 1, \mu_2 = 10, r = 0.1$. График A для $N = 2$, B для $N = 5$, C для $N = 10$, D для $N = 100$, E для $N = 200$.

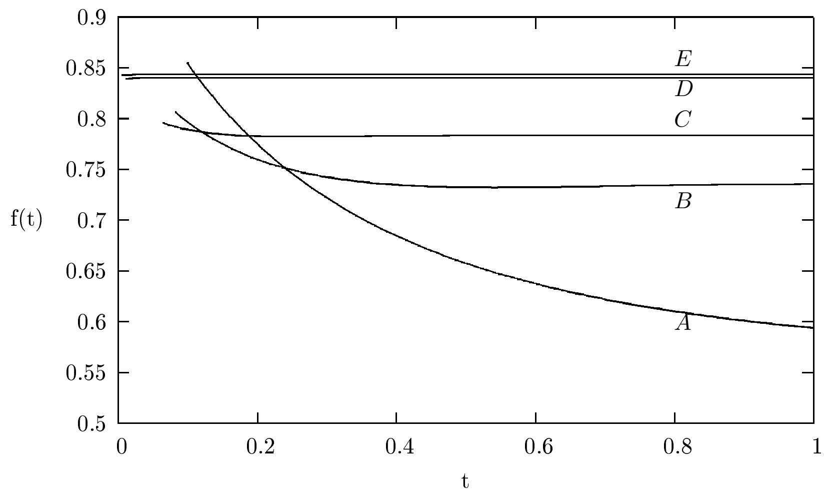

Рисунок 2. Равновесные плотности $f(t)$ при $T = 1, \mu_1 = 1, \mu_2 = 10, r = 0.9$. График A для $N = 2$, B для $N = 5$, C для $N = 10$, D для $N = 100$, E для $N = 200$.

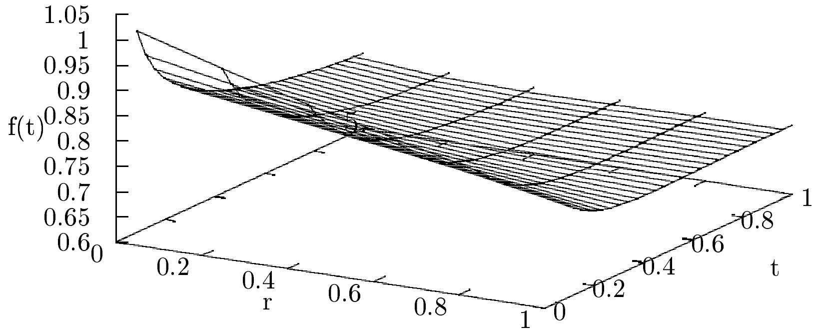

Рисунок 3. Как изменение r изменяет вид стратегии при $T = 1, \mu_1 = 1, \mu_2 = 10, N = 5$

двух серверов. Пусть в двухсерверной системе два сервера с интенсивностями обслуживания $\mu_1 = 1, \mu_2 = 10$. В односерверной системе $\mu = \mu_1 + \mu_2 = 11$. Рассматривается временной интервал работы системы $[0, 1]$. В качестве критерия эффективности системы возьмем равновесный выигрыш игрока, то есть равновесную вероятность получения обслуживания в системе, которая прямо пропорциональна ожидаемому числу обслуженных системой запросов. На рис. 4 сравнивается эффективность систем для различного числа игроков N. Кривые представленных графиков располагаются друг над другом, не пересекаясь, и соответствуют следующим случаям в порядке убывания эффективности сверху вниз:

1. система с двумя серверами с рациональным случайным доступом, высокая вероятность попадания на быстрый сервер $r = 0.1$;
2. система с двумя серверами с рациональным случайным доступом, низкая вероятность попадания на быстрый сервер $r = 0.9$;
3. система с двумя серверами со случайным доступом, высокая вероятность попадания на быстрый сервер $r = 0.1$;
4. система с одним сервером;
5. система с двумя серверами со случайным доступом, низкая вероятность попадания на быстрый сервер $r = 0.9$.

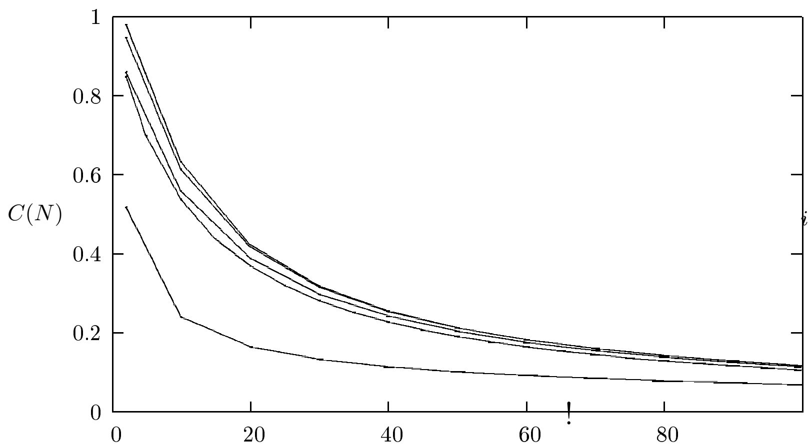

Рисунок 4. Изменение эффективности различных систем с изменением числа игроков.

4. Пуассоновское число игроков

Пусть теперь ни одному игроку не известно число его противников N, известно только, что N – случайная величина, имеющая распределение Пуассона с параметром λ. Пусть с вероятностью p_e каждый из противников обращается к системе в момент времени $t=0$. Случайная величина X_{p_e} – число игроков, обратившихся к серверу в нулевой момент времени. Заметим, что для каждого значения случайной величины N случайная величина X_{p_e} имеет биномиальное распределение $Bin(N, p_e)$. Тогда вероятность получения обслуживания в нулевой момент времени равна, согласно (2.1),

$$C(p) = e^{-\lambda}\sum_{N=0}^{\infty}\frac{\lambda^N}{N!}\cdot$$

$$\left((1-p_e)^N + C_N^1 p_e(1-p_e)^{N-1} + \sum_{=2}^{N} C_N^i p_e^i (1-p_e)^{N-i}\frac{2}{i+1}\right).$$

Сумма в скобках в правой части данного выражения равна правой части (3.1), откуда получаем

$$C(p_e) = e^{-\lambda}\sum_{N=0}^{\infty}\frac{\lambda^N}{N!}\left(2\frac{1-(1-p_e)^{N+1}}{p_e(N+1)} - (1-p_e)^N\right) = \qquad (4.1)$$

$$2\frac{e^{-\lambda}}{p_e\lambda}\left(\sum_{N=0}^{\infty}\frac{\lambda^{N+1}}{(N+1)!}-\sum_{N=0}^{\infty}\frac{\lambda^{N+1}(1-p_e)^{N+1}}{(N+1)!}\right)-e^{-\lambda}e^{\lambda(1-p_e)}=$$
$$2\frac{e^{-\lambda}}{p_e\lambda}\left(e^{\lambda}-e^{\lambda(1-p_e)}\right)-e^{-\lambda p_e}=2\frac{1-e^{-\lambda p_e}}{\lambda p_e}-e^{-\lambda p_e}.$$

Найдем вероятность получения обслуживания в момент $t_e > 0$ при условии, что вероятности обращений к системе на интервале $(0, t_e)$ равны нулю. Эта вероятность равна

$$1-P(X_{p_e}\geq 2)e^{-(\mu_1+\mu_2)t_e}=$$
$$1-(1-P(X_{p_e}=0)-P(X_{p_e}=1))e^{-(\mu_1+\mu_2)t_e}=$$
$$1-(1-e^{-\lambda}\sum_{N=0}^{\infty}\frac{\lambda^N}{N!}(1-p_e)^N-e^{-\lambda}\sum_{N=0}^{\infty}\frac{\lambda^N}{N!}Np(1-p_e)^{N-1})e^{-(\mu_1+\mu_2)t_e}=$$
$$1-(1-e^{-\lambda p_e}-e^{-\lambda}\lambda p_e\sum_{N=1}^{\infty}\frac{\lambda^{N-1}}{(N-1)!}(1-p)^{N-1})e^{-(\mu_1+\mu_2)t_e}=$$
$$1-(1-e^{-\lambda p_e}-e^{-\lambda}\lambda p_e\sum_{N=0}^{\infty}\frac{\lambda^N}{N!}(1-p_e)^N)e^{-(\mu_1+\mu_2)t_e}=$$
$$1-(1-e^{-\lambda p_e}-e^{-\lambda}\lambda p_e e^{\lambda(1-p_e)})e^{-(\mu_1+\mu_2)t}=$$
$$1-(1-e^{-\lambda p_e}-\lambda p_e e^{-\lambda p_e})e^{-(\mu_1+\mu_2)t_e}.$$

Тогда уравнение (2.2) для случая фиксированного числа игроков имеет вид

$$2\frac{1-e^{-\lambda p_e}}{\lambda p_e}-e^{-\lambda p_e}=1-(1-e^{-\lambda p_e}-\lambda p_e e^{-\lambda p_e})e^{-(\mu_1+\mu_2)t_e}. \qquad (4.2)$$

Если $t_e(1)\geq T$, то равновесной является чистая стратегия – обращение в систему в момент $t=0$ с вероятностью 1. Далее будем считать, что $t_e(1)<T$.

Необходимо найти равновесную плотность $f(t)$ распределения моментов обращения к системе на интервале $[t_e, T]$. Определим Марковский процесс с состояниями системы (i, j) в каждый момент времени $t\in[t_e, T]$, где $i, j\in\{0,1\}$ – состояния занятости первого и второго сервера соответственно (0 – свободен, 1 – занят). Интенсивность поступлений запросов в систему в каждый момент времени t равна $\lambda f(t)$. Система Колмогорова для вероятностей состояний системы p_{ij} выглядит следующим образом.

$$\begin{aligned}
p'_{00}(t)&=-\lambda f(t)p_{00}(t)+\mu_1 p_{10}(t)+\mu_2 p_{01i}(t)\\
p'_{10}(t)&=r\lambda f(t)p_{00}(t)-(\overline{r}\lambda f(t)+\mu_1)p_{10}(t)+\mu_2 p_{11}(t)\\
p'_{01}(t)&=\overline{r}\lambda f(t)p_{00}(t)-(r\lambda f(t)+\mu_2)p_{01}(t)+\mu_1 p_{11}(t)\\
p'_{11}(t)&=\lambda f(t)(p_{01}(t)+p_{10}(t))-(\mu_1+\mu_2)p_{11}(t).
\end{aligned} \qquad (4.3)$$

Вероятность получить обслуживание в момент $t \in [t_e, T]$ в равновесии постоянна для всех $t \in [t_e, T]$ и равна $p_{00}(t) + p_{01}(t) + p_{10}(t) = 1 - p_{11}(t) = C(p_e)$. Тогда сумма соответствующих производных должна быть равна нулю. Отсюда, подставив в суммы производные вероятностей состояний из системы Колмогорова (4.3), получаем равновесную плотность

$$f(t) = \frac{(\mu_1 + \mu_2)(1 - C(p_e))}{\lambda(p_{01}(t) + p_{10}(t))}. \tag{4.4}$$

для $t \in [t_e, T]$.

Заметим, что в случае с одним сервером [16] равновесная стратегия на интервале $[t_e, T]$ представляет собой равномерное распределение. В нашем более общем случае это не так.

Определим для системы начальные условия – вероятности состояний в момент t_e, при условии, что на $(0, t_e)$ нет поступлений. Для каждого состояния (i, j) соответствующую вероятность можно найти как

$$p_{ij}(t_e) = e^{-\lambda} \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} \sum_{k=0}^{N} p_{ijk}^N(t_e),$$

где $p_{ijk}^N(t_e)$ – вероятность состояния (i, j, k) в момент t_e для случая N игроков. Тогда

$$\begin{aligned} &p_{00}(t_e) = e^{-\lambda} \sum_{N=0}^{\infty} \tfrac{\lambda^N}{N!} \cdot \\ &\quad \sum_{k=0}^{N} C_N^k p_e^k (1 - p_e)^{N-k} \left[\mathbb{1}_{k=0} + \mathbb{1}_{k=1} \left(r(1 - e^{-\mu_1 t_e}) + \overline{r}(1 - e^{-\mu_2 t_e}) \right) + \right. \\ &\quad \left. \mathbb{1}_{k>1}(1 - e^{-\mu_1 t_e})(1 - e^{-\mu_2 t_e}) \right]. \end{aligned}$$

Преобразуем внутреннюю сумму

$$\begin{aligned} &\sum_{k=0}^{N} C_N^k p_e^k (1 - p_e)^{N-k} \left[\mathbb{1}_{k=0} + \mathbb{1}_{k=1} \left(r(1 - e^{-\mu_1 t_e}) + \overline{r}(1 - e^{-\mu_2 t_e}) \right) + \right. \\ &\qquad \left. \mathbb{1}_{k>1}(1 - e^{-\mu_1 t_e})(1 - e^{-\mu_2 t_e}) \right] = \\ &(1 - p_e)^N + N p_e (1 - p_e)^{N-1} (r(1 - e^{-\mu_1 t_e}) + \overline{r}(1 - e^{-\mu_2 t_e})) + \\ &\left(1 - (1 - p_e)^N - N p_e (1 - p_e)^{N-1} \right) (1 - e^{-\mu_1 t_e})(1 - e^{-\mu_2 t_e}). \end{aligned}$$

Тогда

$$\begin{aligned} p_{00}(t_e) = e^{-\lambda p_e} + \lambda p_e e^{-\lambda p_e} (r(1 - e^{-\mu_1 t_e}) + \overline{r}(1 - e^{-\mu_2 t_e})) + \\ \left(1 - e^{-\lambda p_e} - \lambda p_e e^{-\lambda p_e} \right) (1 - e^{-\mu_1 t_e})(1 - e^{-\mu_2 t_e}). \end{aligned}$$

Аналогично получаем выражение для остальных вероятностей состояний в момент t_e.

$$
\begin{aligned}
&p_{00}(t_e) = e^{-\lambda p_e} + \lambda p e^{-\lambda p_e}(r(1-e^{-\mu_1 t_e}) + \overline{r}(1-e^{-\mu_2 t_e}))+\\
&\quad \left(1-e^{-\lambda p_e} - \lambda p_e e^{-\lambda p_e}\right)(1-e^{-\mu_1 t_e})(1-e^{-\mu_2 t_e})\\
&p_{10}(t_e) = \lambda p_e e^{-\lambda p_e} r e^{-\mu_1 t_e} +\\
&\quad \left(1-e^{-\lambda p_e} - \lambda p_e e^{-\lambda p_e}\right) e^{-\mu_1 t_e}(1-e^{-\mu_2 t_e})\\
&p_{01}(t_e) = \lambda p_e e^{-\lambda p_e} \overline{r} e^{-\mu_2 t_e} +\\
&\quad \left(1-e^{-\lambda p_e} - \lambda p_e e^{-\lambda p_e}\right)(1-e^{-\mu_1 t_e}) e^{-\mu_2 t_e}\\
&p_{11}(t_e) = \left(1-e^{-\lambda p_e} - \lambda p_e e^{-\lambda p_e}\right) e^{-\mu_1 t_e} e^{-\mu_2 t_e}.
\end{aligned}
\tag{4.5}
$$

Таким образом, имеется задача Коши с системой диференциальных уравнений (4.3) и начальными условиями (4.5). Ее решение дает вероятности состояний $p_{ij}(t)$, зависящие от параметра p_e. С известными вероятностями состояний выражение (4.4) определяет равновесную плотность распределения моментов обращения $f(t)$ на интервале $[t_e, T]$, также зависящую от параметра p_e. Необходимо, чтобы положительная плотность распределения моментов обращения заканчивалась точно в момент T, поэтому значение параметра p_e выбирается так, чтобы

$$p_e + \int_{t_e}^{T} f(t)dt = 1. \tag{4.6}$$

В итоге вышеизложенного анализа получаем следующую теорему.

Теорема 4.1. *Любое симметричное равновесное по Нэшу распределение моментов обращения к системе обслуживания с двумя серверами со случайным доступом и потерями с функцией распределения $F(t)$ на интервале $[0, T]$ обладает следующими свойствами.*

1. *Существует строго положительная вероятность обращения в нулевой момент времени $p_e = F(0) > 0$.*

2. *На интервале $(0, t_e)$ игроки обращаются в систему с нулевой вероятностью, где*

$$t_e = \left(\ln \frac{1-e^{-\lambda p_e} - \lambda p_e e^{-\lambda p_e}}{1 - 2\frac{1-e^{-\lambda p_e}}{\lambda p_e} + e^{-\lambda p_e}}\right) / (\mu_1 + \mu_2).$$

3. *Если для* $p_e = 1$ *решение уравнения (4.2) дает* $t_e > T$, *то равновесная стратегия – чистая стратегия, когда все игроки обращаются в систему в нулевой момент времени.*

4. *Иначе,* $p_e < 1$ *и на интервале* $[t_e, T]$ *существует непрерывная положительная плотность распределения моментов обращения в систему* $f(t)$, *определяемая выражением (4.4).*

5. *Равновесная вероятность обращения в нулевой момент времени находится из уравнения (4.6).*

6. $C(p_e) = 2\frac{1-e^{-\lambda p_e}}{\lambda p_e} - e^{-\lambda p_e}$ *– вероятность получить обслуживание на всем носителе стратегии.*

Лемма 4.1. *Функция распределения* $F(t)$, *являющаяся решением (4.2) и (4.4) для заданного значения начального условия* $F(0) = p$ *в любой точке интервала* $[0, T]$ *возрастает по значению* p.

Доказательство. Доказательство полностью аналогично доказательству леммы 3.1 для случая фиксированного числа игроков. □

Тогда, как и в случае для фиксированного числа игроков, справедлива следующая теорема.

Теорема 4.2. *Симметричное равновесное распределение поступлений* F, *определеное в теореме 4.1, существует и единственно.*

4.1. Вычислительные эксперименты

Для нахождения равновесий в рассматриваемой задаче при $\lambda > 0$ используется численный алгоритм, являющийся комбинацией методов дихотомии решения уравнений с одним неизвестным и метода Эйлера решения системы обыкновенных дифференциальных уравнений первого порядка. Сначала проверяется, что $t_e(1) < T$, иначе равновесной является чистая стратегия, когда все игроки обращаются в систему в нулевой момент времени. Далее выбирается некоторое начальное значение p_e, для него находится t_e как решение уравнения (4.2). В точке t_e вычисляются начальные значения для системы (4.3) и $f(t_e)$. Для каждой следующей точки разбиения отрезка $[t_e, T]$ строится решение обеих систем по методу Эйлера: $p_{ij}(t + \delta) \approx p_{ij}(t) + \delta p'_{ij}(t)$, и по формуле (4.4) находится очередное $f(t)$. Затем вычисляется $F(T)$ и сравнивается с 1. Если получено равенство (с точностью до некоторого ϵ, близкого к нулю), алгоритм завершает работу. Если это значение больше 1, то p_e уменьшается, а если меньше, то увеличивается, и алгоритм повторяется для нового значения p_e.

Таблица 2. Равновесные характеристики

	$r = 0.1$			$r = 0.9$		
λ	p_e	t_e	$C(p_e)$	p_e	t_e	$C(p_e)$
1	0.251	0.096	0.99	0.374	0.094	0.98
5	0.268	0.08	0.84	0.311	0.077	0.803
20	0.221	0.045	0.433	0.226	0.045	0.426
100	0.169	0.114	0.118	0.170	0.011	0.118
200	0.162	0.006	0.062	0.162	0.006	0.062
300	0.159	0.004	0.042	0.159	0.004	0.042

В табл. 2 приведены результаты вычисления равновесий для конкретных значений параметров системы. Рассматривается временной интервал работы системы $[0, 1]$. В системе первый сервер значительно уступает в производительности второму: $\mu_1 = 1, \mu_2 = 10$. Сравниваются равновесия для различного числа игроков в случаях высокой ($r = 0.1$) и низкой ($r = 0.9$) вероятностей попадания запроса на быстрый сервер. На рис. 5 и 6 приведены графики соответствующих равновесных плотностей распределения моментов обращения в систему. На примерах видно, что, как и в случае фиксированного числа игроков, при увеличении значения λ распределение моментов обращения в систему стремится к равномерному, причем одному и тому же для различных вероятностей попадания на быстрый сервер. В последних трех случаях для $\lambda = 100, 200, 300$ равновесная плотность распределения моментов обращения в систему примерно равна 0.84 на всем интервале $[t_e, T]$. Можно также заметить идентичность данных результатов для обоих случаев – фиксированного и случайного числа игроков. Рис. 7 показывает, как изменяется вид равновесных плотностей при изменении параметра r – вероятности, что обратившийся в систему игрок пренаправляется на первый сервер.

Сравним, как и в случае фиксированного числа игроков, эффективность системы с рациональным случайным доступом, рассмотренной в данной работе, с двумя рассмотренными ранее. Первая –

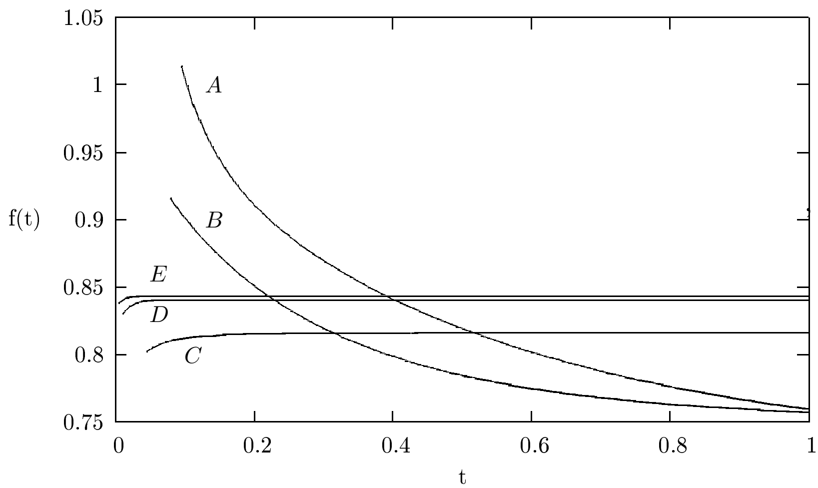

Рисунок 5. Равновесные плотности $f(t)$ при $T = 1, \mu_1 = 1, \mu_2 = 10, r = 0.1$. График A для $\lambda = 1$, B для $\lambda = 5$, C для $\lambda = 20$, D для $\lambda = 100$, E для $\lambda = 200$.

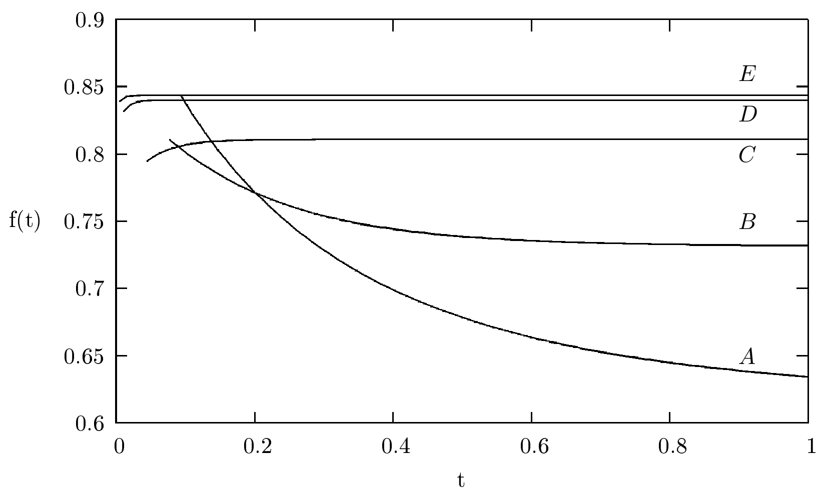

Рисунок 6. Равновесные плотности $f(t)$ при $T = 1, \mu_1 = 1, \mu_2 = 10, r = 0.9$. График A для $\lambda = 1$, B для $\lambda = 5$, C для $\lambda = 20$, D для $\lambda = 100$, E для $\lambda = 200$.

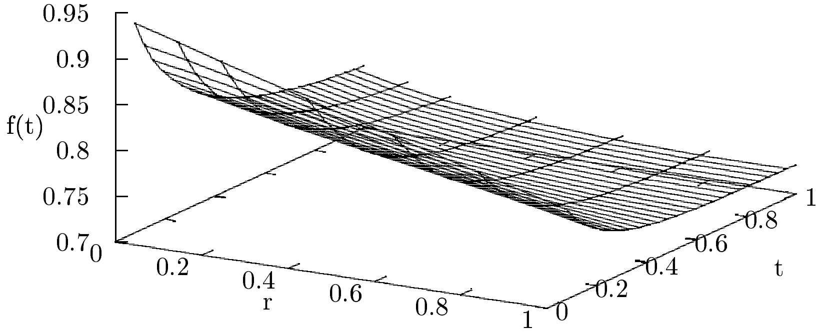

Рисунок 7. Как изменение r меняет вид стратегии при $T = 1, \mu_1 = 1, \mu_2 = 10, \lambda = 5$.

имеет информации о занятости серверов [2]. Рассмотрим на примере проблему: что лучше – использовать два сервера, или один с производительностью, равной сумме производительностей двух серверов. Пусть в двухсерверной системе два сервера с интенсивностями обслуживания $\mu_1 = 1, \mu_2 = 10$. В односерверной системе $\mu = \mu_1 + \mu_2 = 11$. Рассматривается временной интервал работы системы $[0, 1]$. В качестве критерия эффективности системы возьмем равновесную вероятность получения обслуживания в системе, которая прямо пропорциональна ожидаемому числу обслуженных системой запросов. На рис. 8 сравнивается эффективность систем для различных величин λ. Как и в случае фиксированного числа игроков, кривые представленных графиков располагаются друг над другом, не пересекаясь, и соответствуют следующим случаям в порядке убывания эффективности сверху вниз:

1. система с двумя серверами с рациональным случайным доступом, высокая вероятность попадания на быстрый сервер $r = 0.1$;
2. система с двумя серверами с рациональным случайным доступом, низкая вероятность попадания на быстрый сервер $r = 0.9$;
3. система с двумя серверами со случайным доступом, высокая вероятность попадания на быстрый сервер $r = 0.1$;
4. система с одним сервером;

5. система с двумя серверами со случайным доступом, низкая вероятность попадания на быстрый сервер $r = 0.9$.

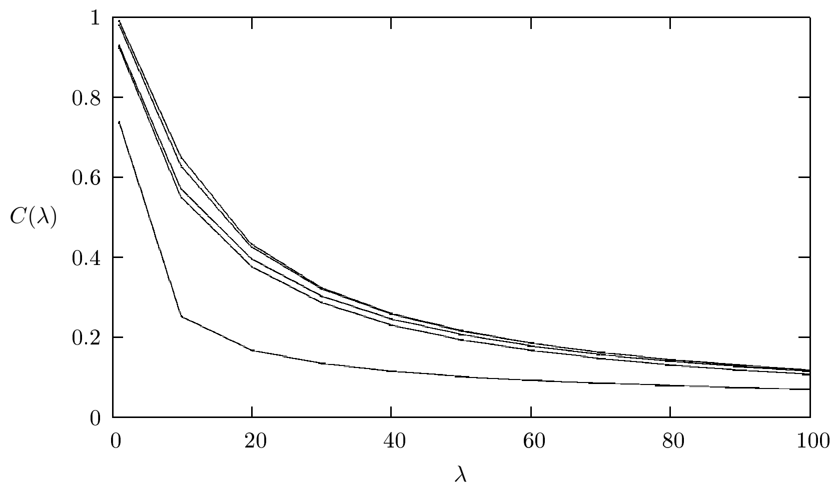

Рисунок 8. Изменение эффективности различных систем с изменением λ.

5. Заключение

Для двухсерверной системы обслуживания с потерями и рациональным случайным доступом, принимающей запросы на интервале времени $[0, T]$ исследованы две модели. В первой число игроков фиксировано, во второй является случайной величиной, имеющей распределение Пуассона. Для обоих случаев доказано существование единственного симметричного равновесия, такого что с некоторой положительной вероятностью p_e пользователи обращаются к системе в нулевой момент времени, и далее существует интервал времени $[t_e, T]$, на котором определена положительная плотность распределения моментов обращения в систему. Для обеих моделей предложены алгоритмы для численного нахождения равновесий. Данные алгоритмы положены в основу реализации программной системы, позволяющей графически представлять равновесные стратегии. Проведены численные эксперименты по сравнению равновесий при различных значениях параметров рассмотренной в данной работе модели, а также сравнению ее эффективности с односерверной моделью [16] и моделью со случайным доступом, в которой система, перенаправляющая запросы на сервера, не имеет информации об их занятости [2].

Возможны следующие пути развития данной модели. Во-первых, результат легко может быть обобщен на произвольное количество серверов. Результаты качественно останутся такими же, увеличится число переменных и уравнений. Численные эксперименты для числа серверов больше двух могут дать интересные результаты. Например, таким образом можно решать прикладные задачи определения оптимального выбора вероятностей доступа к серверам r с целью максимизировать производительность системы, то есть ожидаемое число обслуженных системой запросов. Также возможно ввести в обе модели функцию комфортности [1] для исследования ее влияния на вид равновесия.

СПИСОК ЛИТЕРАТУРЫ

1. Мазалов В.В., Чуйко Ю.В. *Некооперативное равновесие по Нэшу в задаче выбора оптимального момента обращения к системе обслуживания* // Вычислительные технологии. 2006. Т. 11. N. 6. С. 60–71.

2. Чиркова Ю.В. *Оптимальные обращения к 2-серверной системе с потерями и случайным доступом* // Математическая Теория Игр и ее Приложения. 2015. Т. 7. Вып. 3. С. 79–111.

3. Altman E. *Applications of dynamic games in queues* // Advances in Dynamic Games. 2005. V. 7. P. 309–342.

4. Altman E. *A Markov game approach for optimal routing into a queueing network* // INRIA report N 2178, 1994.

5. Altman E., Hassin R. *Non-Threshold Equilibrium for Customers Joining an M/G/1 Queue* // In Proceedings of 10th International Symposium on Dynamic Game and Applications, 2002.

6. Altman E., Jimenez T., Nunez Queija R. and Yechiali U. *Optimal routing among* $\cdot/M/1$ *queues with partial information* // Stochastic Models. 2004. V. 20. N 2. P. 149–172.

7. Altman E., Koole G. *Stochastic scheduling games with Markov decision arrival processes* // Journal Computers and Mathematics with Appl. 1993. V. 26. N 6. P. 141–148

8. Altman E., Shimkin N. *Individually Optimal Dynamic Routing in a Processor Sharing System* // Operations Research. 1998. P. 776–784.

9. Glazer A., Hassin R. *Equilibrium arrivals in queues with bulk service at scheduled times* // Transportation Science. 1987. V. 21. N 4. P. 273–278.

10. Glazer A., Hassin R. $?/M/1$*: On the equilibrium distribution of customer arrivals* // European Journal of Operational Research. 1983. V. 13. N 2. P. 146–150.

11. Johnson O., Goldschmidt C. *Preservation of log-concavity on summation* // ESAIM: Probability and Statistics. 2006. V. 10. P. 206–215.

12. Killelea P. *Web Performance Tuning: Speeding Up the Web*, O'Reilly Media, Inc., 2002.

13. Kopper K. *The Linux Enterprise Cluster: Build a Highly Available Cluster with Commodity Hardware and Free Software*, No Starch Press, 2005.

14. Ou Z., Zhuang H., Lukyanenko A., Nurminen J., Hui P., Mazalov V., Yla-Jaaski A. *Is the Same Instance Type Created Equal? Exploiting Heterogeneity of Public Clouds* // IEEE Transactions on Cloud Computing. 2013. V. PP. Issue 99.

15. Ravner L., Haviv M. *Strategic timing of arrivals to a finite queue multi-server loss system* // Queueing Systems. 2015. V. 81. N 1. P. 71–96.

16. Ravner L., Haviv M. *Equilibrium and socially optimal arrivals to a single server loss system* // In International Conference on NETwork Games COntrol and OPtimization 2014 (NetGCoop'14), Trento, Italy, October 2014.

17. Shaked M., Shanthikumar JG. *Stochastic Orders, Springer Series In Statistics*, Springer, New York, 2007.

OPTIMAL ARRIVALS TO A TWO-SERVER LOSS SYSTEM WITH A RATIONAL RANDOM ACCESS

Julia V. Chirkova, IAMR KarRC RAS, Cand.Sc. (julia@krc.karelia.ru).

Abstract: We consider the 2-server queueing loss-type system which admits requests during a time interval $[0, T]$. Players try to send their requests into the system, which randomly redirects each request to one of its free servers with some probabilities, or to unique free server, or refuses the request. We consider a non-cooperative game for this queuing system. Each player's strategy is a time moment to send his request to the system trying to maximize the probability of successful service obtaining. We use a symmetric Nash equilibrium as an optimality criteria. Two models are considered for this game. In the first model the number of players is deterministic. In the second it follows a Poisson distribution. We prove that there exists a unique symmetric equilibrium for both models. We compare numerically equilibria for different model parameters of the model. Also we compare an efficiency for the one-server model and the two-server model with a random access where the system has no information about servers' being busy.

Keywords: queueing system, optimal arrivals, Nash equilibrium.