

УДК 519.25, 004.75, 519.216.5

ББК 22.18

ЭФФЕКТИВНАЯ ПО ЦЕНЕ СТРАТЕГИЯ АРЕНДЫ ОБЛАЧНЫХ РЕСУРСОВ ПРИ НЕОПРЕДЕЛЕННОСТИ ЦЕНЫ

ЕВГЕНИЙ Е. ИВАШКО

АННА А. ИВАШКО

Институт прикладных математических исследований

Карельский научный центр РАН

185910, Петрозаводск, ул. Пушкинская, 11

Петрозаводский государственный университет

185910, Петрозаводск, пр. Ленина, 33

e-mail: ivashko@krc.karelia.ru, aivashko@krc.karelia.ru

ГЕОРГИЙ Р. САФОНОВ

Петрозаводский государственный университет

185910, Петрозаводск, пр. Ленина, 33

e-mail: jiri.safonov@gmail.com

АНДРЕЙ ЧЕРНЫХ*

Центр исследований и высшего образования

Энсенада, Байя Калифорния, Мексика

e-mail: chernykh@cicese.mx

В работе рассматривается задача оптимизации ставки на облачный спотовый ресурс. Предложена стохастическая оптимизационная модель для минимизации ожидаемой стоимости

©2019 Е.Е. Ивашко, А.А. Ивашко, Г.Р. Сафонов, А. Черных

* Работа частично поддержана CONACYT (Consejo Nacional de Ciencia y Tecnologia, Mexico), grant no. 178415

аренды ресурса в условиях неопределенности цены. Модель основана на хорошо известной задаче наилучшего выбора с полной информацией. На основе предложенной модели представлена стратегия ставок на облачный спотовый ресурс. Стратегия позволяет минимизировать ожидаемую стоимость спотового инстанса в заданный период времени с гарантированным обслуживанием. Результаты моделирования на основе реальных цен Amazon демонстрируют преимущества предложенной стратегии.

Ключевые слова: облачные вычисления, спотовый инстанс, задача наилучшего выбора с полной информацией, Amazon EC2.

Поступила в редакцию: 20.04.18 *После доработки:* 25.07.19 *Принята к публикации:* 30.09.19

1. Введение

Облачные вычисления, предоставляющие доступ по требованию к компьютерным ресурсам (серверам, ресурсам хранения, приложениям и сервисам), стали критичной частью ИТ-отрасли [17]. Облачные провайдеры меняют модели доступа и монетизации предоставляемых ресурсов. Новые бизнес-подходы требуют изучения и выработки рационального (или оптимального с точки зрения каких-то критериев) поведения как для облачных провайдеров, так и для их клиентов.

Одним из основных «законодателей мод» на рынке облачных вычислений является компания Amazon. Платформа Elastic Compute Cloud (Amazon EC2) предоставляет различные варианты аренды виртуальных серверов [4]. В дополнение к инстансам по требованию (On-demand Instances) и резервным (Reserved Instances), в 2009 г. Amazon ввела спотовые инстансы (Spot Instances, SIs) [3]. Предложенный вариант аренды стал новым подходом к ценообразованию и новым способом получить вычислительные ресурсы по сниженной цене.

Используя SI, пользователи могут делать ставки на неиспользуемые мощности Amazon EC2. Текущая спотовая цена периодически меняется, отражая текущий спрос и предложение. SI предназначены для неинтерактивных приложений с гибким временем старта, которые могут быть легко поставлены на паузу. Примерами таких приложений являются видео-рендеринг/обработка, анализ данных, научное моделирование и т.п. В сравнении с ценами по требованию, аукционы SI позволяют пользователям сэкономить до 50-90%.

Мы рассматриваем следующую общую задачу. Облачный провайдер использует аукцион для продажи вычислительных инстансов. В дискретные моменты времени он генерирует спотовую цену и предоставляет компьютерные инстансы всем пользователям, чьи ставки превышают ее. Пользователи, получившие компьютерные инстансы, платят за них одинаковую стоимость, которая равна минимальной выигравшей ставке. Цель заключается в выработке такой стратегии, которая позволит гарантированно получать компьютерный инстанс в заданный период времени, минимизируя ожидаемую цену.

Мы сводим описанную выше задачу к задаче наилучшего выбора с полной информацией; решением является последовательность порогов. Значения порогов зависят от распределения спотовых цен и оставшегося периода, за который необходимо получить инстанс. Такая стратегия гарантирует пользователю минимизацию ожидаемой стоимости инстанса.

Структура статьи имеет следующий вид. В разделе 2 представлен обзор литературы, связанной с изучением аукционов на облачные ресурсы. В разделе 3 кратко описан механизм спотовых аукционов. В разделе 4 представлена математическая модель аукционов SI, предложена стратегия минимизации ожидаемой стоимости инстанса. Мы выполняем проверку предложенной модели и стратегии с помощью численного моделирования в разделе 5. Раздел 6 демонстрирует пример выработки стратегии на основе реальных данных аукционов Amazon на SI. Наконец, в разделе 7 представлено заключение, в котором кратко обсуждаются результаты и направления дальнейших работ.

2. Обзор работ

Облачные вычисления стали жизненно важной частью современной отрасли информационных технологий. Они получили также много внимания со стороны научного сообщества. Li и др., 2014 [16] представили обзор исследований по спотовому рынку облачных вычислений. Более свежий и расширенный обзор сделали Kumar и др., 2018 [14]. Авторы работ демонстрируют два основных направления исследований: изучение существующих механизмов ценообразования, основанных на реальных примерах политики облачных провайдеров, и предложение новых механизмов ценообразования и предостав-

ления ресурсов. В исследованиях подчеркиваются различия моделей, описывающих распределения цен. Помимо вероятностных моделей, исследователи используют марковские или полу-марковские модели, разрабатывают матрицу вероятностей изменения цен (Price Transition Probability Matrix). Вероятностные модели основаны на том, что спотовые цены являются независимыми одинаково распределенными случайными величинами, имеющими нормальное распределение или комбинацию нормальных распределений. Среди этих моделей также существуют различные подходы: статистические, экономические и т.д. В представленной работе предполагается, что спотовые цены являются независимыми одинаково распределенными случайными величинами с некоторым известным законом распределения.

Один из важных исследовательских вопросов – это наиболее выгодная для облачного провайдера политика ценообразования.

Abhishek и др., 2012 [1] обсуждают выбор между фиксированными и определяемыми рынками ценами. Авторы показывают, что динамическое ценообразование – это более гибкий, однако более сложный подход. На основе теории очередей предложены модели спотового и с оплатой по запуску рынков. Данные модели используются для обеспечения теоретического анализа и моделирования результатов, которые показывают, что изолированный рынок с оплатой по запуску более выгодный для облачного провайдера, чем смешанный спотовый рынок. Однако такой вывод верен только для провайдера-монополиста.

Xu и Li, 2012 [30] описывают математическую модель, направленную на максимизацию дохода облачного провайдера. Авторы рассматривают сценарий, в котором облачный провайдер влияет на спотовый спрос, изменяя цены на сервис. Спрос характеризуется стохастическими процессами прихода и ухода с лежащим в основе пуассоновским распределением с параметрами, зависящими от цены. Система рассматривается как марковский процесс рождения-гибели с непрерывным временем. Авторы выводят оптимальную политику облачного провайдера, которая максимизирует доход на бесконечном интервале времени. Необходимо отметить, что при использовании указанной модели в некоторых случаях динамическое ценообразова-

ние оказывается более выгодным.

Song и др., 2012 [22] разрабатывают стратегию ставок для брокеров облачного сервиса. Предложен специальный динамический алгоритм ставок, учитывающий доход. Авторы демонстрируют, что алгоритм дает почти оптимальную производительность; он также может использоваться с другими целевыми функциями брокеров облачного сервиса.

Как отмечалось ранее, Amazon представил свои SI в 2009 г. и продает право их аренды на аукционах, где пользователи делают ставки на желаемые ресурсы. Как анонсировалось, реальные цены зависят от общего объема доступных ресурсов и спроса. SI привлекают большое внимание исследователей как новая модель ценообразования. Один из наиболее интересных вопросов аукционов на SI – это парадигма динамического ценообразования. Ее характеристика является ключевым исследовательским вопросом с точки зрения разработки эффективных стохастических алгоритмов планирования и устойчивых к ошибкам механизмов.

Ben-Yehuda и др., 2011 [5] анализируют реальные цены, основываясь на исторических данных аукционов Amazon на SI. Авторы доказывают, что спотовые цены обычно определяются не рынком, а генерируются компанией Amazon случайным образом из узкого интервала через динамическую скрытую резервную цену. Отмечается, что цены на различные типы инстансов показывают схожее поведение. Такое заключение сделано на основе анализа спотовых цен Amazon EC2 первой половины 2010 г.

Cheng и др., 2016 [7] изучают разницу в ценах между Западным и Восточным рынками Amazon EC2, а также динамику ценообразования.

Javadi и др., 2013 [9] показывают, что динамика спотовых цен каждого SI характеризуется смесью нормальных распределений с тремя или четырьмя компонентами. Авторы выполняют полноценный анализ SI, основываясь на годичной истории цен в четырех центрах обработки данных Amazon EC2. Анализируются все типы SI и динамика ценообразования по часам и по дням недели. Отмечается бимодальность функций распределения спотовых цен и предлагается для моделирования спотовых цен использовать смешанное нормаль-

ное распределение с тремя или четырьмя компонентами. Несколько методов используется для оценки параметров распределений. Предложенная модель проверяется с помощью расширенного численного моделирования.

Wallace и др., 2013 [28] используют нейронные сети для предсказания спотовых цен. При выполнении экспериментов авторы использовали данные за период с декабря 2009 г. по июнь 2010 г. Предсказание выполнялось на основе стандартной модели многослойного перцептрона с обратным распространением ошибки и адаптивным обучением. Выполненные на реальных данных Amazon EC2 численные эксперименты показали хорошие результаты даже для случаев внезапного изменения цены. Результаты работы могут использоваться пользователями для выбора наилучшей стратегии ставок.

Временные ряды являются также популярным инструментом анализа спотовых цен. Chhetri и др., 2018 [8] выполняют декомпозицию истории спотовых цен в компоненты временных рядов; каждый компонент может показывать детерминированные и недетерминированные свойства; затем выполняется прогноз отдельно для каждого компонента, используя различные методы и период прогнозирования; наконец, индивидуальные прогнозы объединяются в единый прогноз спотовых цен. Alkharif и др., 2018 [2] используют модель LSTM для прогнозирования временного ряда для спотовых цен EC2.

Информация по динамике спотовых цен используется в разработке стратегии ставок, например, для минимизации стоимости или повышения надежности.

Karunakaran и др., 2015 [12] сравнивают четыре простые стратегии ставок: (1) ставка близко к стоимости зарезервированного инстанса; (2) ставка выше средней спотовой цены, получаемой из истории; (3) ставка близко к стоимости инстанса по требованию; и (4) ставка выше стоимости инстанса по требованию. Стратегии рассматриваются с точки зрения нескольких метрик: (1) стоимость завершения расчета задачи; (2) время ожидания; (3) прерывания при выполнении расчета задачи. Авторы приходят к выводу, что первая стратегия хорошо подходит для пользователей, для которых цена более важна, чем время завершения расчета задачи, и имеющих возможность делать частые сохранения. Вторая и третья стратегии показы-

вают хороший компромисс между стоимостью, временем ожидания и числом прерываний. Наконец, четвертая стратегия демонстрирует значительное увеличение цены без существенного снижения времени выполнения и числа прерываний. Таким образом, ставка, превышающая стоимость инстанса по требованию бесполезна.

Voorsluys, 2014 [27] предлагает подход для запуска ограниченных по времени вычислительных задач на пуле вычислительных ресурсов, составленном из SI низкой стоимости. Автор строит систему управления ресурсами и планирования, которая называется SpotRMS. Политика планирования использует оценки времени завершения задачи для определения как лучшего типа инстанса для запуска задачи, так и времени запуска задачи для удовлетворения ограничениям по времени. В частности, автор изучает методы запуска вычислительных заданий на прерывающихся SI. Такие методы включают в себя использование различных стратегий ставок и методы обработки ошибок для обхода незапланированной недоступности SI. Предложенный подход проверяется на моделировании по трейсам данных Amazon и демонстрирует свою эффективность. Petcu, 2014 [20] описывает решение по объединению ресурсов нескольких облачных систем.

Menache и др., 2014 [18] представляют самообучающийся алгоритм оптимизации стоимости, основанный на спотовых инстансах и инстансах по требованию. Авторы предлагают алгоритм выделения ресурсов для построения компромисса между высокой стоимостью инстансов по требованию и низкой производительностью SI. Алгоритм использует методы машинного обучения для динамического выделения ресурсов на основе оценки производительности при выполнении предыдущих заданий, истории спотовых цен и характеристик нагрузки.

Kokkinos и др., 2014 [13] представляют механизм оптимизации стоимости и нагрузки при использовании набора инстансов Amazon EC2, формулируемый как задача целочисленного линейного программирования.

Sowmya и Sundarraj, 2013 [23] моделируют стратегии ставок на спотовом рынке как хорошо известную в теории игр дилемму заключенного. Авторы проверяют свои результаты на реальных данных

спотового рынка Amazon EC2. Учитывая, что большинство игроков спотового рынка делают ставки много раз, авторы реализуют кооперативную стратегию, которая соответствует многошаговой дилемме заключенного.

Tang и др., 2012 [25] предлагают стратегии ставок для минимизации стоимости и предоставления хорошей надежности. Авторы используют ограниченный марковский процесс принятия решений (Constrained Markov Decision Process), матрицу вероятностей перехода цен (Price Transition Probability Matrix) и линейное программирование для получения оптимальной рандомизированной стратегии ставок (которая названа AMAZING). Авторы оценивают модель и демонстрируют как пользователи должны оптимально делать ставки на SI для достижения различных целей с желаемыми уровнями надежности.

Kaminski и Szufel, 2015 [10] представляют симулятор спотовых цен EC2; он работает с историческими данными и эмулирует механизм ценообразования Amazon EC2. Симулятор может использоваться для оценки различных стратегий ставок.

Wang и др., 2013 [29] и Тооси и др., 2016 [26] предлагают новые механизмы аукционов, направленных на увеличение дохода и обеспечение достоверности.

Wang и др., 2013 [29] разрабатывают специальный тип динамических аукционов. Указанный тип аукционов определяет число инстансов, которые выставляются на аукцион в каждом периоде, а также внутренние механизмы аукционов, основанные на динамических схемах оплаты. Авторы доказывают, что предлагаемый механизм является достоверным в двух измерениях и асимптотически оптимальным для высокого спроса.

Тооси и др., 2016 [26] адаптируют механизм аукциона Consensus Revenue Estimate для разработки онлайн аукциона аренды большого числа инстансов облачных ресурсов. Авторы комбинируют механизм со схемой динамического вычисления резервных цен, основываясь на энергоэффективности центра обработки данных и стоимости электричества. Отмечается, что финальный механизм имеет высокую вероятность достоверности и генерирует почти оптимальный доход для провайдера. Он также максимизирует доход без необходимости иметь

предварительные данные по распределению ставок. Основываясь на результатах моделирования, авторы показывают, что предложенный механизм превосходит аукционы с равномерной ценой.

В табл. 1 представлено сравнение исследовательских статей на основе их основных характеристик: исследование с точки зрения пользователя или облачного провайдера; использование существующего или разработка нового механизма аукциона; выполнение теоретического или экспериментального анализа; использование реальных или синтетических данных; использование аукциона Amazon Spot Instance или какого-то другого; выполнение характеристики цен.

Таблица 1. Анализ исследовательских статей

Статья / параметр	сторона (клиент/ облако)	способ (существ. /новый)	теорет. анализ	данные (реальные/ сгенерир.)	Amazon SI	описание цен
V. Abhishek и др., 2012 [1]	облако	новый	+	сгенерир.		
H. Xu, B. Li, 2012 [30]	облако	новый	+	сгенерир.		
W. Voorsluys, 2014 [27]	облако	новый	+	сгенерир.		
O. Agmon Ben-Yehuda и др., 2011 [5]		существ.	+	реальные	+	+
B. Javadi и др., 2013 [9]				реальные	+	+
R. M. Wallace и др., 2013 [28]		существ.		реальные	+	+
S. Karunakaran R. P. Sundarraaj, 2015 [12]	клиент	существ.		реальные	+	
K. Sowmya, R.P. Sundarraaj, 2013 [23]	клиент	существ.	+	реальные	+	
Y. Song и др., 2012 [22]	облако	новый	+	реальные сгенерир.	+	
S. Tang и др., 2012 [25]	клиент	существ.	+	реальные	+	+
B. Kaminski , P. Szufel, 2015 [10]				реальные	+	+
W. Wang и др., 2013 [29]	облако	новый	+	сгенерир.		
A.N. Toosi и др., 2016 [26]	облако	новый	+	сгенерир.		
I. Menache и др., 2014 [18]	облако	новый	+			

3. Спотовые аукционы Amazon

Amazon предлагает пользователям три типа механизмов аренды: зарезервированные инстансы, инстансы по запросу и спотовые аук-

ционы. Указанные механизмы отличаются уровнем гарантии, касающейся запуска и завершения инстансов, а также ценой. Зарезервированные инстансы дают возможность пользователю запустить зарезервированную виртуальную машину когда он пожелает. Пользователь может арендовать инстанс по требованию, когда пожелает, однако по более высокой цене, однако без гарантии того, что запуск будет возможен в конкретное время. Как зарезервированный, так и инстанс по требованию остаются активны до тех пор, пока не будут выключены пользователем.

Третий тип аренды – это аукцион Викри на неиспользованные мощности. Инстансы предоставляются без гарантии как времени запуска, так и времени завершения. Пользователи делают ставки на максимальную стоимость часа, которую они готовы платить за SI. Запрос на инстанс удовлетворяется, если сделанная ставка выше, чем спотовая цена, в противном случае – ожидает. Все победившие пользователи платят одинаковую цену, которая равна значению минимальной выигравшей ставки. Спотовая цена периодически изменяется в зависимости от количества неиспользуемых мощностей и спроса на них. Инстанс работает до тех пор, пока клиент не завершит его или пока цена не превысит указанную пользователем ставку.

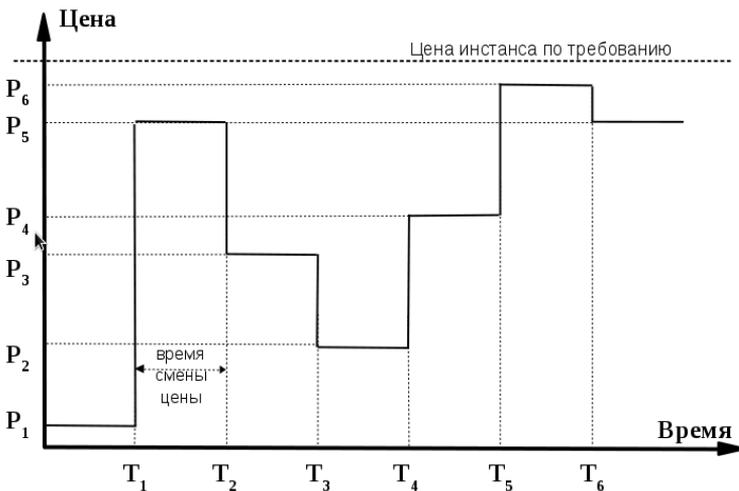


Рисунок 1. Структура спотовых цен

Amazon EC2 взимает плату за полные часы, прошедшие до момента прерывания инстанса из-за изменения спотовой цены. В этом случае незавершенный час считается бесплатным.

На рис. 2 показан пример динамики спотовых цен за три месяца (см. Amazon Spot Instance Pricing History [24]).

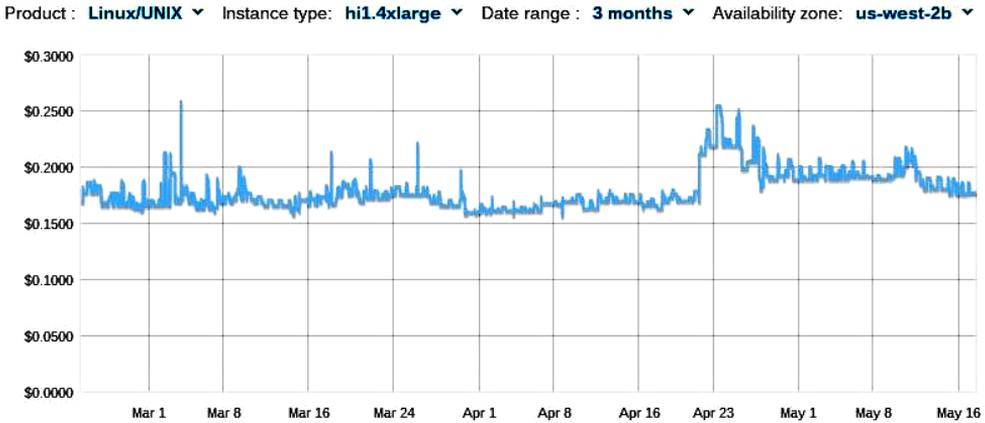


Рисунок 2. Динамика спотовых цен

Очевидно, что более низкая ставка обычно приводит к более низкой стоимости инстанса. Однако при этом ухудшаются другие метрики, такие как время завершения задания и число прерываний. Поэтому спотовый аукцион – это компромисс между надежностью сервиса и его стоимостью.

4. Стратегия ставок

4.1. Математическая модель

Задача наилучшего выбора с полной информацией была впервые сформулирована Cauley [6]. Moser, 1956 [19] переформулировал задачу Cauley следующим образом: наблюдаются последовательно одна за другой независимые одинаково распределенные случайные величины X_1, \dots, X_n из равномерного распределения на интервале $[0, 1]$. Остановившись на наблюдении X_j , наблюдатель получает в качестве

выигрыша значение случайной величины X_j . Необходимо максимизировать ожидаемый выигрыш.

Доказано, что оптимальная стратегия заключается в остановке на наблюдении m , если текущее наблюдение превышает E_m , где E_m определяется рекурсивно как $E_n = 0$ и $E_m = (1 + E_{m+1}^2)/2$, $m = 1, \dots, n-1$. Значения E_m называются порогами для стратегии такого типа.

Последовательность оптимальных порогов является убывающей, она представлена на рис. 3.

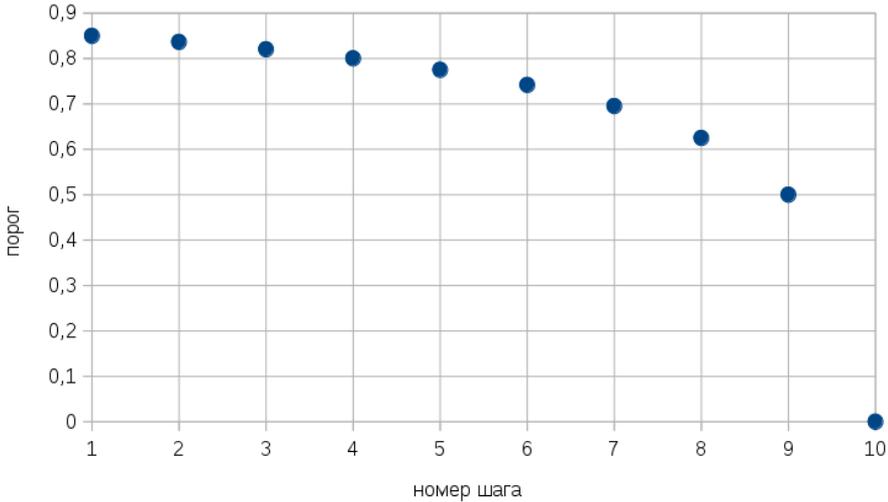


Рисунок 3. Последовательность порогов

Мы используем эти результаты для переформулирования ранее описанной задачи установки оптимальных ставок на спотовых аукционах.

4.2. Стратегия

Как можно заметить, задача наилучшего выбора имеет прямое сходство с задачей определения наилучшей ставки на облачном аукционе. Предположим, что пользователь хочет арендовать инстанс в определенный период времени. Как и в задаче наилучшего выбора, пользователь последовательно просматривает спотовые цены, которые распределены от 0 до некоторого значения p_{max} . Он может ис-

пользовать пороговую стратегию для минимизации ожидаемой стоимости инстанса. Полагая, что цены являются независимыми одинаково распределенными случайными величинами с известной непрерывной функцией распределения, наилучшая стратегия ставок может быть получена так же, как и при решении задачи наилучшего выбора.

Опишем формальную постановку задачи. Проводится спотовый аукцион, где спотовые цены – независимые одинаково распределенные случайные величины (н.о.р.с.в.) x_1, x_2, x_3, \dots с известным непрерывным распределением $F(x)$ на интервале $[0, p_{max}]$. p_{max} соответствует максимально возможной стоимости SI на аукционе (очевидно, что она не превышает стоимость инстанса по требованию). Пользователь имеет n периодов, чтобы выиграть SI. Если пользователь не получает инстанс в этот период, он вынужден покупать инстанс по требованию по максимальной цене p_{max} .

Пусть пользователь установил ставку τ_i перед i -м аукционом на SI. Он хочет минимизировать ожидаемую стоимость аренды инстанса за заданный период времени. Рассматриваемая задача – это хорошо известная задача наилучшего выбора с полной информацией, решаемая методом обратной индукции [19]. Не выиграв SI за период n , пользователь должен купить его по максимальной цене p_{max} . Поэтому любая цена меньшая или равна p_{max} подходит пользователю: $\tau_n = p_{max}$.

Принимая во внимание, что спотовые цены – н.о.р.с.в. с известной непрерывной функцией распределения $F(x)$, ожидаемая стоимость на шаге n равна $\int_0^{\tau_n} x dF(x)$ – интеграл Лебега-Стилтьеса по функции распределения $F(x)$ (см. [11]).

На шаге $n - 1$ уровень цены пользователя – это минимум между текущей стоимостью SI и ожидаемой стоимостью на n -м шаге:

$$\tau_{n-1} = \mathbf{E}[\min \{\tau_n, x\}] = \int_0^{p_{max}} \min \{\tau_n, x\} dF(x) = \int_0^{\tau_n} x dF(x) + \int_{\tau_n}^{p_{max}} \tau_n dF(x).$$

Здесь $\mathbf{E}[X]$ обозначает математическое ожидание случайной величины X .

Продолжая обратную индукцию, получим следующее решение как систему рекуррентных уравнений:

$$\begin{cases} \tau_n = p_{max}, \\ \tau_i = \mathbf{E}[\min \{\tau_{i+1}, x\}] = \int_0^{\tau_{i+1}} x dF(x) + \int_{\tau_{i+1}}^{p_{max}} \tau_{i+1} dF(x), i = 1, \dots, n - 1. \end{cases} \quad (4.1)$$

Используя формулу (4.1), можно получить значение оптимальной ставки τ_i для каждого шага i для заданного периода n . На первом шаге пользователь должен использовать первую (наименьшую) ставку. Если ставка τ_1 не выиграла, следующая ставка τ_2 используется на следующем шаге. Продолжая процесс, пользователь гарантированно получает инстанс за период n с минимальной ожидаемой стоимостью.

5. Результаты численного моделирования

Как было отмечено выше, формула (4.1) описывает решение для любой непрерывной функции распределения вероятностей $F(x)$. При этом, как указывают многие исследователи, реальное распределение спотовых цен Amazon сложное. Например, Javadi и др., 2013 [9] указывает, что спотовые цены – это комбинация нормальных распределений с тремя или четырьмя компонентами. Однако известно, что любое непрерывное распределение может быть сведено к равномерному на $[0, 1]$ с помощью соответствующего преобразования (см., например, Law и Kelton, 1999 [15]). Для стандартного равномерного распределения формула (4.1) преобразуется в более простую форму:

$$\begin{cases} \tau_n = 1, \\ \tau_i = \mathbf{E}[\min \{\tau_{i+1}, x\}] = \int_0^{\tau_{i+1}} x dx + \int_{\tau_{i+1}}^1 \tau_{i+1} dx, i = 1, \dots, n - 1. \end{cases} \quad (5.1)$$

На основе формулы (5.1) построим последовательность оптимальных ставок для 10 шагов (см. табл. 2 и рис. 4). Для оценки полезности предложенной стратегии (Стратегия 1) сравним ее с тремя другими эвристиками с помощью численного моделирования:

Таблица 2. Оптимальные значения ставок, $n = 10$

шаг	1	2	3	4	5	6	7	8	9	10
цена	0.150	0.164	0.180	0.20	0.225	0.258	0.305	0.375	0.5	1.0

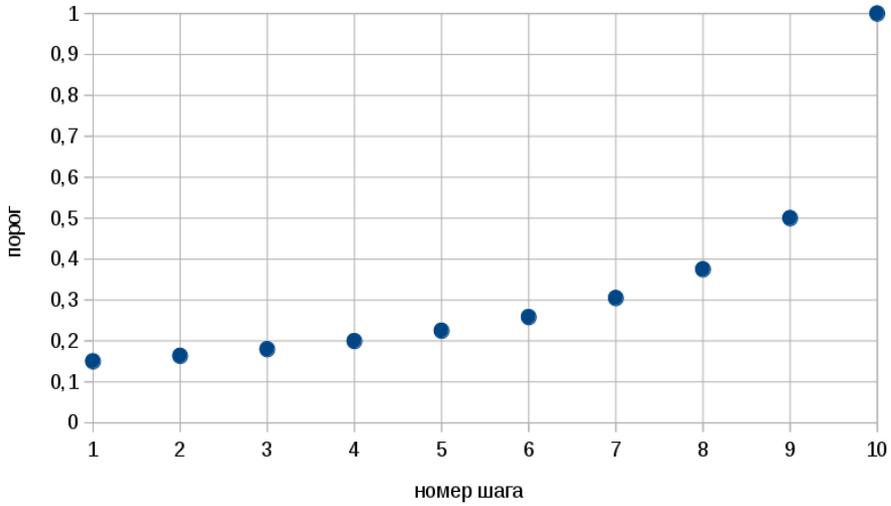


Рисунок 4. Оптимальные значения ставок для периода в 10 шагов и равномерного распределения

- Стратегия 2: постоянная стратегия — пользователь определяет постоянную ставку, которая используется на каждом шаге. Если за n шагов пользователь не получает инстанс, то он покупает инстанс по цене инстанса по требованию.
- Стратегия 3: случайная стратегия — на каждом шаге пользователь генерирует случайную ставку из равномерного на $[0, 1]$ распределения. Если за n шагов пользователь не получает инстанс, то он покупает инстанс по цене инстанса по требованию.
- Стратегия 4: линейная стратегия — пользователь использует пороги, получаемые по формуле $\tau = kn + b$, где n — это номер шага, а коэффициенты k и b зависят от общего числа шагов. Если за n шагов пользователь не получает инстанс, то он покупает инстанс по цене инстанса по требованию.

Перечислим все возможные события для вывода наилучшего порога постоянной стратегии. Имея постоянный порог x , ожидаемая стоимость инстанса определяется следующей формулой:

$$\begin{aligned}
 P(x) &= \int_0^x y dy + \int_x^1 dy \int_0^x y dy + \dots + \underbrace{\int_x^1 dy \dots \int_x^1 dy \int_0^x y dy}_{n-1} + \underbrace{\int_x^1 dy \dots \int_x^1 dy}_{n} \\
 &= (1-x)^n + \frac{x^2}{2} \sum_{i=0}^{n-1} (1-x)^i.
 \end{aligned}
 \tag{5.2}$$

Здесь первое слагаемое соответствует установлению на первом шаге спотовой цены ниже, чем x ; второе слагаемое – на втором шаге, и так далее. Наконец, последнее слагаемое соответствует тому, что на всех шагах спотовая цена была ниже x .

Очевидно, что чем больше x , тем больше шансов получить инстанс, однако при этом ожидаемая стоимость инстанса будет выше. Чем ниже x , тем ниже ожидаемая стоимость инстанса, однако меньше шансов его получить. Поэтому, наилучший постоянный порог для случая равномерно распределенных на $[0, 1]$ цен определяется следующей формулой:

$$x^* = \operatorname{argmin} \left((1-x)^n + \frac{x^2}{2} \sum_{i=0}^{n-1} (1-x)^i \right).
 \tag{5.3}$$

Для случая $n = 10$ формулы (5.2) и (5.3) дают значение порога $x^* = 0.275$ и ожидаемую стоимость $P(x^*) = 0.17$.

Наилучшие коэффициенты k и b линейной стратегии получены с помощью численного моделирования: $k = 0.04$ и $b = 0.09$ для случая $n = 10$, а для $n = 100$ $k = 0.001$ и $b = 0.01$.

Пороги всех четырех стратегий представлены в табл. 3 и на рис. 5. Заметим, что Стратегия 3 представляет собой реализацию случайного процесса, соответственно, в табл. 3 и на рис. 5 представлены псевдослучайные числа, сгенерированные из равномерного на интервале $[0, 1]$ распределения.

Результаты численного моделирования представлены в табл. 4.

Спотовые цены генерируются как псевдослучайные числа из интервала $[0, 1]$. На втором шаге спотовая цена ниже, чем порог Стра-

Таблица 3. Значения ставок в соответствии со стратегиями, период 10 шагов

порог/ стр.	1	2	3	4	5	6	7	8	9	10
Стр. 1	0.15	0.164	0.18	0.2	0.225	0.258	0.305	0.375	0.5	1.0
Стр. 2	0.275	0.275	0.275	0.275	0.275	0.275	0.275	0.275	0.275	0.275
Стр. 3	0.214	0.159	0.207	0.589	0.346	0.626	0.412	0.777	0.078	0.577
Стр. 4	0.13	0.17	0.21	0.25	0.29	0.33	0.37	0.41	0.45	0.49

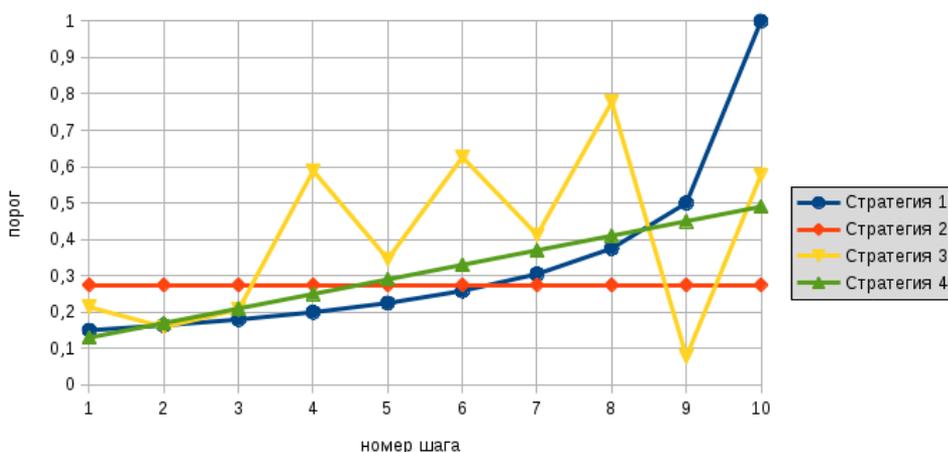


Рисунок 5. Значения ставок в соответствии со стратегиями, период 10 шагов

Таблица 4. Результаты численного моделирования ($n = 10$)

Шаг	1	2	3	4	5	6	7	8	9	10	шаг	спотовая	
Стр.											принятия	цена	
Стр. 1	0.150	0.164	0.18	0.2	0.225	0.258						6	0.161
Стр. 2	0.275	0.275									2	0.253	
Стр. 3	0.214	0.159	0.207	0.589							4	0.302	
Стр. 4	0.13	0.17	0.21	0.25	0.29						5	0.283	
спотовая	0.359	0.253	0.721	0.302	0.283	0.161							
цена													

тегии 2: используя эту стратегию, пользователь получает инстанс на втором шаге по цене 0.253. Используя Стратегию 3, пользователь получает инстанс на шаге 4, когда сгенерированный случайный порог превышает спотовую цену (шаг принятия 4 и спотовая цена 0.302). Спотовая цена ниже порога линейной стратегии на шаге 5 (шаг принятия 5 и спотовая цена 0.283). Наконец, последняя выигравшая ставка соответствует Стратегии 1: значение порога, превышающее спотовую цену, равно 0.258, шаг принятия 6 и спотовая цена 0.161.

Заметим, что в табл. 4 представлен только пример реализации стохастического процесса. Для оценки ожидаемых значений мы провели численное моделирование с усреднением 10^6 экспериментов. Среднее время принятия и средняя стоимость инстанса представлены в табл. 5.

Таблица 5. Сравнение стратегий ($n = 10$)

	Средний шаг принятия	Средняя спотовая цена
Стратегия 1	4.604	0.139
Стратегия 2	3.546	0.185
Стратегия 3	1.999	0.334
Стратегия 4	4.365	0.148

Как можно заметить, Стратегия 1, определяемая рекуррентной формулой (5.1) дает лучшие результаты, чем три другие стратегии с точки зрения средней стоимости инстанса.

Тот же эксперимент проведен для периода в 100 шагов. Стратегии представлены на рис. 6; в табл. 6 представлено сравнение по среднему шагу принятия и средней стоимости.

Таблица 6. Сравнение стратегий ($n = 100$)

	Средний шаг принятия	Средняя спотовая цена
Стратегия 1	35.298	0.0188
Стратегия 2	19.182	0.031
Стратегия 3	2.0	0.333
Стратегия 4	31.127	0.022

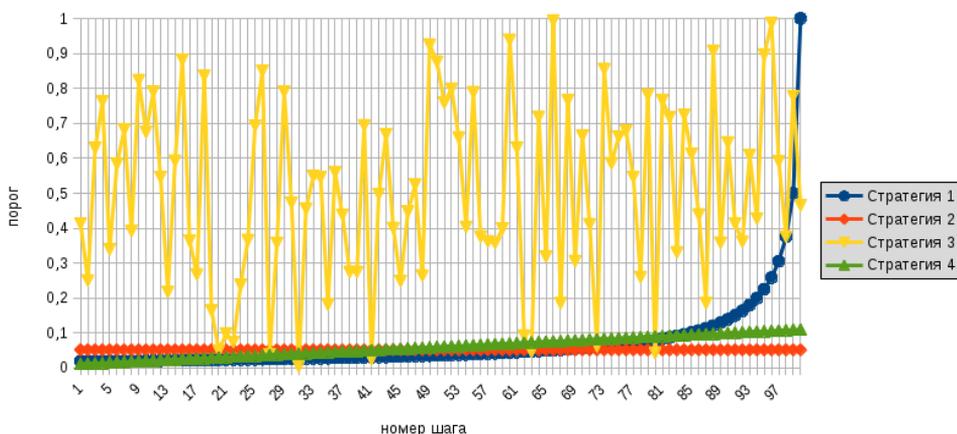


Рисунок 6. Значения ставок в соответствии со стратегиями, период 100 шагов

Стратегия 3 не зависит от номера шага и, в большинстве случаев, значительно превышает пороги остальных стратегий. Очевидно, что следуя этой стратегии стоимость инстанса будет значительно больше, чем при использовании трех других стратегий, при этом выигрыш на аукционе происходит на ранних шагах. Следуя постановке задачи, пользователь должен выиграть один аукцион за определенный период времени. Поэтому если период большой, то разумно устанавливать низкие ставки, надеясь на низкую стоимость инстанса.

Как можно заметить, предложенные рекуррентные пороги минимизируют среднюю стоимость инстанса. Также, более длинный период приводит к большему выигрышу в стоимости.

6. Моделирование спотовых цен Amazon

В данном разделе предложен метод получения оптимальной последовательности ставок на основе реальных данных спотовых цен Amazon [24]. Мы строим функцию распределения цен и получаем последовательность порогов, используя формулу (4.1).

Сначала, используя статистику спотовых цен, построим функцию распределения цен. Например, мы используем статистику инстанса

Linux р3.8xlarge за период с 1 по 22 февраля 2019 г. На основе этих данных построим гистограмму; анализируя данные, можно построить функцию распределения (используя соответствующие статистические методы, см., например, [21]), см. рис. 7. Мы используем усеченное нормальное распределение:

$$f_{b_1, b_2}(x) = \begin{cases} \frac{C}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, & x \in [b_1, b_2], \\ 0, & x \notin [b_1, b_2], \end{cases} \quad (6.1)$$

здесь $C = \frac{1}{\int_{b_1}^{b_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} dx}$, $b_1 = p_{min} = 3.672$, $b_2 = p_{max} = 4.307$ и $a = 4.002$, $\sigma = 0.141$.

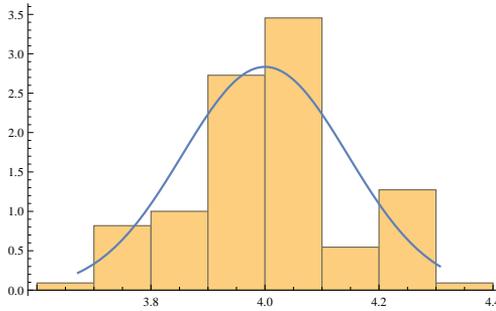


Рисунок 7. Гистограмма и функция распределения спотовых цен

Используя формулу (4.1), получим оптимальные пороги τ_i , $i = 1, \dots, n$ (см. табл. 7).

Таблица 7. Значения ставок в соответствии с оптимальной стратегией τ_i , период 10 шагов

i	1	2	3	4	5	6	7	8	9	10
τ_i	3.838	3.846	3.855	3.866	3.879	3.895	3.916	3.947	4.0	4.307

Для сравнения предложенной стратегии (Стратегия 1) с тремя другими стратегиями, описанными в разделе 5 (их параметры адаптируем к распределению (6.1), см. табл. 8), воспользуемся численным моделированием.

Таблица 8. Результаты численного моделирования, $n = 10$

шаг	1	2	3	4	5	6	7	8	9	10
Стр.1	3.838	3.846	3.855	3.866	3.879	3.895	3.916	3.947	4.000	4.307
Стр.2	3.916	3.916	3.916	3.916	3.916	3.916	3.916	3.916	3.916	3.916
Стр.3	3.862	3.936	3.890	3.913	3.873	3.895	4.051	3.950	3.973	4.013
Стр.4	3.819	3.843	3.866	3.890	3.914	3.938	3.962	3.985	4.010	4.033
	Средний шаг	Средняя цена								
Стр.1	8.364	3.990								
Стр.2	7.364	4.094								
Стр.3	7.273	4.073								
Стр.4	7.727	4.018								

Как можно заметить, Стратегия 1 (оптимальная) дает наименьшую стоимость инстанса. Та же процедура может использоваться для других временных периодов аукционов, где цены соответствуют некоторому непрерывному вероятностному распределению.

7. Заключение

Рассматривается задача установления ставок пользователем облачного сервиса для минимизации стоимости аренды инстанса. Предполагается, что пользователю не требуется получить инстанс немедленно, но есть ограниченный временной период, за который необходимо арендовать инстанс. Мы демонстрируем сходство между задачей установления ставок на облачные спотовые аукционы и задачей наилучшего выбора с полной информацией. На основе этого анализа мы предлагаем пороговую стратегию ставок, которая минимизирует ожидаемую стоимость аренды инстанса.

Проводится сравнение предложенной стратегии с тремя эвристическими стратегиями. Результаты численного моделирования демонстрируют, что стратегия дает лучшую стоимость за счет времени ожидания аренды. Также предложен метод выработки стратегии на основе реальных данных спотовых цен Amazon.

Для оценки реальной эффективности и эффекта требуются дальнейшие исследования на основе реальных распределений спотовых цен. Такие исследования также помогут оценить влияния вероятностного распределения на эффективность стратегии ставок. Реальное

распределение сложное, его параметры изменяются во времени; тип распределения зависит от типа инстанса, региона, сезонности и т.д. Предложенная модель может быть объединена с динамическим анализом типа вероятностного распределения текущих спотовых цен.

Другое возможное направление исследований связано с учетом возможности прерывания инстанса. Аренда инстанса не гарантирует ее непрерывной работы, соответственно, пользователь может не успеть закончить расчет своей задачи. Повышение вероятности непрерывной работы может быть достигнуто специальным набором порогов. Как говорилось ранее, высокие пороги повышают вероятность выигрыша в аукционе (и непрерывности работы инстанса), однако повышают среднюю стоимость инстанса; с другой стороны, низкие пороги дают более низкую стоимость инстанса, но снижают шансы на победу в аукционе (и на непрерывность работы инстанса). Необходимость непрерывной работы также может быть учтена в математической модели.

СПИСОК ЛИТЕРАТУРЫ

1. Abhishek V., Kash I.A., Key P. *Fixed and Market Pricing for Cloud Services* // Proceedings of the 7th Workshop Economics of Networks System Computer (NetE-con 2012), IEEE Computer Society, Orlando, 20 March 2012. P. 157–162.
2. Alkharif S., Lee K., Kim H. *LSTM model to forecast time series for EC2 cloud price* // 16th IEEE Int Conf on Dependable, Autonom and Secure Computing, 16th IEEE Intl Conf on Pervasive Intelligence and Computing, 4th IEEE Intl Conf on Big Data Intelligence and Computing, 3rd IEEE Cyber Science and Technology Congress (DASC/PICOM/DATACOM/CYBERSCITECH), Athens. 2018. P. 1085–1088.
3. Amazon EC2 Spot Instances. <http://aws.amazon.com/ec2/spot/>
4. Amazon Inc. Amazon Elastic Compute Cloud (Amazon EC2). <http://aws.amazon.com/ec2>

5. Ben-Yehuda A. O., Ben-Yehuda M., Schuster A., Tsafirir D. *Deconstructing Amazon EC2 Spot Instance Pricing* // Cloud Computing Technology and Science (Cloud-Com), 2011 IEEE Third International Conference on, Athens. 2011. P. 304–311.
6. Cayley A. *Mathematical questions with their solutions* // The Educational Times 23 18-19. 1875. See The Collected Mathematical Papers of Arthur Cayley 10 587-588 (1896). Cambridge Univ. Press, Cambridge.
7. Cheng H.K., Li Z., Naranjo A. *Cloud computing spot pricing dynamics: Latency and limits to arbitrage* // Information Systems Research. 2016. V. 27. N 1. P. 145–165.
8. Chhetri M. B., Lumpe M., Quoc B. V., Kowalczyk R. *On Forecasting Amazon EC2 Spot Prices Using Time-Series Decomposition With Hybrid Look-Backs* // 2017 IEEE 1ST INTERNATIONAL CONFERENCE ON EDGE COMPUTING (IEEE EDGE), eds: Goscinski, AM and Luo, M. 2017. P. 158–165.
9. Javadi B., Ruppa K. Thulasiram, Rajkumar Buyya *Characterizing spot price dynamics in public cloud environments* // Future Generation Computer Systems, V. 29, Issue 4, June 2013, P. 988–999, ISSN 0167-739X, <http://dx.doi.org/10.1016/j.future.2012.06.012>.
10. Kaminski B., Szufel P. *On optimization of simulation execution on Amazon EC2 spot market* // Simulation Modelling Practice and Theory. 2015. V. 58. P. 172–187.
11. Karr A.F. *Probability*. Springer-Verlag, 1993.
12. Karunakaran S., Sundarraj R. P. *Bidding Strategies for Spot Instances in Cloud Computing Markets* // IEEE Internet Computing. May-June 2015. V. 19. N 3. P. 32–40.
13. Kokkinos P., Varvarigou T.A., Kretsis A., Soumplis P., Varvarigos E.A. *SuMo: Analysis and Optimization of Amazon EC2 Instances* // Journal of Grid Computing. 2014. V. 13. P. 255–274.

14. Kumar D., Baranwal G., Raza Z., Vidyarthi D. P. *A Survey on Spot Pricing in Cloud Computing* // Journal of Network and Systems Management. 2018. P. 809–856.
15. Law A.M., Kelton D.M. *Simulation Modeling and Analysis (3rd ed.)*. McGraw-Hill Higher Education, 1999.
16. Li Z., O'Brien L., Ranjan R., Flint S., Zomaya A. *On Cloud Spot Market: An Overview of the Research* // In A. Al-Saidi, R. Fleischer, Z. Maamar, & O.F. Rana (Eds.), *Intelligent Cloud Computing: First International Conference, ICC 2014 Muscat, Oman, February 24–26, 2014 Revised Selected Papers*. 2015. P. 46–61.
17. Mell P., Grace T. *The NIST definition of cloud computing* // NIST Special Publication 800145 (2011). 2011. <http://dx.doi.org/10.6028/NIST.SP.800-145>
18. Menache I., Shamir O., Jain N. *On-demand, Spot, or Both: Dynamic Resource Allocation for Executing Batch Jobs in the Cloud* // 11th International Conference on Autonomic Computing (ICAC 14). 2014. isbn 978-1-931971-11-9, P. 177–187, <https://www.usenix.org/conference/icac14/technical-sessions/presentation/menache>
19. Moser L. *On a problem of Cayley* // Scripta Math. 1956. V. 22. N 5. P. 289–292.
20. Petcu D. *Consuming resources and services from multiple clouds* // J. Grid Comput. 2014. V. 12 (2). P. 321–345.
21. Rohatgi V.K. *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley & Sons, 1976.
22. Song Y., Zafer M., Lee K.-W. *Optimal bidding in spot instance market* // Proceedings – IEEE INFOCOM. 2012. art. no. 6195567. P. 190–198.
23. Sowmya K., Sundarraj R.P. *Strategic bidding for cloud resources under dynamic pricing schemes* // (2013) Proceedings – 2012 International Symposium on Cloud and Services Computing, ISCOS 2012. art. no. 6481231. P. 25–30.

24. Spot Instance Pricing History Amazon Elastic Compute Cloud
<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances-history.html>
25. Tang S., Yuan J., Li X.Y. *Towards Optimal Bidding Strategy for Amazon EC2 Cloud Spot Instance* // Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on, Honolulu, HI. 2012. P. 91–98.
26. Toosi A.N., Vanmechelen K., Khodadadi F., Buyya R. *An auction mechanism for cloud spot markets* // ACM Transactions on Autonomous and Adaptive Systems. 2016. 11 (1), art. no. 2. P. 1–33.
27. Voorsluys W. *Resource provisioning in spot market-based cloud computing environments*. 2014.
28. Wallace R. M. et al. *Applications of neural-based spot market prediction for cloud computing* // Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2013 IEEE 7th International Conference on, Berlin. 2013. P. 710–716.
29. Wang W., Liang B., Li B. *Revenue maximization with dynamic auctions in IaaS cloud markets* // Quality of Service (IWQoS), IEEE/ACM 21st International Symposium on, Montreal, QC. 2013. P. 1–6.
30. Xu H., Li B. *Maximizing revenue with dynamic cloud pricing: The infinite horizon case* // 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, 2012. P. 2929–2933.

COST-EFFICIENT STRATEGY IN CLOUDS WITH SPOT PRICE UNCERTAINTY

Evgeny E. Ivashko, Institute of Applied Mathematical Research of Karelian Research Centre of RAS, Petrozavodsk State University, Cand.Sc. (ivashko@krc.karelia.ru),

Anna A. Ivashko, Institute of Applied Mathematical Research of Karelian Research Centre of RAS, Petrozavodsk State University, Cand.Sc. (aivashko@krc.karelia.ru),

Georgy R. Safonov, Petrozavodsk State University (jiri.safonov@gmail.com).

Andrey Tchernykh, CICESE Research Center, Prof. (chernykh@cicese.mx).

Abstract: We address a cloud spot bidding problem for user cost optimization. We propose stochastic optimization model to minimize the expected resource rental cost in the presence of spot price uncertainty. The model is based on the well-known full-information best-choice problem. Based on the model, we derive the strategy for cloud spots bidding. The strategy allows to minimize the expected cost for a spot instance in a specific period of time with quality of service guarantee. Our simulation analysis based on realistic settings clearly demonstrates the advantages of the proposed optimization solutions.

Keywords: cloud computing, spot instance, mathematical modeling, full-information best-choice problem, Amazon EC2.