

Математическая модель метода разделения сетевых ресурсов для видео- и веб-трафика¹

Н.С. Жирнов, А.И. Ляхов, Е.М. Хоров

Институт проблем передачи информации им. А.А. Харкевича РАН, Москва

Поступила в редколлегию 10.09.2018 г.

Аннотация—В работе исследуются алгоритмы адаптивного управления битрейтом для управления качеством видеопотока, переданного от камер видеонаблюдения до удаленных серверов через беспроводную сеть с видео- и веб-трафиком. В таком сценарии алгоритмы адаптивного управления битрейтом могут принимать решения на основе текущего состояния буфера камер. В работе описывается математическая модель передачи видеопотока, позволяющая оценивать производительность различных алгоритмов управления битрейтом.

КЛЮЧЕВЫЕ СЛОВА: разделение сетевых ресурсов, xStream, система видеонаблюдения, веб-трафик.

1. ВВЕДЕНИЕ

На протяжении последнего десятилетия стремительно растет объем мультимедийных данных, передающихся по телекоммуникационным сетям, при этом основной вклад вносит передача видеоданных. Существует множество сервисов, в основу которых положена доставка видеоданных с использованием телекоммуникационных сетей. В качестве примеров можно привести сервисы просмотра сохраненных на сервере видеоизображений (Netflix, YouTube), IP-телевидение, видеоконференцсвязь (Skype, FaceTime, всевозможные мессенджеры), онлайн трансляции в социальных сетях, организация систем видеонаблюдения. Появление огромного числа гаджетов, которые зачастую не имеют проводного сетевого интерфейса, а также размещение приемника или передатчика в труднодоступных местах способствовали бурному развитию беспроводных технологий передачи данных. Вследствие чего такие технологии, как LTE и Wi-Fi, в настоящее время широко используются для передачи видеоданных.

Для обеспечения высокого качества видеопотока необходимо обладать информацией об объеме предоставляемых ресурсов сети. Однако, скорость передачи данных, доступная различным видеопользователям, постоянно меняется. Это связано как с непостоянством состояния канала при использовании беспроводной сети для передачи данных, так и с наличием непостоянного числа пользователей в сети и ее гетерогенностью (т.е. наличием пользователей, передающих разные типы данных, например, видео и веб-данные). В таких условиях для обеспечения приемлемого качества предоставляемого сервиса используется технология адаптивного изменения битрейта (битрейт определяется как объем видеоданных, содержащих одну секунду видеопотока). Технология заключается в уменьшении битрейта передаваемого видеопотока в случае уменьшения доступной пропускной способности и в увеличении битрейта в противоположном случае. Без использования технологии адаптивного битрейта видеопоток будет либо воспроизводиться с прерываниями (из-за нехватки данных для проигрывания), либо записываться с низким битрейтом.

¹ Исследование выполнено в ИППИ РАН за счет гранта Правительства Российской Федерации (Договор No 14.W03.31.0019)

В данной работе рассматривается применение технологии адаптивного изменения битрейта для организации восходящей передачи системы видеонаблюдения, подключаемой к действующей гетерогенной беспроводной централизованной сети (например, LTE или 802.11ax[1]). Камеры видеонаблюдения генерируют видеопоток, записываемый с определенным битрейтом, и передают его на удалённый сервер через базовую станцию беспроводной сети, рис. 1. На сервере видеопоток воспроизводится оператором в режиме реального времени.



Рис. 1. Схема системы видеонаблюдения

Наша работа посвящена построению математической модели передачи видеопотока с камер на удаленный сервер через гетерогенную централизованную беспроводную сеть. Модель позволяет оценивать целевые показатели качества передаваемого видеопотока для определенного класса адаптивных алгоритмов выбора битрейта. В работе показано, что построенная модель может быть применена для оптимизации подобных алгоритмов.

Дальнейшее изложение построено следующим образом. В разделе 2 приводится обзор работ, исследующих алгоритмы адаптивного выбора битрейта. Раздел 3 подробно описывает процесс работы рассматриваемой системы видеонаблюдения, а в разделе 4 приводится формальная постановка задачи. В разделе 5 строится математическая модель передачи видеосегментов, которая применяется для оптимизации определенного алгоритма выбора битрейта в разделе 6.

2. ОБЗОР ЛИТЕРАТУРЫ

Множество работ посвящено улучшению качества передаваемого видеопотока. При этом для оценки предоставляемого сервиса используются различные критерии качества восприятия или QoE (англ. Quality of Experience) и обслуживания или QoS (англ. Quality of Service). Существуют различные объективные и субъективные критерии качества восприятия. MOS (англ. Mean Opinion Score) является примером субъективного критерия. Для его оценки группа экспертов оценивает различные параметры полученного видеопотока. Для объективной оценки качества восприятия можно сравнивать исходный видеопоток с полученным, используя такие метрики, как MSE (англ. Mean Squared Error) и SSIM (англ. Structural SIMilarity), которые описываются в работах [2] и [3]. Для оценки качества обслуживания измеряются такие параметры, как задержка между моментом записи видеоданных и моментом их воспроизведения, длительность замираний при проигрывании, средний битрейт видеопотока, частота переключений битрейта проигрываемых видеоданных и т.д.

Для управления объемом видеопотока в нашей работе используются алгоритмы адаптивного выбора битрейта. Существует несколько работ, предлагающих различные адаптивные алгоритмы. Работа [4] описывает алгоритм, в котором предлагается собирать данные о загрузке видеосегментов и оценивать пропускную способность канала. Основываясь на этой оценке, алгоритм для следующего видеосегмента выбирает максимальный битрейт, при котором будут отсутствовать замирания. Работа [5] — пример алгоритма, основанного на использовании информации о текущем объеме буфера проигрывателя. Алгоритм максимизирует целевую функцию, основанную на среднем битрейте получаемого видеопотока и частоте замираний при проигрывании. В работе доказывается оптимальность такого подхода для видеопотоков бесконечной длительности. Однако, в представленных работах не производится анализа работы алгоритмов в гетерогенной сети.

Существует несколько работ, посвященных применению адаптивных алгоритмов выбора битрейта в системах видеонаблюдения. Общая концепция системы видеонаблюдения, подключаемой к беспроводной сети, может быть найдена в работе [6]. Следуя этой концепции, работа [7] предлагает детальное описание системы видеонаблюдения. Видеопоток записывается с фиксированным битрейтом, а для уменьшения замираний при проигрывании изменяется частота кадров воспроизводимого видеопотока. Такой подход является приемлемым для компьютерной обработки воспроизводимого видеопотока, но может вызывать ухудшение качества восприятия, испытываемого человеком, просматривающим такой видеопоток. В работе [8] рассматривается передача видеопотока в режиме реального времени с мобильного робота. Для уменьшения длительности замираний при сохранении высокого битрейта используется адаптивный алгоритм выбора битрейта, основанный на оценке доступной полосы передачи для отдельной камеры. В случае гетерогенной сети (особенно в случае наличия высокоприоритетных пользователей) такой подход может приводить к неверной оценке доступной полосы передачи и, как следствие, её неэффективному использованию.

Качество видеопотока, переданного по беспроводной сети, зависит от синхронизации решения приложения по выбору битрейта и решения сетевой инфраструктуры по выделению сетевых ресурсов для передачи данных. Для повышения качества видеопотока недавно был предложен ряд решений, таких как SAND[9] и xStream[10], обеспечивающих взаимодействие между сетевой инфраструктурой и приложениями. В рассматриваемой в статье задаче обмен информацией между сетевой инфраструктурой и приложением может быть осуществлен косвенным образом. Если выбран слишком высокий битрейт по сравнению с имеющимися сетевыми ресурсами, то на камерах будет расти объем данных в буфере отправителя. В противном случае он будет малым.

3. ОПИСАНИЕ РАБОТЫ СИСТЕМЫ ВИДЕОНАБЛЮДЕНИЯ

Мы рассматриваем систему видеонаблюдения, состоящую из N камер, подключенных к базовой станции беспроводной централизованной сети (рис. 1). В качестве трафика, отличного от видеоданных, мы рассматриваем веб-данные как один из наиболее распространенных видов трафика [11]. Из-за того, что веб-пользователи оплачивают предоставляемый сервис, они являются первичными пользователями сети и обслуживаются с абсолютным приоритетом (т.е. при наличии данных от веб-пользователей все частотно-временные ресурсы сети выделяются для их обслуживания). Заметим, что значительную долю восходящего веб-трафика составляют запросы для получения данных с удаленных серверов. Размеры запросов различаются незначительно. В данной работе мы пренебрегаем такими различиями и считаем, что веб-данные представляют собой запросы постоянного размера.

Каждая камера записывает видео в виде видеосегментов, каждый из которых содержит запись D секунд реального времени. Записанные видеосегменты кладутся в буфер камеры для

последующей передачи на базовую станцию. Размер каждого видеосегмента определяется его битрейтом. Формально, каждые D секунд видеочасть выполняет следующую последовательность действий:

1. Определение битрейта следующего видеосегмента (выбор производится с помощью алгоритма выбора битрейта).
2. В течение следующих D секунд происходит передача данных из буфера камеры на базовую станцию и одновременная запись нового видеосегмента в выбранном битрейте.
3. Добавление записанного видеосегмента в буфер камеры в конце периода времени D .

Камеры способны генерировать видеосегменты с битрейтом, лежащим в интервале $[b_{min}, b_{max}]$.

Оператор загружает данные с базовой станции в свой буфер и осуществляет воспроизведение данных из буфера. Формально, каждые D секунд оператор выполняет следующие действия:

1. Загрузка данных с базовой станции в буфер оператора (мы пренебрегаем временем этой загрузки, считая, что самым узким местом рассматриваемой сети является соединение между камерами и базовой станцией).
2. В течение следующих D секунд оператор осуществляет воспроизведение очередного видеосегмента из буфера. Если в буфере не оказывается полностью загруженного видеосегмента, оператор ничего не воспроизводит в течение D секунд. Такая ситуация называется замиранием видеопотока.

Сразу после подключения камеры к сети и начала записи видеоданных, оператор обеспечивает начальную задержку, равную $(K_0 - 1)D$ секунд, не проигрывая данные из буфера в течение этого времени (после данного периода времени оператор начинает воспроизводить данные из буфера). Присутствие такой начальной задержки обеспечивает наполнение буфера оператора данными для проигрывания. Объем буфера камеры считается ограниченным: камеры не записывают очередной видеосегмент, если их буфер содержит K_0 видеосегментов.

4. ПОСТАНОВКА ЗАДАЧИ

В данной работе исследуется описанная система видеонаблюдения в произвольной гетерогенной централизованной беспроводной сети. В качестве адаптивного алгоритма выбора битрейта рассматриваются произвольные алгоритмы, выбор битрейта которых описывается функцией состояния буфера камеры (например, текущего объема данных в буфере). Поскольку эффективность применения различных адаптивных алгоритмов выбора битрейта для обслуживания системы видеонаблюдения в гетерогенных сетях не была исследована в литературе, в нашей работе ставятся следующие задачи:

1. Разработка математической модели передачи видеосегментов на сервер через произвольные гетерогенные централизованные беспроводные сети для оценки среднего битрейта видеосегментов и вероятности замирания при проигрывании оператором.
2. Применение разработанной математической модели для оптимизации адаптивных алгоритмов выбора битрейта, основанных на состоянии буфера камеры.

5. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПЕРЕДАЧИ ВИДЕОСЕГМЕНТОВ

Мы предполагаем, что камеры и оператор ведут синхронный отсчет времени. Тогда в рассматриваемой системе видеонаблюдения количество полностью загруженных видеосегментов L в буфере оператора в моменты времени перед выбором битрейта очередного видеосегмента

(напомним, что эти моменты кратны периоду D), можно определить, исходя из количества l непереданных видеосегментов, находящихся в буфере камер видеонаблюдения в эти моменты: $L = K_0 - l$. Таким образом, замирание в течение следующих D секунд происходит в том случае, если $L = 0$ или $l = K_0$. Рассинхронизация часов камер и оператора (по сравнению с синхронизированными часами) приводят к появлению постоянной добавки к задержке между моментом записи видеосегмента и моментом начала его воспроизведения. Такая добавка не влияет на основные результаты модели.

Мы полагаем, что качество канала между пользователями (камерами и веб-пользователями) и базовой станцией остается постоянным, а базовая станция равномерно распределяет частотно-временные ресурсы между всеми видеокамерами, имеющими данные для передачи. Тогда объем частотно-временных ресурсов, доступных для передачи одной камеры, составляет $\frac{1}{N}$ от объема ресурсов, свободных от обслуживания веб-заявок. В таких предположениях мы можем считать скорость передачи данных камерой R_{up} в моменты, когда в системе отсутствуют веб-заявки, постоянной (если R_{up} — скорость передачи данных при $N = 1$, то $\frac{R_{up}}{N}$ — скорость передачи при $N > 1$). Таким образом, камеры оказываются независимы друг от друга и мы можем рассматривать систему с одной камерой видеонаблюдения без потери общности. Поток веб-заявок от веб-пользователей описывается пуассоновским потоком интенсивности λ . Предполагается, что каждый веб-пакет передается за одинаковое время, равное τ секундам.

Работу рассматриваемой системы видеонаблюдения можно представить в виде системы массового обслуживания (СМО), в которой присутствуют две очереди:

1. Высокоприоритетная очередь веб-пакетов типа M/D/1.
2. Низкоприоритетная очередь видеосегментов типа D/G/1. Видеосегменты появляются в системе каждые D секунд. Время передачи видеосегментов зависит от времени, затраченного на обслуживание высокоприоритетной очереди веб-пакетов.

Поскольку выбор битрейта очередного видеосегмента осуществляется на основе состояния буфера камеры (например, его объема), опишем это состояние. Для этой цели дискретизуем непрерывный интервал $[b_{min}, b_{max}]$ допустимых битрейтов видеосегментов (обозначим упорядоченное дискретное множество допустимых битрейтов через \mathcal{B}):

$$\mathcal{B} = \{b_i\}_{i=0}^{N_b}, \quad b_0 = 0, \quad b_{min} \leq b_1 \leq \dots \leq b_{max}.$$

Кроме того, мы дискретизуем долю от объёма текущего видеосегмента, которая не была передана к текущему моменту, в виде упорядоченного множества \mathcal{Q} :

$$\mathcal{Q} = \{q_j\}_{j=0}^{N_q}, \quad q_0 = 0, \quad 0 \leq q_1 \leq \dots \leq q_{N_q} \leq 1.$$

При фиксированном значении доли $q \in \mathcal{Q}$ передаваемого видеосегмента, остающейся в буфере камеры на текущий момент, и битрейте этого видеосегмента $b \in \mathcal{B}$, объём непереданных данных данного видеосегмента оценивается как $q \cdot b \cdot D$ байт.

Состояние буфера камеры полностью описывается K_0+1 -мерным вектором $X = (j, i_1, \dots, i_{K_0})$. Здесь представлены номера i_j битрейтов b_{i_j} , $b_{i_j} \in \mathcal{B}$, всех видеосегментов, располагающихся в буфере камеры видеонаблюдения (напоминаем, что в буфере не может находиться более K_0 видеосегментов). Поскольку часть первого видеосегмента может быть уже передана, мы используем номер j для указания значения доли q_j от объёма этого видеосегмента, которая остаётся в буфере на данный момент. Если в буфере камеры содержится $r < K_0$ видеосегментов, то $i_{r+1} = \dots = i_{K_0} = 0$.

Рассматриваемую СМО можно описывать марковской цепью дискретного времени с единицей времени, равной D секундам. При этом состояния данной цепи являются состояниями

буфера видеокамеры в моменты перед выбором битрейта очередного видеосегмента. Множество состояний, в которые возможен переход из текущего состояния, определяется используемым алгоритмом выбора битрейта. Для описания марковского процесса необходимо оценить вероятности перехода между состояниями. Рассмотрим различные типы переходов.

Пусть выбор битрейта определяется функцией $B(X)$, где X обозначает текущее состояние буфера (т.е. марковской цепи). Если $X = (j, i_1, \dots, i_n, 0, \dots, 0)$, то камера записывает новый видеосегмент с битрейтом $b_{i_{n+1}} = B(X)$. В общем случае при $n < K_0$ возможны следующие переходы:

$$X \rightarrow \begin{cases} (l, i_1, \dots, i_n, i_{n+1}, 0, \dots, 0), & 1 \leq l \leq j; \\ (l, i_m, \dots, i_n, i_{n+1}, 0, \dots, 0), & 1 < m \leq n, 1 \leq l \leq N_q; \\ (N_q, i_{n+1}, 0, \dots, 0). \end{cases}$$

Если $X = (j, i_1, \dots, i_{K_0})$, $i_{K_0} > 0$, то, согласно принятым соглашениям о работе системы видеонаблюдения, новый видеосегмент не записывается (в таком случае мы полагаем, что $B(X) = 0$), т.е. допустимы следующие переходы:

$$X \rightarrow \begin{cases} (l, i_1, \dots, i_{K_0}), & 1 \leq l \leq j; \\ (l, i_m, \dots, i_{K_0}, 0, \dots, 0), & 1 < m \leq K_0, 1 \leq l \leq N_q; \\ (0, \dots, 0). \end{cases}$$

Состояние $(0, \dots, 0)$ возникает в том случае, если в течение очередных D секунд были переданы все данные из буфера камеры.

Из состояния $X = (0, \dots, 0)$ с опустошенным буфером камеры возможен только один переход ($b_{i_1} = B(0, \dots, 0)$):

$$X \rightarrow (N_q, i_1, 0, \dots, 0).$$

Суммируя вышесказанное, мы описываем поведение рассматриваемой СМО как марковскую цепь дискретного времени. Состояния цепи представляют собой $K_0 + 1$ -мерный вектор $X = (j, i_1, \dots, i_{K_0})$, $0 \leq i_k \leq N_b$, $0 \leq j \leq N_q$, являющийся детализацией состояния буфера камеры видеонаблюдения в моменты выбора битрейта очередного видеосегмента. Данная цепь апериодическая и неприводимая, что приводит к эргодичности и существованию стационарного распределения $\hat{p}(X)$, описывающего вероятность нахождения цепи в состоянии X .

Для определения $\hat{p}(X)$ необходимо оценить значения переходных вероятностей между всеми возможными состояниями цепи. Для этой цели мы вводим функцию плотности вероятности $f(x)$, описывающую вероятность события, заключающегося в том, что, внутри произвольно взятого интервала длительности D секунд, времени, свободного от обслуживания веб-пакетов, достаточно для передачи x байт. Для оценки $f(x)$ мы рассматриваем функцию плотности вероятности $e(t)$, выражающую вероятность того, что в системе массового обслуживания с очередью типа M/D/1 (напомним, что такая система описывает очередь веб-заявок, представленную в системе видеонаблюдения) в течение времени D ровно t секунд будут свободны от обслуживания заявок.

Теорема 1. Для СМО с очередью типа M/D/1, интенсивностью появления заявок λ и фиксированным временем обслуживания заявки τ , функция плотности вероятности $e(t)$, описывающая вероятность события, заключающегося в том, что, в течение произвольно взятого интервала времени длительности D секунд, ровно $0 \leq t \leq D$ секунд будут свободны от обслуживания заявок, оценивается следующим образом:

$$e(t) = e_0 \delta(t) + \pi_0 \sum_{i=0}^{\lfloor \frac{D}{\tau} \rfloor} P_D^i \delta(t - D + i\tau) + \sum_{j=0}^{\lfloor \frac{D-t}{\tau} \rfloor} M(t + j\tau, P_{t+j\tau}^j), \quad 0 \leq t \leq D; \quad (1)$$

где

$$e_0 = 1 - \int_0^D l(x)dx; P_t^k = \frac{\lambda^k(t - k\tau)t^{k-1}}{k!}e^{-\lambda t}; R(t) = 1 - \sum_{r=0}^{\lfloor \frac{t}{\tau} \rfloor - 1} B_{r+1};$$

$$M(t, \Pi(t)) = (\pi_0 \cdot \lambda \cdot R(D - t) + l(D - t) + I(D - t)) \cdot \Pi(t);$$

$$l(t) = \pi_0 \delta(t) + \sum_{q=0}^{\lfloor \frac{t}{\tau} \rfloor} \frac{\pi_{q+1}}{\tau} P_t^{\lfloor \frac{t}{\tau} \rfloor - q} = \pi_0 \delta(t) + \hat{l}(t); I(A) = \int_0^A \hat{l}(x) \lambda R(A - x) dx;$$

δ — дельта-функция Дирака; $B_i = \frac{(i\tau\lambda)^{i-1}}{i!} e^{-i\tau\lambda}$ — распределение Бореля [12]; π_μ обозначает стационарную вероятность обнаружения в СМО с очередью типа $M/D/1$ ровно μ заявок.

Доказательство теоремы приведено в разделе 5.2.

С помощью замены переменных в функции плотности вероятности $e(t)$ мы получаем выражение для $f(x) = \frac{N}{R_{up}} e\left(\frac{xN}{R_{up}}\right)$. Зафиксируем определенное состояние X рассматриваемой цепи Маркова. Объём данных в буфере камеры $V(X)$ в состоянии X оценивается следующим образом: $V(X) = q_j \cdot b_{i_1} \cdot D + D \sum_{w=2}^{K_0} b_{i_w}$. Пусть S — множество всех состояний цепи, для которых значения переходной вероятности из состояния X отлично от нуля: $S = \{S_1, \dots, S_{K(X)}\}$, где $K(X)$ — мощность множества S . Объём данных $V(X \rightarrow S_i)$, который необходимо передать для перехода из состояния X в состояние S_i , равен $V(X \rightarrow S_i) = V(X) - (V(S_i) - B(X) \cdot D)$. Здесь учтено, что видеосегмент с битрейтом $B(X)$ только записывается при переходе из X в S_i и не влияет на объём передаваемых данных.

Пусть состояния множества S расположены в порядке убывания по объёму данных в буфере. Напомним, что единица времени марковской цепи соответствует D секундам передачи данных с камеры видеонаблюдения. В течение этого времени возможна передача произвольного объёма данных от 0 до $\min\left\{V(X), \frac{R_{up}}{N}D\right\}$ байт, при этом конкретный объём переданных данных зависит от времени, свободного от обслуживания веб-пакетов. Поскольку для описания всех возможных переходов из состояния X рассматривается множество S , значения $V(X \rightarrow S_i)$ которого в общем случае не полностью заполняют интервал $\left[0, \min\left\{V(X), \frac{R_{up}}{N}D\right\}\right]$ из-за дискретизации доли $q \in \mathcal{Q}$ передаваемого видеосегмента, остающейся в буфере камеры, мы оцениваем вероятность перейти из состояния X в состояние S_i как вероятность передачи произвольного количества байт, лежащего в интервале, который включает значение $V(X \rightarrow S_i)$, рис. 2 (интервалы не пересекаются и покрывают всё множество $[0, V(X)]$):

$$p(X, S_i) = \begin{cases} \frac{V(X \rightarrow S_1) + V(X \rightarrow S_2)}{2} \int_0^2 f(x) dx, & i = 1; \\ \frac{V(X \rightarrow S_i) + V(X \rightarrow S_{i+1})}{2} \int_0^2 f(x) dx, & 1 < i < K(X); \\ \frac{V(X \rightarrow S_{i-1}) + V(X \rightarrow S_i)}{2} \\ 1 - \sum_{z=1}^{K(X)-1} p(X, S_z), & i = K(X). \end{cases}$$

Для нахождения значений стационарных вероятностей рассматриваемой цепи Маркова при известных значениях переходных вероятностей, используются численные методы.

Напомним, что замирание при проигрывании оператором возникает в том случае, если буфер камеры содержит K_0 видеосегментов. Таким образом, для оценки вероятности замирания P_{stall} необходимо оценить вероятность нахождения цепи в состояниях с $b_{i_{K_0}} \neq 0$:

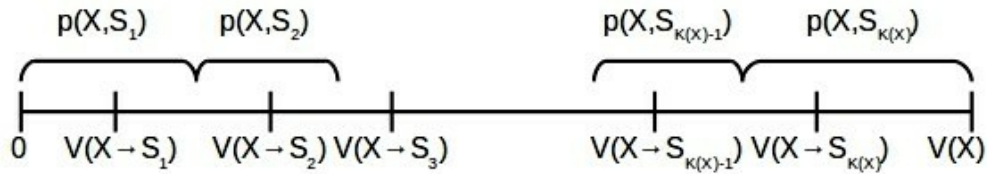


Рис. 2. Оценка вероятности перехода из состояния X

$P_{stall} = \sum_{i_{K_0} \neq 0} \hat{p}(j, i_1, \dots, i_{K_0})$. Среднее значение битрейта видеосегментов оценивается следующим образом: $B_{av} = \sum_X \hat{p}(X) B(X)$.

5.1. Оценка длительности времени, свободного от обслуживания заявок, в СМО с очередью типа M/D/1

Раздел посвящен вычислению функции плотности вероятности $e(t)$, выражающей вероятность того, что в СМО с очередью типа M/D/1 в течение произвольного интервала времени длительности D ровно t секунд будут свободны от обслуживания заявок. В общем случае занятость обслуживающего прибора СМО описывается ON/OFF-процессом со следующими состояниями:

- Время, свободное от обслуживания заявок, называемое свободными периодами. Их длительность описывается экспоненциальным распределением с параметром λ .
- Время непосредственного обслуживания заявок, поступивших в систему, называемое периодами занятости системы. Длительность периода занятости кратна τ , а вероятность B_i того, что в течение периода занятости будет обслужено ровно i заявок, описывается распределением Бореля [12]:

$$B_i = \frac{(i\tau\lambda)^{i-1}}{i!} e^{-i\tau\lambda}.$$

Стационарные вероятности обнаружения μ требований в СМО с очередью типа M/D/1 оцениваются согласно [13]:

$$\pi_\mu = \begin{cases} \bar{\pi}_\mu - \bar{\pi}_{\mu-1}, & \mu \geq 1; \\ \bar{\pi}_0, & \mu = 0; \end{cases} \quad \bar{\pi}_\mu = (1 - \lambda\tau) \sum_{i=0}^{\mu} e^{i\lambda\tau} \frac{(-i\lambda\tau)^{\mu-i}}{(\mu-i)!}, \quad \lambda\tau < 1.$$

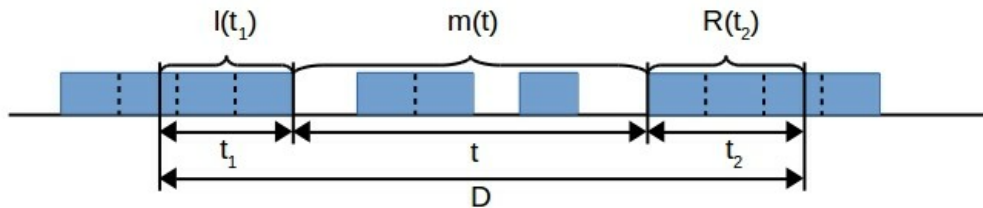


Рис. 3. Модель передачи в течение одного периода времени

Рассмотрим СМО на фиксированном интервале времени длительности D (без ограничения общности можно считать, что начальный момент времени равен 0). Для описания рассматриваемой системы удобно рассматривать 3 этапа эволюции процесса в течение фиксированного интервала времени (рис. 3):

1. Интервал в общем случае начинается с периода занятости (стартовый этап), начавшегося до момента времени 0. Для описания длительности этого этапа вводится функция плотности вероятности $l(t_1)$, описывающая вероятность того, что текущий период занятости закончится через t_1 секунд, $t_1 \geq 0$.
2. Заканчивается интервал в общем случае периодом занятости (завершающий этап), продолжающимся после окончания рассматриваемого интервала времени. Для описания длительности этапа используется вероятность $R(t_2)$ события, заключающегося в том, что произвольный период занятости будет продолжаться более t_2 секунд (отсчет времени начинается со старта данного периода).
3. Между стартовым и завершающим этапами система эволюционирует как неограниченная очередь типа M/D/1 (промежуточный этап).

Заметим, что стартовый и завершающий этапы могут отсутствовать (если в моменты времени 0 и D соответственно в СМО отсутствуют заявки).

Рассмотрим подробнее промежуточный этап эволюции. Как следует из описания этапов, промежуточный этап начинается и заканчивается периодами времени, свободными от обслуживания заявок. В остальном данный этап представляет собой уже описанный ON/OFF-процесс.

Лемма 1 ((Обобщение распределения Бореля)). *Рассмотрим СМО с очередью типа M/D/1 на протяжении интервала времени, равного t . Пусть в начале данного интервала времени система свободна от обслуживания заявок. Вероятность того, что в течение времени t поступят и будут полностью обслужены k заявок, вычисляется следующим образом:*

$$P_t^k = \frac{\lambda^k (t - k\tau)^{k-1}}{k!} e^{-\lambda t}.$$

При этом в конце рассматриваемого отрезка времени система снова свободна от обслуживания заявок.

Доказательство. Для удобства доказательства, введем дополнительную переменную $t' = t - k\tau$, обозначающую суммарную длительность периодов времени, свободную от обслуживания заявок. Появление и обслуживание заявок в течение t секунд можно рассматривать как прерывание отсчета свободного времени длиной t' секунд для обслуживания заявок. Каждое прерывание отодвигает завершение отсчета на τ (k прерываний отодвигают конец отсчета до момента t). При этом заявки должны приходить таким образом, чтобы отсрочить конец отсчета свободного времени до момента t , а после этого в системе не должно появиться ни одной заявки.

Рассмотрим передачу с произвольным числом заявок k . Пусть в течение t' секунд в системе появятся q заявок. К моменту времени t' конец отсчета свободного времени откладывается до $t' + q\tau$ секунды. Заметим, что в течение интервала $[t', t]$, $t - t' = k\tau$, должно появиться и быть полностью обслужено $r = k - q$ заявок.

Сначала рассмотрим систему на интервале $[t', t]$ и вычислим вероятность $P_{k\tau}^r$ того, что за время $k\tau$ поступят и будут полностью обслужены $r = k - q$ заявок. В работе [12] показано, что $P_{k\tau}^r = e^{-(k-r)\lambda\tau} (\lambda\tau e^{-\lambda\tau})^r \frac{k-r}{r!} k^{r-1}$.

Вернемся к оценке величины P_t^k через рассмотрение интервала времени $[0, t']$. Учитывая, что вероятность появления q заявок на интервале времени t' равна $\frac{(\lambda t')^q}{q!} e^{-\lambda t'}$, величина P_t^k оценивается следующим образом:

$$P_t^k = \begin{cases} \sum_{q=1}^k \frac{(\lambda t')^q}{q!} e^{-\lambda t'} P_{k\tau}^{k-q}, & k > 0; \\ e^{-\lambda t}, & k = 0. \end{cases}$$

Подставляя сюда выражение для $P_{k\tau}^{k-q}$, получаем следующую цепочку равенств:

$$P_t^k = e^{-\lambda t'} \sum_{q=1}^k \frac{(\lambda t')^q}{q!} e^{-k\tau\lambda} (\lambda\tau)^{k-q} \frac{q}{(k-q)!} k^{k-q-1} = \frac{\lambda^k t' (t' + \tau k)^{k-1}}{k!} e^{-\lambda(t'+k\tau)}.$$

Осталось заметить, что полученное выражение справедливо и в случае $k = 0$. Учитывая, что $t' = t - \tau k$, получаем искомое утверждение.

Замечание 1. Распределение Бореля является частным случаем полученного в лемме 1 распределения, а именно $B_i = P_{i\tau}^{i-1}$.

Теперь рассмотрим стартовый этап, представляющий собой период занятости системы, стартовавший до начала рассматриваемого отрезка времени.

Лемма 2. *Функция плотности вероятности $l(t_1)$, описывающая вероятность события, заключающегося в том, что стартовый этап закончится в момент времени t_1 , может быть вычислена следующим образом:*

$$l(t_1) = \pi_0 \delta(t_1) + \sum_{q=0}^{\lfloor \frac{t_1}{\tau} \rfloor} \frac{\pi_{q+1}}{\tau} P_{t_1}^{\lfloor \frac{t_1}{\tau} \rfloor - q} = \pi_0 \delta(t_1) + \hat{l}(t_1),$$

где δ — дельта-функция Дирака.

Доказательство. С вероятностью π_0 в момент времени 0 в системе массового обслуживания будут отсутствовать заявки, т.е. стартовый этап будет отсутствовать. Вклад в функцию $l(t_1)$ от этого события описывается членом $\pi_0 \delta(t_1)$.

Для того, чтобы стартовый период длился $t_1 \geq 0$ секунд, начиная с момента 0 ровно $\lfloor \frac{t_1}{\tau} \rfloor$ заявок должно быть непрерывно обслужено системой, рис. 4. Пусть $q + 1$ из $\lfloor \frac{t_1}{\tau} \rfloor$ заявок накапливаются в системе к моменту времени 0 (вероятность этого события равна π_{q+1}). Одна из этих заявок (появившаяся раньше остальных в системе) непосредственно обслуживается и с вероятностью $\frac{1}{\tau}$ до конца её обслуживания остается время, равное $t_1 - \lfloor \frac{t_1}{\tau} \rfloor \tau$. Остальные $\lfloor \frac{t_1}{\tau} \rfloor - q$ заявок должны появиться во время передачи очереди из $q + 1$ пакетов (данное событие эквивалентно событию, рассмотренному в лемме 1 при $t = t_1$ и $k = \lfloor \frac{t_1}{\tau} \rfloor - q$ веб-пакетов, т.е. его вероятность равна $P_{t_1}^{\lfloor \frac{t_1}{\tau} \rfloor - q}$). Суммируя по всем возможным значениям $0 \leq q \leq \lfloor \frac{t_1}{\tau} \rfloor$, получаем искомую оценку функции $l(t_1)$.

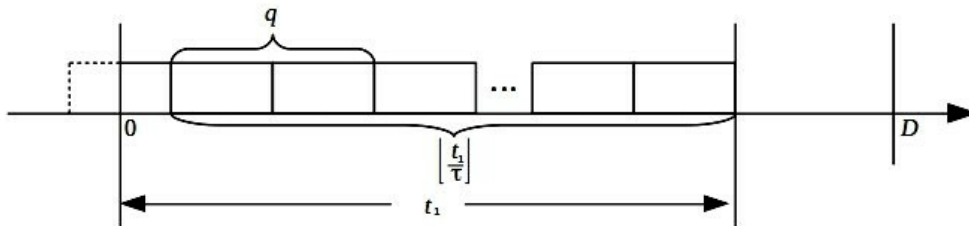


Рис. 4. Передача веб-пакетов, накопившихся в предыдущем периоде

Перейдем к рассмотрению завершающего этапа и оценке вероятности $R(t_2)$ того, что завершающий период занятости будет продолжаться более t_2 секунд.

Лемма 3. *Рассмотрим только что начавшийся период занятости СМО с очередью типа M/D/1. Вероятность $R(t_2)$ события, заключающегося в том, что данный период будет продолжаться более t_2 секунд, оценивается следующим образом:*

$$R(t_2) = 1 - \sum_{r=1}^{\lfloor \frac{t_2}{\tau} \rfloor} B_r,$$

где $B_r = \frac{(r\tau\lambda)^{r-1}}{r!} e^{-r\tau\lambda}$ — распределение Бореля [12].

Доказательство. Для того, чтобы только что начавшийся период занятости длился более t_2 секунд, необходимо, чтобы он состоял из не менее чем $\lfloor \frac{t_2}{\tau} \rfloor$ заявок (при этом одна заявка уже присутствует в СМО). Распределение Бореля B_r позволяет оценить вероятность того, что период занятости состоит из r заявок, тогда вероятность $R(t_2)$ того, что период содержит не менее $\lfloor \frac{t_2}{\tau} \rfloor$ заявок можно оценить, вычитая из 1 суммарную вероятность событий, в которых данное условие не выполняется.

Рассмотрев все этапы эволюции СМО с очередью типа M/D/1 на ограниченном интервале времени, перейдем к доказательству теоремы 1.

5.2. Доказательство теоремы 1

Введем в рассмотрение функцию плотности вероятности $m(t, \Pi(t))$ события, заключающегося в том, что между стартовым и завершающим этапами происходит событие, длящееся t секунд, вероятность которого равна $\Pi(t)$ (рис. 3). Заметим, что функция $R(t_2)$ описывает событие, которое происходит при условии появления заявки. Вероятность появления заявки после промежуточного этапа равна λ . Для оценки функции $m(t, \Pi(t))$ классифицируем всевозможные реализации эволюции рассматриваемой системы на интервале времени D по наличию или отсутствию стартового и завершающего этапов:

1. Все этапы присутствуют. В таком случае необходимо оценить вероятность того, что завершающий этап начинается через t секунд после стартового, длительность t_1 которого может принадлежать интервалу $[0, D - t]$. Вклад в функцию $m(t, \Pi(t))$ от таких событий оценивается членом $\int_0^{D-t} \hat{l}(t_1) \cdot \Pi(t) \cdot \lambda \cdot R(D - t - t_1) dt_1$.
2. Стартовый этап отсутствует, завершающий этап присутствует. Стартовый этап отсутствует с вероятностью π_0 . При этом завершающий этап должен начинаться в момент времени t . Вклад в функцию $m(t, \Pi(t))$ от таких событий оценивается членом $\pi_0 \Pi(t) \cdot \lambda \cdot R(D - t)$.
3. Стартовый этап присутствует, завершающий этап отсутствует. В таком случае стартовый этап должен продолжаться до момента времени $D - t$. Вклад в функцию $m(t, \Pi(t))$ от таких событий оценивается членом $l(D - t) \cdot \Pi(t)$.
 - Стартовый этап с конечной вероятностью $e_0 = 1 - \int_0^D l(x) dx$ может продолжаться на протяжении всего интервала времени D . В таком случае промежуточный этап отсутствует. Вклад в функцию $m(t, \Pi(t))$ от таких событий оценивается членом $e_0 \cdot \delta(t)$.
4. Стартовый и завершающий этап отсутствуют. Данные события происходят с конечной вероятностью и входят в функцию $m(t, \Pi(t))$ в виде взвешенных дельта-функций Дирака. Обсуждение их конкретного вида отложим до момента конкретизации функции $\Pi(t)$, обозначив данную поправку через $\chi(t)$.

Суммируя вышесказанное, функция $m(t, \Pi(t))$ может быть оценена так:

$$\begin{aligned} m(t, \Pi(t)) &= e_0 \cdot \delta(t) + \chi(t) + \pi_0 \Pi(t) \cdot \lambda \cdot R(D-t) + l(D-t) \cdot \Pi(t) \\ &\quad + \int_0^{D-t} \hat{l}(t_1) \cdot \Pi(t) \cdot \lambda \cdot R(D-t-t_1) dt_1 \\ &= e_0 \cdot \delta(t) + \chi(t) + M(t, \Pi(t)). \end{aligned}$$

Значения интегралов

$$I = \int_0^A \hat{l}(x) \cdot \lambda \cdot R(A-x) dx$$

и

$$e_0 = 1 - \int_0^D l(x) dx$$

могут быть получены в явном виде.

Функция $e(t)$, выражающая вероятность того, что в СМО с очередью типа M/D/1 в течение времени D ровно t секунд будут свободны от обслуживания заявок, может быть оценена через функцию $m(t, \Pi(t))$. Свободное время является частью промежуточного этапа эволюции. Для того, чтобы ровно t секунд были свободны от обслуживания, необходимо, чтобы промежуточный этап длился $t + j\tau$ секунд, $0 \leq j \leq \lfloor \frac{D-t}{\tau} \rfloor$, в течение которых приходят и полностью обслуживаются j заявок (вероятность такого события $\Pi(t + j\tau)$ по лемме 1 равна $\Pi(t + j\tau) = P_{t+j\tau}^j$). При суммировании функций $m(t, \Pi(t))$ учтем, что событие, заключающееся в том, что стартовый этап охватывает весь интервал времени D (описываемое членом $e_0\delta(t)$), необходимо учесть только один раз. Остаётся заметить, что события, описываемые функцией $\chi(t)$ и заключающиеся в отсутствии стартового и завершающего этапов, возможны в том случае, если на протяжении D секунд появляются и обслуживаются $0 \leq i \leq \lfloor \frac{D}{\tau} \rfloor$ заявок (конечная вероятность каждого из таких событий равна $\pi_0 P_D^i$, при этом $D - i\tau$ секунд остаются свободными от обслуживания заявок). Объединяя всё вышесказанное, получаем оценку функции $e(t)$ из выражения (1).

6. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Построенная математическая модель может применяться для оптимизации параметров адаптивных алгоритмов выбора битрейта, основанных на текущем состоянии буфера. Это может быть как просто объем данных, содержащихся в буфере, так и более детальная информация (количество видеосегментов в буфере, средний битрейт видеосегментов в буфере и т.д.). В данном разделе приводится пример использования математической модели для оптимизации кусочно-постоянной функции выбора битрейта, аргументом которой является объем данных в буфере камеры.

Опишем кусочно-постоянную функцию выбора битрейта $B = B\left(\frac{\omega}{\omega_{max}}\right)$, где ω и ω_{max} — текущий и максимальный размеры очереди видеосегментов на камере. Для этой цели определим набор отсчетных значений битрейтов $\mathcal{B} = [b_1, \dots, b_G]$ и отношений текущего размера очереди видеосегментов на камере к максимальному $\mathcal{C} = [C_1, \dots, C_{G-1}]$. Тогда кусочно-постоянную функцию выбора битрейта можно описать следующим образом:

$$B = \begin{cases} B_1, & \frac{\omega}{\omega_{max}} \leq C_1; \\ B_i, & C_{i-1} < \frac{\omega}{\omega_{max}} \leq C_i; \\ B_G, & \frac{\omega}{\omega_{max}} > C_{G-1}. \end{cases}$$

Параметрами кусочно-постоянной функции является набор отсчетных значений битрейта \mathcal{B} и отношения текущего размера очереди видеосегментов к максимальному C . Результатом оптимизации является кусочно-постоянная функция выбора битрейта, описываемая набором \mathcal{B}_{opt} и C_{opt} , применение которой максимизирует среднее значение битрейта $B_{av}(\mathcal{B}_{opt}, C_{opt})$ среди всех рассмотренных кусочно-постоянных функций, описываемых наборами \mathcal{B}' , C' , при использовании которых удовлетворяются ограничению по вероятности замирания на операторе $P_{stall}(\mathcal{B}', C') \leq \theta$:

$$(\mathcal{B}_{opt}, C_{opt}) = \arg \max_{\mathcal{B}', C'} \{B_{av}(\mathcal{B}', C') : P_{stall}(\mathcal{B}', C') \leq \theta\}. \quad (2)$$

Для рассмотрения различных кусочно-постоянных функций определим допустимые значения кусочно-постоянных функций в виде конечного множества допустимых битрейтов B_{pool} и допустимые отсчетные значения отношения текущего размера очереди видеосегментов к максимальному в виде конечного множества C_{pool} . На основе множеств B_{pool} и C_{pool} строятся произвольные невозрастающие кусочно-постоянные функции выбора битрейта, для которых множество \mathcal{B}' состоит из G элементов множества B_{pool} (мы считаем величину G постоянной), а множество C' состоит из $G - 1$ элементов множества C_{pool} . Для каждой из полученных функций выбора битрейта оценивается средний битрейт видеосегментов и вероятность замирания при проигрывании на операторе, получаемые при использовании алгоритма выбора битрейта, основанного на данной функции. В завершении согласно (2) определяется оптимальная комбинация $\mathcal{B}_{opt}, C_{opt}$, описывающая оптимальную кусочно-постоянную функцию выбора битрейта.

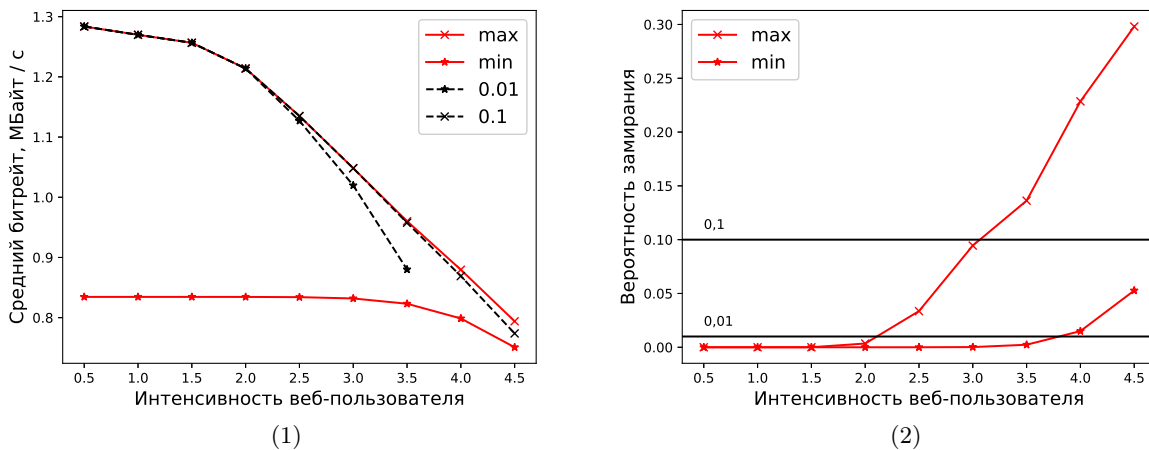


Рис. 5. Результаты оптимизации алгоритма выбора битрейта, основанного на кусочно-постоянной функции, для (1) среднего битрейта видеосегментов и (2) вероятности замираний на операторе

На рис. 5 приводятся результаты оптимизации, проведенной для сети с $N = 1$ камерой видеонаблюдения и начальной задержкой воспроизведения на операторе, равной $K_0 = 3$. При проведении оптимизации мы рассматривали следующие значения используемых величин: $B_{pool} = [0, 55; 0, 7; 0, 85; 1, 0; 1, 15; 1, 3; 1, 4]$ Мбайт/с, $C_{pool} = [0, 15; 0, 3; 0, 45; 0, 6; 0, 75; 0, 9]$, $G = 5$. При вычислениях принималось $N_q = 20$.

Рис. 5 демонстрирует максимальный (обозначается через max) и минимальный (обозначается через min) средний битрейт видеосегментов, полученные в результате перебора различных кусочно-постоянных функций выбора битрейта, а также значения среднего битрейта (обозначаются через 0,01 и 0,1), получаемые в результате применения оптимального в смысле уравнения (2) алгоритма управления битрейтом для значений ограничения θ по вероятности

замираний, равных $\theta_1 = 0,01$ и $\theta_2 = 0,1$. Например, для $\lambda = 3,5$ и ограничения $\theta_1 = 0,01$ оптимизированная версия алгоритма описывается множествами $B_{opt} = [1, 4; 1, 15; 0, 85; 0, 7; 0, 55]$ Мбайт/с и $C_{opt} = [0, 15; 0, 3; 0, 45; 0, 6]$.

При $\lambda = 4,5$ максимальный средний битрейт превосходит средний битрейт оптимизированной версии алгоритма для значения ограничения $\theta_2 = 0,1$ не более чем на 5%. Однако вероятность замираний для оптимизированного алгоритма не превышает 10%, в то время как значение данной величины для алгоритма с максимальным битрейтом превосходит 30%.

7. ЗАКЛЮЧЕНИЕ

В данной работе представлен способ организации системы видеонаблюдения, работающей в гетерогенных беспроводных централизованных сетях в том случае, когда пользователи, отличные от камер видеонаблюдения, обслуживаются с абсолютным приоритетом, т.е. являются основными пользователями сети. Камеры записывают видеопоток и осуществляют передачу записанных данных на удаленный сервер через базовую станцию сети. Оператор, расположенный на сервере, проигрывает получаемый видеопоток в режиме реального времени. Из-за присутствия в сети приоритетных пользователей объём частотно-временных ресурсов, доступных для передачи видеокamer, постоянно меняется, поэтому для обеспечения высокого качества видеопотока используются различные адаптивные алгоритмы выбора битрейта.

В работе разработана математическая модель передачи видеосегментов на удаленный сервер, позволяющая оценивать средний битрейт видеопотока и вероятность замираний при проигрывании оператором. Данная модель может быть использована для оптимизации параметров различных адаптивных алгоритмов выбора битрейта, основывающихся на информации о буфере камеры, например, [5] или [14]. На примере кусочно-постоянной функции выбора битрейта продемонстрирована эффективность такой оптимизации.

СПИСОК ЛИТЕРАТУРЫ

1. *Khorov E., Kiryanov A., Lyakhov A., Bianchi G.* A Tutorial on IEEE 802.11ax High Efficiency WLANs. IEEE Communications Surveys and Tutorials (In press). DOI: 10.1109/COMST.2018.2871099
2. *Wang Y.* Survey of Objective Video Quality Measurements // Computer Science Faculty Publications, Department of Computer Science, Worcester Polytechnic Institute, 2006.
3. *Bankov D., Khorov E., Lyakhov A.* Fast Quality Assessment of Videos Transmitted over Lossy Networks // Proceedings of International Conference on Engineering and Telecommunication (EnT), 2014.
4. *Romero L.* A Dynamic Adaptive HTTP Streaming Video Service for Google Android, Royal Institute of Technology (KTH), 2011.
5. *Spiteri K., Urgaonkar R., Sitaraman R.* BOLA: Near-Optimal Bitrate Adaptation for Online Videos // Proceedings of IEEE INFOCOM, 2016.
6. *Mahonen P.* Wireless video surveillance: system concepts // Proceedings of International Conference on Image Analysis and Processing, 1999.
7. *Gualdi G., Prati A., Cucchiara R.* Video Streaming for Mobile Video Surveillance // IEEE Transactions on Multimedia, 2008. V. 10. P. 1142–1154.
8. *Rosu V., Marcu M.* Real time adaptive video streaming // Proceedings of the 12th International Conference on Intelligent Computer Communication and Processing (ICCP), 2016.
9. *Khorov E., Krasilov A., Liubogoshchev M., Tang S.* SEBRA: SAND-Enabled Bitrate and Resource Allocation algorithm for network-assisted video streaming. // Proceedings of WiMob 2017, Rome, Italy, 2017.

10. *Akyildiz I. F., Khorov E., Kiryanov A., Kovkov D., Krasilov A., Liubogoshchev M., Shmelkin D., and Tang S.* xStream: A New Platform Enabling Communication between Applications and the 5G Network // Proceedings of IEEE Globecom - CA5GS, Abu Dhabi, UAE, 2018
11. Cisco Visual Networking Index: Forecast and Methodology, 2016–2021. Cisco White Paper, 2017.
12. *Tanner J.C.* A derivation of the Borel distribution // *Biometrika*, 1961. V. 48. P. 222–224.
13. *Pack C.D.* The Output of an M/D/1 Queue // *Operations Research*, 1975. V. 23. No. 4. P. 750–760.
14. *Rahman W., Chung K.* Buffer-Based Adaptive Bitrate Algorithm for Streaming over HTTP // *KSI Transactions on Internet and Information Systems*, 2015. V. 9. No. 11.

Mathematical Model of a Network Slicing Approach for Video and Web Traffic

Khorov E.M., Lyakhov A.I., Zhirnov N.S.

The paper studies adaptive bitrate control algorithms used to control the quality of the video stream from video surveillance cameras to remote servers through a live wireless network with video and web traffic. In such a scenario, bitrate control algorithms may select an appropriate video bitrate taking into account the current state of the camera buffers. The paper introduces a mathematical model of video streaming, which allows evaluating the performance of different bitrate control algorithms.

KEYWORDS: network slicing, xStream, video surveillance system, web traffic, video traffic.