An Approach to the Modeling of the P2P Streaming Network Based on Peers' Geolocation and Activity

Yuliya Gaidamaka, Ivan Vasiliev,
Konstantin Samouylov
Applied Probability and Informatics Department
Peoples' Friendship University of Russia
Moscow, Russia
E-mail: ygaidamaka@sci.pfu.edu.ru, iuvasiliev@gmail.com,
ksam@sci.pfu.edu.ru

Sergey Shorgin
Institute of Informatics Problems,
Federal Research Center
"Computer Science and Control"
of the Russian Academy of Sciences
Moscow, Russia
E-mail: sshorgin@ipiran.ru

Abstract— **In this paper, an approach to the simulation of peer-to-peer (P2P) live streaming network is presented. The simulation model considers peers' geolocation and their daily activity, time lags between the video server and a peer, lags between peers, collisions, and three types of selection strategies: neighbor selection strategy, peer selection strategy and chunk selection strategy. Geolocation is considered as distribution of users by time zones and the users' daily activity is defined as the distribution of the number of online users by the hour of the day. Initial data for the geolocation of users and their daily activity are taken from known Internet sources. The aim of this study is to show how the parameters of the geolocation and users' daily activity affect the performance of the network. To do this, we compare the download probability obtained analytically without geolocation and daily activity with the results of simulation. Preliminary numerical analysis carried out for the case of the Rarest First chunk selection strategy, shows difference in results of up to 30%.**

Keywords—P2P live streaming network; P2P tecnology; playback continuity; playback lags; positive analysis

I. INTRODUCTION

Peer-to-peer (P2P) technology is used by major service providers in the market of online TV, such as BBC iPlayer, Zattoo, PeerCast, Pulse, QQLive and many others [1][2]. A fairly complete overview on P2P technical aspects with an extensive list of references was done by Yue et al. [3]. One of the main advantages of P2P is high performance-cost ratio, that allows commercial companies to minimize the costs for technical equipment. Furthermore, P2P technology not only enables efficient use of network resources, but also reduces the load on the server that is the source of the video data. Thus, due to high load balancing, there is no need to install additional servers, as well as in the growth of network bandwidth to handle a large number of users. Other advantages of P2P technology are scalability and high network robustness: it gives an opportunity to have a trouble-free operation of at least one video server to provide services for all users at the acceptable level of Quality of Experience (QoE).

However, P2P networks also have disadvantages. The main shortcomings include insufficient level of security, relatively long start-up latency and data transmission delay, inter-peer playback lag, playback discontinuity. While information security is not required, the goal is to minimize the transmission delay, and hence lack of data on the network. To solve the optimization problem for these performance measures, mathematical models and simulators should be developed. Known mathematical and simulation models for streaming P2P-network pay much attention to investigation of the buffering mechanism [11][12][13][14][15][16] and usually take into account lags [7][10]. A lag means the delay of data transmission from the server to the user, as well as the data transmission delays between peers (inter-peer playback lag). These models allow to carry out a qualitative analysis of the key performance measures of video streaming P2P networks - the probability of playback continuity, that is the probability of watching video with no pauses, and the probability of chunk availability. In previous works [8][9][10] these models allowed to formulate and to obtain the solution of the optimization problems in a choice of the selection strategies [13][14] that are used in a network.

Each of the models in the aforementioned sources reflects one or more features of the analyzed network and enables to estimate and optimize the corresponding network parameters and strategies. For example, the problem is to estimate which chunk selection strategy is better: Rarest First, where a peer downloads the rarest chunk in the network, or Greedy, where a peer downloads the most popular chunk in the network, or Mixed strategy that combines these two strategies. Previously, [8][9][10] we build a stand-alone analytical model to select the optimal strategy. Having analyzed results, we have come to the conclusion that the Rarest First strategy is better if we want to increase probability of playback continuity, and the Greedy chunk selection strategy will most likely be used if we want to reduce the startup latency. Finally, it is better to use Mixture strategy in real networks, where both quality parameters, playback continuity and startup latency, are important to the users. Using a particular model, one can only examine how the behavior of peers in the network influence its target characteristics, but cannot evaluate what we get in real life when we measure these characteristics in a commercial P2P network. In order to get closer to reality and assess the adequacy of the individual models, to give at least the recommendations against some of the problems before they occur in a real P2P networks, we propose a new model that can help to avoid troubleshooting. The presented model takes into account the geographical location of each peer,

Copyright (c) IARIA, 2016. ISBN: 978-1-61208-521-0

remoteness of peers from the server and from each other and their daily activity. In contrast to recent results (see, i.e. [9][10]), in this paper, the problem is solved taking into account the distribution of users across time zones.

Previously, we have studied various models of streaming P2P networks that have focused on the study of the optimal downloading strategy with the criterion of maximizing the probability of playback continuity of the video stream [9][10]. Then, the model was modified to take into account the users' behavior by introducing the probability of arrival of new users, as well as the probability of their leaving the network. [10][12]. The negative effect of peers churns on playback continuity was observed and investigated. The analytical model [8][9] also gives the correct understanding of lags' impact on the network performance. Usually, it is believed that the lag affects only the shift of the data location in the buffers of different users. We have shown that there is at least one more important aspect of lags. It is that the remote user is unable to receive a chunk within a number of time slots and thus play the video stream continuously. Due to the problem complexity, each main feature of the P2P network was analyzed in the known models separately. In this study, we present an approach to simulation that allows to combine in a single model the basic aspects of functioning of the streaming P2P-networks.

This paper is organized as follows. In Section II, a simulation model of video data distribution in a P2P live streaming network with buffering mechanism, geolocation and daily peers' activity is proposed. Also, the detailed algorithm of chunk exchange between buffers of peers in P2P live streaming network is determined and the main performance measures are defined. In Section III, the numerical analysis and case study is performed. The conclusion of this paper is presented in Section IV.

II. SIMULATION MODEL

In this section, a simulation model of video data distribution in a P2P live streaming network with buffering mechanism is proposed. A previously developed model [12] was improved by taking in consideration buffer selection strategies [13][14], peers' geolocation [15] and activity [16]. In contrast to the previous model, in this paper, besides chunk selection strategy, two more strategies were considered - neighbor selection strategy and peer selection strategy. The choice of strategies has a significant effect on the P2P-network performance measures, including the probability of playback continuity; the probability of chunk availability; the probability of chunk selection, and also the probability of collision – a situation when a peer cannot download a chunk because the target peer does not have enough capacity for uploading. There are various definitions of collision in the P2P network. In the present model, it is considered that, in the case of collisions, the peer that requested the chunk will not receive it in the current time slot.

We consider the basic model of a P2P network with N users and a single server, transmitting only one video stream, which we developed in [12]. The process of video stream playback is divided into time slots, the length of each time slot corresponding to the playback time of one chunk. Each user has a buffer designed to accommodate M+1 chunks, where the buffer positions are numbered from 0 to M: 0-position is to store the freshest chunk just received from the server, the other m-positions, $m = 1, ..., M-1$, are to store chunks, already received during the past time slots or that will be downloaded in the coming time slots. The buffer M-position is to store the oldest chunk that will be moved out from the buffer for playback during the next time slot. In Figure 1, a model of peer's buffer is illustrated. Thus, a state of n-th user is represented in the form of a vector $\boldsymbol{x}(n) = \left(x(n,0), x(n,1), .., x(n,M)\right)$ where $x(n,m) = 1$ if the n-th peer has a chunk at the buffer's position m, and $x(n,m) = 0$ otherwise.

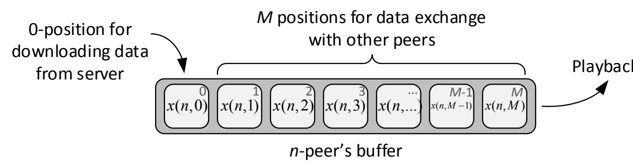

Figure 1. Model of n-th peer's buffer

Initially, a set of original parameters for each of N users is determined. They are upload U and download D rates, value of a lag LAG, and a set of neighbors $\mathbf{B}$ – the users from whom downloading is acceptable. A lag is the number of time slots between sending and receiving a chunk, thus, lag reflects the quantitative characteristics of a chunk delay. The algorithm works in such a way that, within a group, the user selects a neighbor to download data from using the criterion of minimum lag between the neighbors, regardless of their time zones distribution. Here is a simple example. Suppose that Peer 1 is in Poland and its neighbor, Peer 2, is located in Moscow, i.e. they are from different time zones. Suppose that Peer 3 is located in Angola, in the same time zone as Peer 1, and they are also neighbors. In this example, Peer 1 selects Peer 2 because they have the smallest lag, although Peer 2 is located in a different time zone. The set of neighbors for each user is formed according to the neighbor selection strategy depending on upload and download rates and lags. The neighbor selection is one of the target function parameters for optimization problems.

The algorithm of peers' actions at each time slot is described below according to the protocol of data distribution in P2P live streaming networks.

1) At the beginning of each time slot, the chunk at the M-position of the buffer is going to be played if it is present. Video data in the buffers shifts one position towards the end of the buffer. 0-position is nulled.
2) The server randomly chooses a peer and loads the newest chunk to 0-position of its buffer.
3) Each peer that was not chosen by the server selects a target peer from the set of neighbors to download a chunk during the current time slot. Target peer's selection is carried out in accordance to the peer selection strategy.

Copyright (c) IARIA, 2016. ISBN: 978-1-61208-521-0

4) If collision takes place, the peer gets nothing during the current time slot. Otherwise, it selects a chunk to download according to the chunk selection strategy. If there is an available chunk to download, the loading starts. Otherwise, the peer gets no chunks during the current time slot.

It should be noticed that, in a real network, each peer is able to join the video stream and to disjoin it at any time slot of modelling and at any step of the algorithm, but we specify it by saying that the peers join or disjoin immediately after the first step of the algorithm.

The first difference from the previous model [12] is that in the presented model three strategies were considered: neighbor selection strategy, peer selection strategy, and chunk selection strategy. The second difference is that the presented model takes into account peers' geolocation [15] and twenty-four hours peers' activity [16].

The number of peers in the network is not constant. Every peer stays online a random amount of time each day, with the average value of $0<HO<86\,400=24h*60min*60s$:
HoursOnline(n)~P(HO), n=1,…,N.
Here, 86 400 is the number of time slots when modelling one astronomical day with one second as one time slot: $24h\times 60min\times 60s$=86 400 time slots.

Let us introduce the parameter of users' activity, which reflects the behavior of peers in the streaming network:
UserActivity(n)~RAND(1..UA), n=1,…,N; UA>1,
where UA is the maximal number of peers joining the network within a day.

This parameter shows how often a peer joins the network, disjoins from it and switches channels. Here $UserActivity(n) = 1$ means that within a day n-th peer once came to a network and was online during the random time $HoursOnline(n)$ without switching to other channels. $UserActivity(n) = i, 1 \le i \le UA$, means that the n-th peer joined the network i times per day including switching channels, and each session lasted exactly $HoursOnline(n)/i$ time slots. Thus, e.g. if $UserActivity(n) = 100$ and $HoursOnline(n) = 15\,000$ the n-th peer per day during 15 000 time slots (seconds) carries out 100 sessions of 150 time slots each.

Figure 2 presents the distribution of the number of online peers versus the time of the day [15]. Here, the number of users of the mobile applications is depicted by the solid line. The dashed line represents the number of Internet users while the dotted line shows the number of TV watchers. To simplify the modeling process in this research, we investigate the influence of TV watchers' behavior only. The graph shows that the peak of the online users is between 6 p.m. and 12 a.m. while the minimum number of TV watchers is from 2 a.m. to 7 a.m.

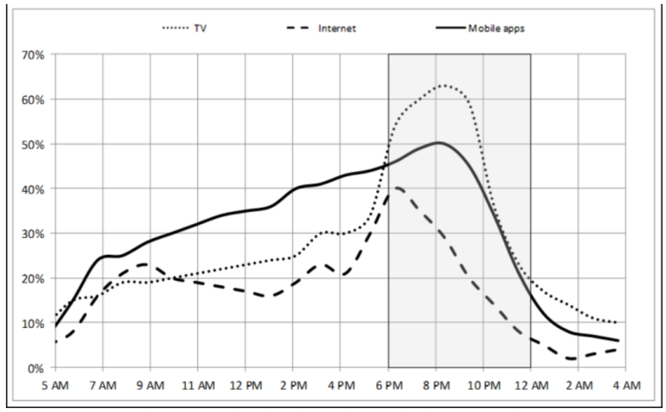

Figure 2. Distribution of users in the network

In accordance with the distribution in Figure 2, parameters $UserActivity = (UserActivity(n)), n = 1, \dots, N$, and $HoursOnline = (HoursOnline(n)), n = 1, \dots, N$, correspond to randomly generated intervals when peers are online. Let $TimeOnline = (TimeOnline(n,t)), n = 1, \dots, N;\ t = 1, \dots, T$, be a binary matrix of the $N \times T$ size, where T is the number of time slots in the simulation. The matrix indicates time slots when peers are online: $TimeOnline(n,t) = 1$ if the n-th peer at the t-th time slot is online, and $TimeOnline(n,t) = 0$ otherwise. Thus, if $TimeOnline(n,t) = 1$ and $TimeOnline(n,t+1) = 0$, then the n-th peer left the network at the $(t+1)$-th time slot, and $TimeOnline(n,t) = 0$ and $TimeOnline(n,t+1) = 1$ say that the n-th peer joined the network at the $(t+1)-th$ time slot.

Peers churns significantly influence the key performance measures. So, when a new peer joins the network it still has no data for exchange with other peers, but it uses other peers' resources to download content. Similarly, when a peer disjoins the network, it stops to participate in distribution of already downloaded video chunks. For a proper peers churns simulation, it is important to take into account the distribution of users by time zones. In this paper, the total of *N* peers in the network are divided in a random way across time zones according to the distribution shown in Figure 3 [16]. One can see that the majority of the users is located in -5, +1, and +8 time zones, which include the USA, Canada, Europe, and China - the most populated and technologically developed regions. In the model, splitting peers across time zones allows to reduce the probability of the global splashes corresponding to mass connections and disconnections of users.

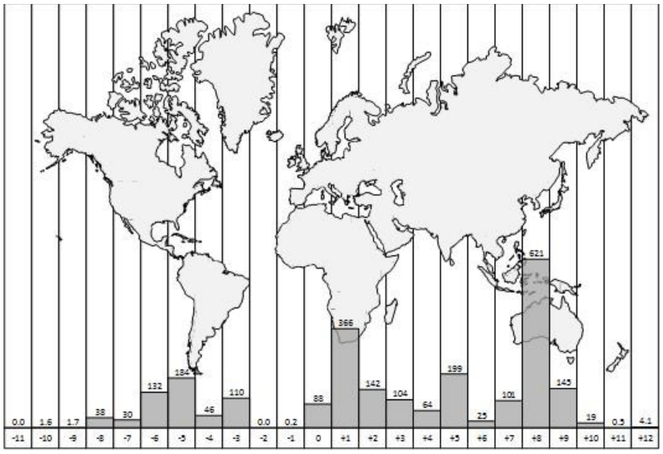

Figure 3. Distribution of users across the time zones

III. Numerical Analysis

In this paper, the aim of numerical analysis is to compare results of the models that do not consider distribution of users across time zones (we call it the basic model, [8][9][10]) to the results which are obtained by means of the model described above. We analyzed the network with $N = 300$ users and the size of the users' buffer $M = 40$. The time of modeling is $T = 1\ 000\ 000$ time slots, which corresponds to about 12 days. The basic model assumes values of lags equal to 0, 10, 20 between groups of neighbors. The Rarest First (RF) has been chosen as a chunk selection strategy, and neighbors, as well as the target user, were selected randomly.

As seen in Figures 4 and 5, the results of the simulation with splitting users across time zones (dashed line) qualitatively repeat the results of the basic model. The nature of the behavior of curves, including flexes in points of $m = 10$ and $m = 20$ is described in [9][10]. The probability of playback continuity for the model with time zones is much lower because the delivery of chunks takes, on average, more than one time slot. Note that, in Figure 4, value $p(40)$ corresponds to the probability of playback continuity. Numerical analysis shows that the basic model gives a very rough upper bound of the performance measures compared to the model with splitting users across time zones. On the other hand, the simulation model shows unsatisfactory (less than 70%) value of the probability of playback continuity. This means that it is necessary to use some other combinations of selection strategies, which will be the subject of further study. It is also necessary to explore a new way of peer's communication to improve the playback continuity, which is a key performance of the network.

IV. Conclusion and Future Works

We try to construct a model of a P2P streaming network, which is the most approximate to the reality. Preliminary numerical analysis showed that more research is necessary to find the optimal strategies; buyout will improve the quality of service parameters of streaming P2P-network. To do this, it is necessary to formulate the appropriate optimization problems, to find ways to solve them, even if numerical, and conduct computer experiments using the simulator described above. It is already clear that it is necessary to modify the strategies used in streaming P2P-network.

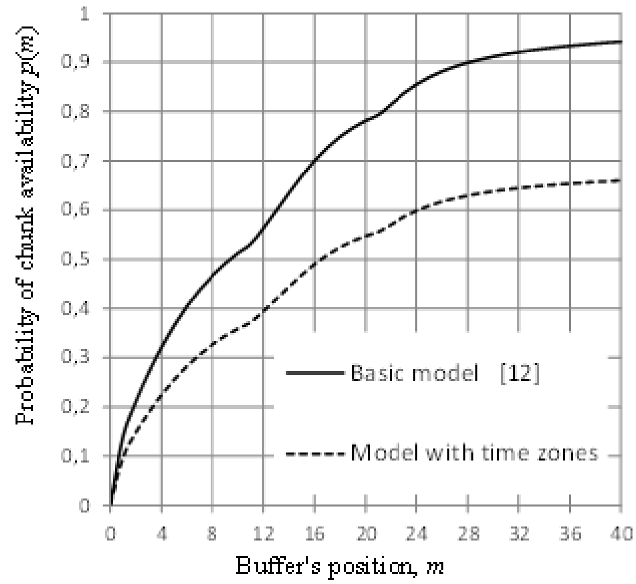

Figure 4. Probability of chunk availability at the buffer's positions. Value $p(40)$ corresponds to the probability of playback continuity.

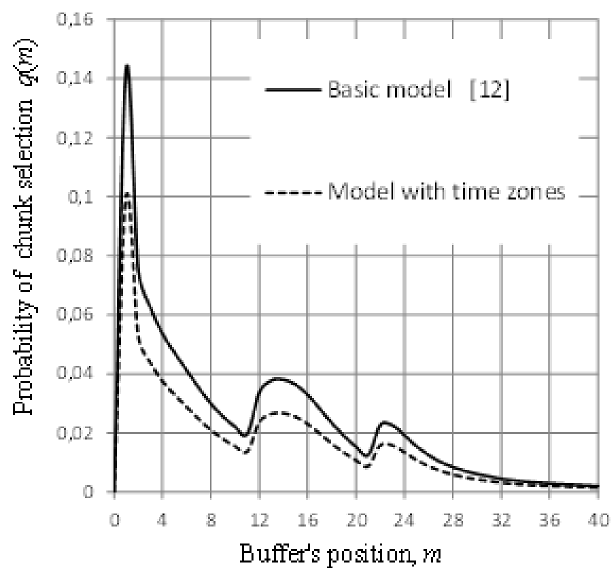

Figure 5. Probability of chunk selection to download at the buffer positions.

ACKNOWLEDGMENT

The reported study was partially supported by the RFBR, research project No. 14-07-00090.

REFERENCES

[1] X. Shen, H. Yu, J. Buford, and M. Akon, “Handbook of Peer-to-Peer Networking. Springer,” pp. 1421, 2010.

[2] S.M. Thampi “A Review on P2P Video Streaming,” arXiv preprint arXiv:1304.1235, 2013.

[3] G. Yue, N. Wei, J. Liu, X. Xiong, and L. Xie, “Survey on Scheduling Technologies of P2P Media Streaming,” Journal of Networks, Vol. 6, No. 8, pp. 1129-1136, August 2011.

[4] F. Baccelli, F. Mathieu, and I. Norros, “Performance of P2P networks with spatial interactions of peers. Networks and Telecommunications Networks, Systems and Services, Distributed Computing,” Equipes-Projets GANG, TREC, Centre de recherche INRIA Paris, Rapport de recherche, n°7713, pp. 1-23, August 2011.

[5] L. Ying, R Srikant, and S Shakkottai, “The Asymptotic Behavior of Minimum Buffer Size Requirements in Large P2P Streaming Networks”, Information Theory and Applications Workshop (ITA), pp. 1-6, 2010.

[6] Y. Zhou, M. Dah, Chiu, and J. C. S. Lui, “A Simple Model for Analyzing P2P Streaming Protocols,” Proc. of IEEE Int. Conf. IN Network Protocols (ICNP 2007), pp. 226-235, Oct. 19, 2007.

[7] Y. Zhao and H. Shen, “A simple analysis on P2P streaming with peer playback lags,” Proc. of the 3rd IEEE Int. Conf. on Communication Software and Networks (ICCSN 2011), Xi’an, China, pp. 396-400, May 2011.

[8] A. Adamu, Yu. Gaidamaka, and A. Samuylov, “Discrete Markov Chain Model for Analyzing Probability Measures of P2P Streaming Network,” Lecture Notes in Computer Science. Germany, Heidelberg: Springer. Vol. 6869, pp. 428-439, 2011.

[9] Yu. Gaidamaka and A. Samuylov, “Analytical Modeling of Playback Continuity in P2P Streaming Network with Latest First Download Strategy”, Lecture Notes in Computer Science, Germany, Heidelberg, Springer, Vol. 8121, pp. 363-370, 2013.

[10] Yu. Gaidamaka, I. Vasiliev, A. Samuylov, K. Samouylov, and S. Shorgin. “Simulation of Buffering Mechanism for Peer-to-Peer Live Streaming Network with Collisions and Playback Lags”, Proc. of the 13th Int. Conf. on Networks (ICN), Nice, France, pp. 86-91, 2014.

[11] N. Sluijs, T. Wauters, C. Develder, F. Turck, P. Demeester, B. Dhoedt, “Using topology information for quality-aware Peer-to-peer multy layer video streaming,” Int. J. Commun. System, Published online:, DOI: 10.1002/dac.2919, January 2015.

[12] Z. Li, M.A. Kaafar, K. Salamatian, G. Xie “User Behavior Characterization of a Large-Scale Mobile live Streaming System,” WWW 2015 Companion, Florence, Italy. ACM 978-1-4503-3473-0/15/05, May 2015.

[13] B. Fan, D. Andersen, M. Kaminsky, and K. Papagiannaki, “Balancing Throughput, Robustness, and In-Order Delivery in P2P VoD,” In Proc. ACM CoNEXT, Dec. 2010.

[14] P. Romero, M.E. Bertinat, D. Padula, P. Rodr´ıguez-Bocca, and F.R. Amoza, “A cooperative model for multi-class peer-to-peer streaming networks,” Proc. of the 1st Int. Conf. on Operations Research and Enterprise Systems (ICORES), pp. 274-282, 2012.

[15] How many internet users are there in your time zone, http://www.slideshare.net/pingdom/how-many-internet-users-are-there-in-your-time-zone [retrieved April 2016].

[16] F. Alan, “Study shows that in-app advertising is more effective than television or internet ads,” http://www.phonearena.com/news/Study-shows-that-in-app-advertising-is-more-effective-than-television-or-internet-ads_id22556, [retrieved April 2016].