

Why are Peers Less Stable in Unpopular P2P Streaming Channels?

Zimu Liu, Chuan Wu, Baochun Li, and Shuqiao Zhao*

Abstract. In large-scale P2P live streaming systems, it is shown that peers in an unpopular channel often experience worse streaming quality than those in popular channels. In this paper, by analyzing 130 GB worth of traces from a large-scale P2P streaming system, UUSee, we observe that a large number of “unpopular” channels, those with dozens or hundreds of concurrent peers, tend to experience inferior streaming quality. We also notice a short lifespan in these channels, which further exacerbates streaming quality. To derive useful insights towards improving streaming performance, we seek to thoroughly characterize important factors that may cause peer volatility in unpopular channels. Specifically, we conduct a comprehensive statistical analysis on the impact of various factors on peer lifespan, using survival analysis techniques. We found that the initial buffering level, the variance of peer indegree, and the peer joining time all have important effects on the lifespan of peers.

Keywords: Measurement, Peer-to-Peer, Live Streaming

1 Introduction

Real-world live P2P multimedia streaming systems have been successfully deployed in the Internet at a large scale, with hundreds of channels and hundreds of thousands of users at any given time. With the large number of concurrent channels, practical experiences have revealed the widely uneven distribution of peers across different channels: there may be thousands of concurrent users watching a popular channel, and no more than a few hundred of peers in an unpopular channel. These unpopular channels, usually representing the majority of the available channels in the streaming system, generally experience lower streaming quality, as compared to large popular channels. While many research efforts have been made to guarantee the performance of large popular channels, *e.g.*, to accommodate a flash crowd scenario where a large number of peers join in a short period of time, little attention has been devoted to the improvement of streaming quality in unpopular channels.

* Zimu Liu and Baochun Li are affiliated with the Department of Electrical and Computer Engineering, University of Toronto. Their email addresses are {*zimu, bli*}@*eecg.toronto.edu*. Chuan Wu is affiliated with Department of Computer Science, The University of Hong Kong. Her email address is *cwu@cs.hku.edu*. Shuqiao Zhao is affiliated with Multimedia Development Group, UUSee Inc. His email address is *zhaosq@uusee.com*.

In this paper, we focus on unpopular channels in large-scale P2P streaming systems. Using more than 130 GB worth of run-time traces from hundreds of streaming channels in a large-scale real-world P2P live streaming system, UUSee [1] (among the top three commercial systems in mainland China, along with PPLive and PPStream), we have investigated the distribution of peer population and streaming quality across different channels, and observed inferior streaming qualities that are empirically experienced by unpopular channels. We have further discovered a short peer lifespan (severe peer volatility) in the unpopular channels, which reveals a less than desirable situation that may lead to a downward spiral of peer population: On one hand, the low streaming quality in an unpopular channel may lead to short peer stay in the channel; on the other hand, the more severe peer churn further exacerbates the streaming quality of existing peers. To promote peer stability for a better streaming quality, it is critical to thoroughly understand and characterize the important factors that may have caused the peer volatility in unpopular channels.

Towards this objective, we conduct a comprehensive and in-depth statistical analysis using the UUSee traces. Our objective is very clear: we wish to identify critical performance metrics as risk factors that may influence the lifespan of peers, in order to derive useful insights towards the improvement of stability of peers in unpopular channels. To achieve this, we have parsed and imported all run-time traces into a database, where we apply survival analysis techniques such as the Cox proportional hazards model and the Mantel-Haenszel test to discover such influential factors. We have found that the initial buffering level, peer indegree and peer joining time all have important effects on the lifespan of individual peers. We are able to discover these influential performance factors from the real-world traces, which has not been possible in the existing literature. Based on the results of our regression analysis, we have derived a number of useful insights to guide the design of better P2P streaming protocols that promote the stability of peers in unpopular channels.

The remainder of this paper is organized as follows. In Sec. 2, we present our research methodologies with respect to collecting and parsing UUSee run-time traces. In Sec. 3, we show our survival analysis of the traces to identify influential factors on the peer lifespan. In Sec. 4, we model the impact of influential factors using the Cox regression model. In Sec. 5, we discuss the implication of our model and its usage in promoting peer stability. We discuss related work and conclude the paper in Sec. 6 and Sec. 7, respectively.

2 Trace Overview: A First Glance at Unpopular Channels

2.1 Collecting Real-World Traces

Starting from 2006, we have been monitoring the performance statistics of a real-world commercial P2P streaming system, offered by UUSee Inc. [1]. Similar to all current-generation mesh pull-based P2P streaming protocols, UUSee's streaming protocol design is based on the principle of allowing peers to serve each other by exchanging blocks of data, that are received and cached in their

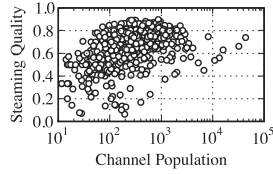
local playback buffers. To dynamically monitor the entire system, we have implemented detailed measurement and reporting capabilities within the UUSee client application. Each peer collects a set of its vital statistics, and encapsulates them into “heartbeat” reports to be sent to the tracking servers every 5 minutes via UDP. The statistics include its IP address, the channel it is watching, its buffer availability map, the number of consecutive blocks in its current playback buffer (henceforth referred to as the *buffering level*), instantaneous aggregate download and upload throughput from and to all partners, as well as its download and upload bandwidth capacities.

Though we have been continuously monitoring the performance of UUSee, the study in this paper features a most recent set of run-time traces, collected between Thursday, May 29, 2008 (GMT+8) and Monday, June 2, 2008 (GMT+8), which contains continuous-time snapshots of the streaming system throughout the period, featuring over 16 million peer sessions. We believe these recent traces best captured the up-to-date characteristics of peers in the millions-of-users scale, to which the application has expanded over the years. Here, a *peer session* refers to the lifespan between the joining and the departure of a peer.

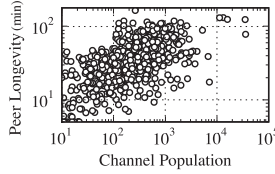
2.2 Observations on Unpopular Channels

Why do we need to investigate the streaming performance in unpopular channels and popular channels distinctively in such a large-scale system? First of all, we observe that the popularity differs significantly across channels: a small number of most popular channels ($\approx 2\%$) with an average peer population over 5000, a small percentage of less popular channels ($\approx 31\%$) with a population in the range of 500 to 5000, and the majority of UUSee channels accommodating a peer population less than 500 ($\approx 67\%$). Fig. 1 plots the correlation between the streaming quality and peer population in all UUSee channels in two representative snapshots, 9 a.m. on May 30 and 9 p.m. on May 30. Here, we evaluate the streaming quality in a channel at each time as the *percentage of high-quality peers in the channel*, where a high-quality peer has a buffering level of more than 80% of the total size of its playback buffer (buffer size in UUSee is 500 media blocks). The criterion of the buffering level (*i.e.*, the number of consecutive blocks in the playback buffer of a peer, starting from the current playback position) has been extensively used in the actual UUSee streaming protocol to evaluate the current streaming quality of a peer. Accordingly, we also use the peer buffering level as our basic streaming quality metric, based on the rationale that the more blocks a peer has cached in its buffer, the higher chance it has to enjoy a smooth playback. We can observe from Fig. 1 that unpopular channels generally represent worse streaming quality, as compared to large popular channels. The less than satisfactory streaming performance in unpopular channels—which represent the majority of streaming channels in UUSee—exposes a critical challenge in improving the performance of real-world streaming systems: How shall we boost the streaming quality of unpopular channels?

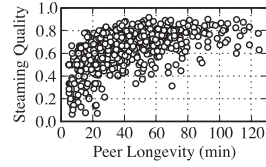
In P2P streaming systems, peer instability represents a “killer” factor that negatively affects the achievable streaming quality. It is even more so in unpop-



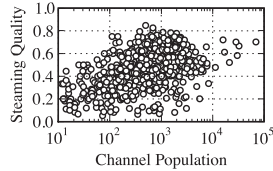
(a) 9 a.m., May 30



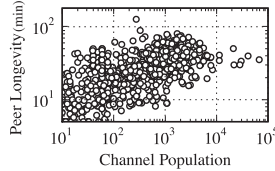
(a) 9 a.m., May 30



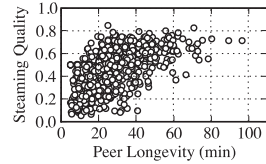
(a) 9 a.m., May 30



(b) 9 p.m., May 30



(b) 9 p.m., May 30



(b) 9 p.m., May 30

Fig. 1: Correlation between channel population and streaming quality.

Fig. 2: Correlation between channel population and peer longevity.

Fig. 3: Correlation between peer longevity and streaming quality.

ular channels, as observed by our trace studies in Fig. 2 and Fig. 3. We found that the peer lifespan (also referred to as peer *longevity*) tends to be shorter in the unpopular channels from Fig. 2, while in most cases the more severe peer churns further exacerbate the streaming quality in those channels, as shown in Fig. 3. All these observations have pointed to the following fact: To promote the streaming quality in unpopular channels, a key step is to improve the peer stability in these channels, by promoting peer online times. In order to promote peer stability, we find it important to obtain a thorough and in-depth understanding of the critical factors that influence peer online times in unpopular streaming channels, which constitutes the major objective of our study in this paper.

3 Deciphering Peer Instability in Unpopular Channels

3.1 Survival Analysis and Censoring

Survival analysis [2] represents a set of statistical methods for the analysis of death or failure events and involves the modeling of time to event data, *i.e.*, the survival time. In our analysis of peer longevity in each streaming channel, a peer's departure represents a *failure* or *death* event, and the time between its joining and departure, the peer longevity, is the survival time to be considered. In survival analysis, a *survival function* is frequently used to describe the probability that an individual survives to a specific time t . Let a random variable T represent the longevity of a peer session, the survival function is defined as: $S(t) = \Pr(T > t) = 1 - \Pr(T \leq t) = 1 - F(t)$, where $F(t)$ is the cumulative distribution function

(CDF) of the longevity. A standard estimator of the survival function, based on a number of measured survival times, is proposed by Kaplan and Meier, referred to as the product-limit estimator or the K-M estimator [2]. We now seek to investigate critical factors that influence peer longevity in unpopular channels based on correlation plotting and survival function K-M estimator.

3.2 Buffering Level

Intuitively, the higher buffering level a peer experiences, the smoother its streaming is, and the more likely it will stay longer in the channel. Therefore, we start by investigating: *Do peer longevity patterns differ significantly under different buffering levels?* To answer this question, we explore the relevance between peer longevity and various statistical metrics of the buffering level, including the average buffering level during a peer session, the standard deviation of the buffering level throughout a peer session and the initial buffering level, as the first buffering level measured when a peer starts its playback. As a statistics to represent the distribution of peer longevity among a group of sessions, we define an $EDR(t)$ function, *i.e.*, Early Departure Rate function, as the percentage of peers whose lifespan is less than or equal to t minutes within a group. In each sub-figure in Fig. 4, we plot the $EDR(15 \text{ min})$ of each session group categorized according to different levels of the respective buffering level metric, as well as the smoothed lowest curves. Note that in all our studies hereinafter, we use peer session data from all the unpopular channels, *i.e.*, channels with less than 500 peers most of the time, in order to derive insights useful for their performance enhancement.

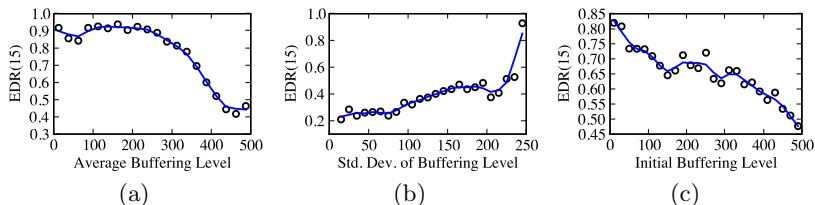


Fig. 4: Correlation of $EDR(15 \text{ min})$ with metrics of buffering level.

Fig. 4(a) reveals a negative correlation between the early departure rate and the average buffering level within the buffering level range of 200 – 475, showing that the departure rate is higher (peer lifespan is shorter) when the average buffering level is lower in this majority range. The positive correlation between the standard deviation of buffering level and early departure rate, as shown in Fig. 4(b), meets our expectation that the less stable the buffering level is, the shorter the peers are staying. We further investigate the tolerance of peers towards the initial buffering level, by plotting the correlations in Fig. 4(c). A strong negative correlation is observed in Fig. 4(c) between the early departure

rate and initial buffering level within the buffering level range of 0 – 120 and 320 – 500, respectively. This reveals that at the two ends of the spectrum, an excellent initial buffering level brings a longer peer online time, and a very poor initial buffering level will almost definitely result in an early departure. In our study, we have varied t in the $EDR(t)$ function from 5 minutes to 60 minutes, and made similar observations.

3.3 Peer Incoming Degree

In P2P streaming, the number of supplying peers a peer can obtain in a streaming channel, *i.e.* its incoming degree or indegree, and how stable these incoming connections are, affect the streaming quality it obtains, and thus affect the longevity of the peer in the channel. To investigate such impact of peer indegree, we plot in Fig. 5 the $EDR(15 \text{ min})$ values of session groups at different levels of the average indegree during a peer session and the standard deviation of the indegree throughout a peer session, respectively.

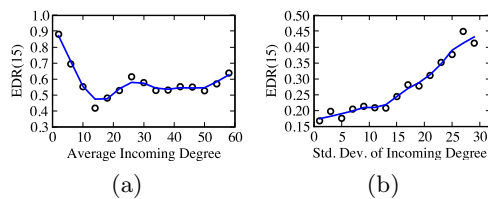


Fig. 5: Correlation of $EDR(15 \text{ min})$ with metrics of peer indegree.

Fig. 5(a) shows an interesting phenomenon: When the peer indegree is at a smaller value (< 20), a negative correlation exists between the average indegree and the early departure rate, meaning that the more suppliers a peer has, the longer it may stay; however, when the indegree goes up, a positive correlation result, showing that the departure rate is high even when peers know many others in the same channel. To explain the latter part of the observation, we have further observed that the majority of peers in unpopular channels in UUSEE have an indegree lower than 30, and only a few may have large indegree up to one hundred. Interesting enough, the peers with large indegrees are generally those with poor buffering levels, which thus strive to find more possible suppliers, but are nevertheless unable to get a satisfactory streaming quality. Fig. 5(b) plots a positive correlation between the standard deviation of indegree and the early departure rate. This reveals the following: when the number of incoming connection fluctuates significantly at a peer, the peer may not be experiencing smooth streaming, and thus is more prone to early departure.

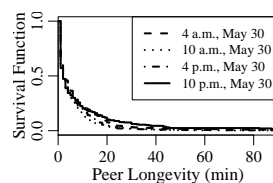


Fig. 6: Detecting the time of the day effect.

3.4 Time Effects

Intuitively, the time when a peer surfs the Internet also influences its viewing behavior. We then explore any possible effect of the time of the day, using sessions starting at different times on a same day and classify sessions according to the hours they start. Fig. 6 exhibits visible differences among survival curves for session groups of four different starting times on May 30. We further statistically validate our observations using the Mantel-Haenszel test [2], also referred to as the log-rank test. The log-rank test is commonly applied to test the null hypothesis that a set of survival functions are statistically equivalent, in which the null hypothesis is rejected if the result p -value is lower than the significance level of 0.05. The log-rank test result of $p \approx 0$, rejects the null hypothesis that survival functions are equivalent and validates our observations.

4 Modeling Peer Longevity: Cox Regression

4.1 The Cox Regression Model

The Cox proportional hazards model [2] is a classical regression model for the analysis of survival times with respect to their relationship with covariates (which are the terminologies in Cox modeling for influential factors). It models the relationship between the covariates and survival times based on the *hazard function*. A hazard function $\lambda(t)$, also referred to as the *hazard rate*, represents the instantaneous failure rate for a session that has survived to time t . Let T denote the duration of a survival session. The hazard function is defined as:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}.$$

In Cox regression modeling, it models the hazard rate at time t for a session with covariate vector $\mathbf{z} = (z_1, \dots, z_p)$ as a function of a baseline hazard function and the influential factors. The basic Cox model is:

$$\lambda(t; \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}) = \lambda_0(t) \exp\left(\sum_{k=1}^p \beta_k z_k\right), \quad (1)$$

where $\lambda(t; \mathbf{z})$ is the hazard rate at time t for a session with covariate vector \mathbf{z} ; $\lambda_0(t)$ is an arbitrary non-negative baseline hazard function, which is computed during the regression process; and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ is a column p -vector of coefficients corresponding to the covariates in \mathbf{z} . A major property of the Cox model is that given two sessions with covariate vector \mathbf{z}_1 and \mathbf{z}_2 , the ratio of their hazard rate is independent of the time:

$$\frac{\lambda(t; \mathbf{z}_1)}{\lambda(t; \mathbf{z}_2)} = \frac{\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_1)}{\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_2)} = \exp(\boldsymbol{\beta}^T (\mathbf{z}_1 - \mathbf{z}_2)).$$

Such a property also imposes the *proportional hazards assumption* for applying the Cox model, that the hazard rate ratio for any two sessions should be always proportional, *i.e.*, dependent *only* on their covariate values.

In our regression modeling, the potential covariates are selected corresponding to the influential factors we have observed. The potential covariates, along with their description and type, are listed in Table 1. In order to use the Cox regression model in (1) to formulate the relationship among these covariates and the hazard rate, we first need to check if the proportional hazards assumption is satisfied, and may adjust the form of the covariates in Table 1, in order to meet the proportional requirement. Once the assumption check is passed, we proceed to derive the values of regression coefficients $\beta_k, k = 1, \dots, p$ in the model. Using the Cox model, we can then estimate the probability that a session lasts to any specific time t (*i.e.*, the survival curve of the session), given the values of the covariates for the session and using the derived coefficients.

Table 1: Potential covariates in Cox Regression Model

Covariate	Description	Type
BUFAVG	Average buffering level of the session	Continuous
BUFSTD	Standard deviation of buffering level during the session	Continuous
BUFINIT	Initial buffering level of the session	Continuous
INDAVG	Average incoming degree during the session	Continuous
INDSTD	Standard deviation of incoming degree during the session	Continuous
TOD	Joining time of the day (TOD $\in \{0, 1, \dots, 23\}$, corresponding to the hours of the day)	Categorical

4.2 Proportional Hazards Assumption Check

Categorical Factor One approach to check the proportional hazards assumption for a categorical covariate, *i.e.*, whether or not the hazard ratio of sessions with different values of a categorical factor is a constant, is to group the sessions based on the values of the corresponding categorical factor, and plot the values of $-\log(\hat{S}(t))$ against t for each session group [2], where $\hat{S}(t)$ is the estimated survival function of the group. The plot for the categorical covariate TOD in our model is shown in Fig. 7(a). If the hazard ratios do not change with time, the curves in the figure should be approximately *parallel*, *i.e.* there is an approximate constant vertical distance between each pair of them at all times. However, in Fig. 7(a), we observe that the curves intersect with each other, indicating the violation of the proportional assumption for TOD.

Given the non-proportionality of the categorical factor, we modify our model in (1) to the stratified Cox model [2], in order to accommodate the categorical factor. The stratified Cox model extends the basic Cox model by incorporating strata, where each stratum corresponds to one hazard rate function, that models the hazard rate of sessions corresponding to one specific value of each categorical factor. The stratified Cox model with n strata ($i = 1, \dots, n$) is given by:

$$\lambda_i(t; \mathbf{z}) = \lambda_{0,i}(t) \exp(\boldsymbol{\beta}^T \mathbf{z}), i = 1, \dots, n. \quad (2)$$

In our modeling, we have one categorical variable with 24 possible values, and thus the total number of strata n is 24. We note that in such a stratified Cox model, each stratum may have different baseline hazard functions, but all strata share the same coefficient vector β as all other non-stratified factors are required to have a constant influence to the hazard functions.

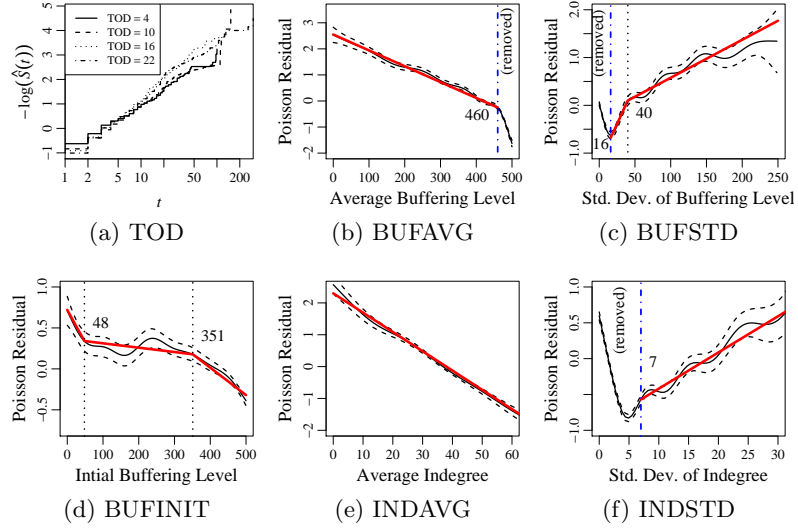


Fig. 7: Proportional hazards assumption check for covariates.

Continuous Factors For a continuous covariate, the proportional hazards assumption implies that it should have a linear influence on the hazard ratio, *i.e.*, the hazard ratio between a session with BUFINIT = 300 and one with BUFINIT = 340 should be the same as that between a session with BUFINIT = 400 and one with BUFINIT = 440. The approach to conduct such an assumption check is to plot the Poisson residual curve [2] for each continuous factor. The Poisson residual of a covariate reflects the *impact* of this specific factor in the hazard rate function: a positive Poisson residual implies a positive impact, *i.e.*, the hazard rate is greater with a larger value of the covariate, while a negative Poisson residual indicates a negative impact, vice versa. The Poisson residual curve should be approximately linear if the hazard ratio between any two sessions with two specific values of the factor is a time-independent constant. We plot the Poisson residual curves (black solid lines) for all our continuous factors in Fig. 7(b)–(f), along with their standard error confidence bands (black dashed lines), and the linear approximation lines (in red). The plots show that many of the Poisson residual curves are not satisfactorily linear, reflecting violation of the proportionality in certain value ranges of the factors. We thus seek

to make necessary adjustments for the form of the covariates, such that all the new covariates have a linear influence on the hazard ratio.

Fig. 7(b) shows an approximated linear curve in the majority range of the average buffering level, except in the range of 460 to 500. To include BUFAVG into the Cox model, we only keep its value range of $[0, 460)$, *i.e.*, we exclude sessions with BUFAVG in the range of $[460, 500]$ when we derive the model coefficients, which nevertheless only represent a small portion ($\approx 7\%$) of all the sessions based on our measurement study. The major part of the Poisson curve in Fig. 7(c) can be approximated by two linear segments. To include BUFSTD into our Cox model, we exclude sessions with BUFSTD in the range of $[0, 16]$ (which only represent a few extremely short sessions), and include a new variable BUFSTD_L to describe the section of BUFSTD with larger values, corresponding to the range of the second linear segment we approximate in Fig. 7(c):

$$\text{BUFSTD_L} = \begin{cases} \text{BUFSTD} - 40 & \text{if } \text{BUFSTD} \geq 40 \\ 0 & \text{otherwise} \end{cases}$$

The Poisson curve in Fig. 7(d) can be approximated by three line segments connected at two knots at 48 and 351, respectively. Since each section of BUFINIT includes a substantial number of sessions, we include two new covariates, BUFINIT_M and BUFINIT_L, to describe the sections of BUFINIT corresponding to linear segments in the middle and to the right, respectively:

$$\text{BUFINIT_M} = \begin{cases} \text{BUFINIT} - 48 & \text{if } \text{BUFINIT} \geq 48 \\ 0 & \text{otherwise;} \end{cases}$$

$$\text{BUFINIT_L} = \begin{cases} \text{BUFINIT} - 351 & \text{if } \text{BUFINIT} \geq 351 \\ 0 & \text{otherwise.} \end{cases}$$

In Fig. 7(e), the Poisson curve of INDAVG can be nicely fitted by one line, revealing the proportionality of the factor on the hazard ratio. In Fig. 7(f), we remove the leftmost part (corresponding to INDSTD in the range of $0 - 7$ with a few sessions), and fit the rest of the curve with a line.

After the adjustment, a covariate vector $\mathbf{z} = (z_1, \dots, z_p)$ with $p = 8$ components is used in our stratified Cox model. The covariates are summarized in Table 2. We note that in Fig. 7(b)–(f), all the linear approximation lines fall within the confidential bands of the original Poisson residual curves, indicating that the Poisson curves in those sections can be effectively approximated by the linear segments. Therefore, after the adjustment, all the covariates we now use in the Cox modeling satisfy the proportional hazards assumptions.

4.3 Estimation of the Coefficients and Model Validation

We next use a specific Cox regression technique proposed by Andersen and Gill [3], to estimate the stratified baseline functions $\lambda_{0,i}, i = 1, 2, \dots, 24$ and the coefficient vector β in our stratified Cox model in (2).

Table 2 gives the coefficients of the covariates along with their standard errors, estimated using information of 12866 session from 20 unpopular channels

in our traces. The 20 channels, whose average concurrent population varies from 48 to 457, are randomly chosen from all 530 unpopular channels contained in our traces. The purpose for such sampling is not only to expedite the speed of the regression process, but also to exhibit the usefulness of our model, trained using only a limited set of samples, as is to be illustrated in the following subsection. We have also computed the p values to test the significance of all the coefficients, which are all far below 0.05, suggesting the significance of the covariates.

Table 2: Covariates and Coefficients for the Cox Model in (2)

Covariate	β	Std. Err.	Covariate	β	Std. Err.
BUFAVG	-0.0074	1.7e-3	BUFINIT	0.011	1.8e-3
BUFSTD	0.059	2.9e-3	BUFINIT_M	-0.012	1.9e-3
BUFSTD_L	-0.044	2.9e-3	BUFINIT_L	-0.0029	4.9e-4
INDAVG	-0.051	1.9e-3	INDSTD	0.046	2.1e-3

With the Cox regression model established, we can now derive the survival curve of a session with covariate vector \mathbf{z} in a certain category of TOD. The estimator of the survival function with covariate vector \mathbf{z} at time t is given by $S_i(t; \mathbf{z}) = \exp(-\int_0^t \lambda_i(u; \mathbf{z}) du)$, where $\lambda_i(u; \mathbf{z})$ is the stratum corresponding to a specific category of TOD of sessions. We may use the expected session time of the survival curve corresponding to a session with \mathbf{z} , as the most probable duration of the session. In this way, given a covariate vector \mathbf{z} and the corresponding TOD, we are able to predict the most probable duration of a session using our stratified Cox model. Recall that our regression model is trained using only a limited set of session data from 20 randomly selected unpopular channels. We now evaluate its accuracy in estimating the duration of sessions in other channels.

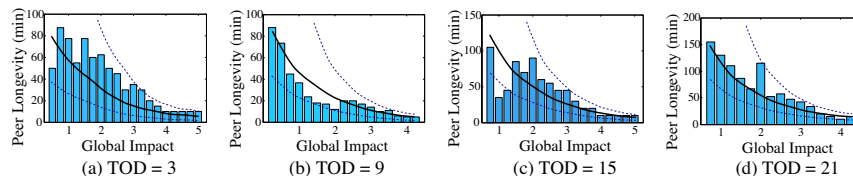


Fig. 8: Regression model validation by prediction of the peer longevity.

We calculate the overall influence of the continuous covariates by $\beta^T \mathbf{z}$, referred to as the *global impact* to the hazard rate. We group all the sessions in the unpopular channels in UUSee (other than the 20 channels for the regression), by their values of global impact and TOD, and plot the median session duration for session groups at different levels of their global impact in different cases of TOD,

as shown by the bar plotting in Fig. 8. In solid curves, we also plot the session duration predicted using our Cox model in (2) at each global impact level, and the 75% confidence intervals (in dashed lines). The actual median session durations in all four figures fall into the confidence intervals, which validates the usefulness of our regression model—derived using a small portion of session data—in accurately capturing the session duration patterns for unpopular channels.

5 Model Implications to Unpopular Streaming Channels

5.1 Impact of the Streaming Quality Factors

Three streaming quality factors are involved in our Cox regression model. We seek to investigate the relative significance of their impact on peer longevity, by calculating an impact value, $\beta_k z_k - \min_{\mathbf{y}}(\beta_k y_k)$, for each individual session with a specific z_k , with respect to the three streaming quality factors of average buffering level, standard deviation of buffering level and initial buffering level, respectively. Here, $\min_{\mathbf{y}}(\beta_k y_k)$ is the minimal impact value for the corresponding streaming quality factor over all the sessions. We then compute the ratio of the impact values of the three streaming quality factors for each session, and derive the average ratio across all sessions. The normalized average ratio in the percentage format is 16% : 33% : 51%, which exhibits the relative level of user’s *intolerance* to the three streaming quality factors, respectively.

An intriguing discovery is that the initial buffering level is the most important streaming quality factor affecting peer longevity. Using the coefficients in Table 2, we derive the coefficient β corresponding to the initial buffering level factor in the range of 48 to 351 is $\beta_{\text{BUFINIT}} + \beta_{\text{BUFINIT}_M} = -0.001$, and the coefficient corresponding to the range of 351 and 500 is $\beta_{\text{BUFINIT}} + \beta_{\text{BUFINIT}_M} + \beta_{\text{BUFINIT}_L} = -0.0039$. The more negative coefficient in the latter case illustrates that when the initial buffer is relatively full, a small increase of buffering level induces more significant decrease of failure probability, *i.e.*, more evident effect in keeping peers longer in the system. Therefore, to promote the stability of high-contribution peers, efforts should be made to guarantee them a high initial buffering level.

5.2 Impact of the Incoming Degree Factors

Following a similar methodology, we further compare the impact of the average peer indegree and the standard deviation of peer indegree, and derive a user’s intolerance ratio of 38% : 62% to the two factors. It confirms that in unpopular channels, peers are much less tolerant to neighbor churns than the average level of neighbor numbers. Therefore, the P2P protocol should always try to find stable good neighbors for each peer, the number of which may be small, but is much desirable than a large number of transient neighbors.

6 Related Work

With respect to P2P measurements related to peer longevity, Hei *et al.* [4] characterized the distribution of peer life time in PPLive, and exhibited different

life-time patterns among popular and unpopular channels. Li *et al.* [5] have also observed a heavy-tailed peer lifetime distribution in their measurement study of Coolstreaming. Based on simulations, Tang *et al.* [6] have shown that the longer peers stay, the better the overall streaming quality is in a streaming channel. Focusing on P2P applications other than live streaming, Stutzbach *et al.* [7] have characterized peer arrivals and departures in three popular P2P file-sharing systems (BitTorrent, Kad and Gnutella). Chen *et al.* [8] have investigated the influence of network QoS metrics on peer session lengths in a P2P VoIP application, Skype. Our work distinguishes itself from all the existing measurement work, by focusing on a thorough understanding of the causes to peer lifetime patterns in unpopular channels, in order to improve their streaming quality.

7 Concluding Remarks

This paper focuses on improving the streaming quality in the large number of unpopular channels in real-world P2P live streaming systems. Utilizing over 130 GB worth of traces from a large-scale commercial system, UUSee, we thoroughly characterize the important factors that influence peer longevity. Our key contributions include: *first*, we successfully identify the key factors that decide the duration of peer sessions, including the initial buffering level, incoming degree and peer joining time; *second*, we model their relationship into a Cox regression model, using a survival analysis approach; *third*, we discuss implications of our model and derive a number of useful insights to promote peer stability in unpopular channels. As important applications of our Cox model, we can compute and compare the relative stability of peers during the peer selection process, and can promote the online time of high-contribution peers by guaranteeing them a better initial buffering level and stable download bandwidth. All these assist in improving the stability and streaming quality in unpopular channels.

References

1. UUSee (<http://www.uusee.com/>)
2. Machin, D., Cheung, Y.B., Parmar, M.: Survival Analysis: A Practical Approach. 2 edn. John Wiley and Sons Ltd. (2006)
3. Andersen, P., Gill, R.: Cox's regression model for counting processes: A large sample study. *Annals of Statistics* **10**(4) (1982) 1100–1120
4. Hei, X., Liang, C., Liang, J., Liu, Y., Ross, K.: A measurement study of a large-scale P2P IPTV system. *IEEE Trans. Multimedia* **9**(8) (2007) 1672–1687
5. Li, B., Xie, S., Keung, G., Liu, J., Stoica, I., Zhang, H., Zhang, X.: An empirical study of the Coolstreaming+ system. *IEEE JSAC* **25**(9) (2007) 1627–1639
6. Tang, Y., Sun, L.F., Zhang, K.Y., Yang, S.Q., Zhong, Y.Z.: Longer, better: On extending user online duration to improve quality of streaming service in P2P networks. In: Proc. IEEE ICME '07. (July 2007) 2158–2161
7. Stutzbach, D., Rejaie, R.: Understanding churn in peer-to-peer networks. In: Proc. ACM IMC '06. (2006) 189–202
8. Chen, K.T., Huang, C.Y., Huang, P., Lei, C.L.: Quantifying Skype user satisfaction. In: Proc. ACM SIGCOMM '06. (2006) 399–410