P2P Live Seeding: Efficiency of Seeders in P2P Live Streaming

Fabien Mathieu

P2P Live Seeding: Efficiency of Seeders in P2P Live Streaming

Fabien Mathieu [†]

Orange Labs, Issy-les-Moulineaux, France

En pair-à-pair, l'utilisation de *seeders* (pairs contributeurs sans être demandeurs) permet de créer un effet de levier qui améliore considérablement les performances du système. Naturel dans le contexte du partage de fichiers ou de vidéo-à-la-demande, le concept de *seeding* peut sembler contre-intuitif en Live Streaming. Cet article vise à étudier la faisabilité et les performances théoriques que l'on peut attendre d'un système de diffusion en temps réel utilisant le seeding.

Notre approche consiste à partir d'un modèle simple que l'on va complexifier pour le rendre de plus en plus réaliste. Pour un système idéal où le nombre de connexions est illimité et leur coût nul, nous montrons que le surcoût dû à l'utilisation de seeders est marginal. Nous étudions ensuite l'impact d'une connectivité limitée sur l'efficacité des seeders. Enfin, à partir d'un modèle affine de coût des connexions, nous donnons une formule simple décrivant l'efficacité des seeders, nous permettant d'avoir un modèle réaliste de dimensionnement d'un système de Live Streaming.

Keywords: P2P Live Streaming, Seeders, Efficiency

1 Introduction

Bandwidth may be the most critical resource in P2P content distribution. In order to increase the available resources, a standard technique is to leverage the capacity of the system by using *seeders*, *i.e.* peers that contribute to the system but are (currently) not needing anything. Using seeders is a natural idea when considering file-sharing or Video-on-Demand systems (after a peer has downloaded its file/video-on-demand, it becomes a potential seeder for that content), but it may be counter-intuitive for Live Streaming systems, where the content cannot have been pre-fetched to some seeders before others ask for it : even if they are willing to contribute, seeders have no access to the live content unless something specific is done about that. The goal of this paper is to describe the feasibility and performance one can expect from P2P Live Seeding.

We propose to start from a very simple model, and to increase progressively its realism : in Section 3, we show that for an ideal P2P system where unlimited connections can be spawned at no cost, Live Seeding can be deployed for a marginal cost compared to the theoretical performance given by the Bandwidth Conservation Law (BCL) ; Section 4 introduces a more realistic model where connectivity has an implicit cost and must be limited ; then that cost is explicitly considered in Section 5, where we use a simplified overhead model to derive the optimal efficiency of seeders. This results in a realistic model for dimensioning Live Streaming systems that use seeders, which is illustrated by a small example in Section 6.

The next Section presents the model and existing results used in this paper.

2 Model and related work

We consider a live content of constant streamrate r, and a P2P live streaming system used to deliver that content. We classify the nodes of the system into three categories :

Servers are in charge of injecting initial copies of the stream into the system. We assume they have a cumulated capacity $U_C = rN_C$, with $N_C \geq 1$, so at least one copy of the stream can be injected.

Leechers are the nodes that want to watch the live content.

Seeders are idle nodes that do not want to watch the live content, but can provide bandwidth to the system [‡].

[†]This work has been supported by the Collaborative Research Contract Mardi II between INRIA and Orange Labs, and by the European Commission through the NAPA-WINE Project, ICT Call 1 FP7-ICT-2007-1, Grant Agreement no.: 214412.

‡. Keeping seeders in a P2P content distribution system is a standard issue, as they have no direct incentive to stay in the system. Experience from filesharing and VoD tells us that if the altruistic behavior of the users is not enough, share-ratio policies can be deployed to ensure a fair population of seeders.

r	Streamrate of the content (constant)	α	upload/streamrate ratio
u_n	Available upload bandwidth of peer n	β	Seeders/Leechers ratio
$U_X/\bar{u}_X$	Total/average upload capacity of population X	η	Efficiency coefficient
N_L (resp. N_S)	Number of leechers (resp. seeders)	a	Proportional cost of a connection
N_C	Normalized capacity of servers ($U_C = N_C r$)	b	Constant cost of a connection

TABLE 1: Table of notation

We denote by C, L and S the sets of servers, leechers and seeders respectively. The number of leechers (resp. seeders) is denoted by N_L (resp. N_S). Every peer n in L or S has an upload capacity u_n devoted to the service and a download rate of at least r (the download capacity is not a bottleneck for the broadcast of the live content). U_X and $\bar{u}_X$ are respectively the total and average upload bandwidths of set X ($\bar{u}_X = \frac{U_X}{N_X}$).

According to BCL [BMHP08], if all available bandwidth resources can be used to useful content transfer, then Live Streaming feasibility is given by

$$\alpha_L + \beta\alpha_S + \frac{N_C}{N_L} \geq 1, \text{ with } \begin{cases} \alpha_X = \frac{\bar{u}_X}{r}, \\ \beta = \frac{N_S}{N_L}. \end{cases} \tag{1}$$

Of course, in real systems, not all available bandwidth can be used, because : some overhead is required ; bandwidth may be wasted to data transfers that are not directly useful ; some nodes may not be able to fully use their upload bandwidth.

For BitTorrent-like file-sharing systems, Qiu and Srikant proposed to introduce an efficiency parameter $\eta \in [0,1]$ that takes these issues into account [QS04]. Formally, we define the efficiency of a node (or set of nodes) as the ratio between the maximal effective bandwidth it can add to the system and its bandwidth capacity. Transposed to our live streaming system, this gives the feasibility condition

$$\eta_L\alpha_L + \eta_S\beta\alpha_S + \eta_C\frac{N_C}{N_L} \geq 1, \text{ where } \eta_X \text{ is the efficiency of set } X. \tag{2}$$

For peer-assisted live streaming (systems with no seeders, i.e. $\beta = 0$), it has been shown that if we don't take overhead into account, then one can have $\eta_L = \eta_C = 1$, i.e. a perfect use of available bandwidth can be achieved [LZSJ+08]. In the rest of this paper, we study what η_S can be, or in other words how efficient seeders can be in a P2P live streaming system.

3 Live seeding for ideal systems

Following [LZSJ+08], we first consider an ideal system that we define by the following properties :
– no overhead, so in the bandwidth budget we only consider the effective data transfer (goodput) ;
– unlimited connectivity, so one peer can send data to all others if it has the necessary upload bandwidth ;
– stream continuity : the live stream can be divided into arbitrary small substreams of constant rate.

Theorem 1 *The feasibility condition for an ideal P2P live streaming system with seeders is*

$$\alpha_L + \beta\alpha_S + \frac{N_C}{N_L} \geq 1 + \frac{\min(\beta\alpha_S, 1)}{N_L}. \tag{3}$$

Sketch of proof If (3) is verified, one can construct a working live streaming system by giving to each seeder s a distinct substream of rate $\frac{u_s}{N_L}$ (if $\alpha_S\beta \leq 1$) or $\frac{u_s}{U_S}r$ (otherwise). s is in charge of broadcasting its own substream to the N_L leechers. The remaining rate, if any, is served by the leechers and the remaining capacity of servers following a perfect peer-assisted diffusion [LZSJ+08]. The cost of using seeders is the rate received by S, which is $r\min(\beta\alpha_S, 1)$ and corresponds to the right term in (3). Reciprocally, if (3) is not verified, one can check that the maximal bandwidth usable by leechers is less than required. □

The efficiency of S, measured when seeders are used at their full capacity (case $\alpha_S\beta \leq 1$) by comparing Equations (2) and (3), is

$$\eta_S = 1 - \frac{1}{N_L}. \tag{4}$$

In other words, seeders are asymptotically optimal in an ideal P2P live streaming system, as the only bandwidth indirect cost relies in at most one streamrate redirected to S for replication.

4 Limited connections

We now propose a more realistic model where overhead is indirectly taken into account by assuming that each seeder s has a limit c_s on the number of connections it can maintain.

Theorem 2 *The efficiency of a seeder s with limited connections c_s is*

$$\eta_s = (1 - \frac{1}{c_s}) \min(1, \frac{rc_s}{u_s}). \tag{5}$$

Sketch of proof The most bandwidth-efficient way to use s is to feed it with a substream of rate $\min(r, \frac{u_s}{c_s})$ that will be usefully sent to c_s other peers (leechers or further relaying seeders). This will "cost" the input rate $\min(r, \frac{u_s}{c_s})$, as it will not be directly used by leechers, leading to the result. □

Equation (5) naturally extends (4). However, summing it over S, which gives

$$\eta_S = \frac{\sum_{s \in S} \eta_s u_s}{U_S}, \tag{6}$$

is no longer the real value of S' efficiency, but an upper approximation [§].

5 Modeling the overhead

In real systems, connectivity is not arbitrarily limited. We propose to model the bandwidth cost of a connection that sends content at a rate e to a peer by a linear function $ae + b$. a is the proportional cost (for instance negotiation messages exchanged anytime a given quantity of data has been sent) and b the additive cost (like periodic messages). We assume for simplicity that the overhead cost is supported only by the sender. Also, let $R := (1+a)r + b$ be the bandwidth required for sending one copy of the stream through a single connection (one can easily check that $\frac{r}{R}$ is the maximal efficiency achievable in our model).

Theorem 3 *If the overhead follows a linear function, then the efficiency of a seeder s is*

$$\eta_s = \begin{cases} 0 & \text{if } u_s \leq 2b, \\ \frac{(1-\sqrt{\frac{b}{u_s}})^2}{1+a} & \text{if } u_s \leq \frac{R^2}{b}, \\ \frac{r}{R} - \frac{r}{u_s} & \text{otherwise.} \end{cases} \tag{7}$$

Sketch of proof In order to be useful, a seeder must have $c \geq 2$, which implies $u_s > 2b$. Then, the input content rate that a seeder of bandwidth u maintaining c connections can sustain is $i = \min(r, \frac{\frac{u_s}{c} - b}{1+a})$, corresponding to an efficiency $\eta = (c-1)\frac{i}{u_s}$. For $u_s \leq \frac{R^2}{b}$, the optimal value for c is $\sqrt{\frac{u_s}{b}}$ [¶], which corresponds to $\eta = \frac{(1-\sqrt{\frac{b}{u_s}})^2}{1+a}$. Otherwise, we should take $c = \frac{u_s}{R}$, which leads to $\eta = \frac{r}{R} - \frac{r}{u_s}$. □

Theorem 3 and proof lead to two interesting remarks.

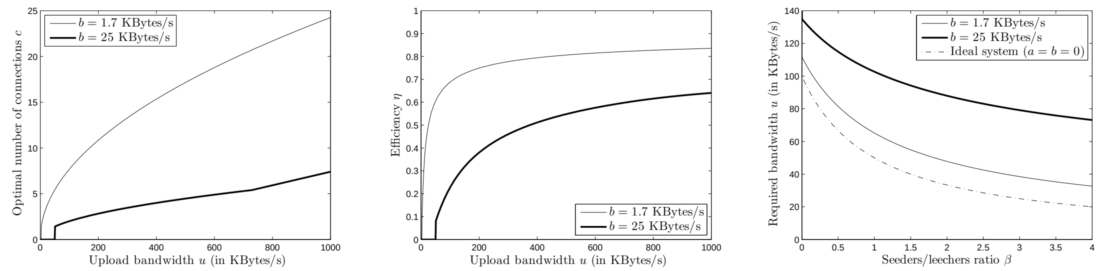
- For bandwidths lower than $\frac{R^2}{b}$, the optimal number of connections and efficiency are functions of u, a, and b, but not of r. Moreover, the formula for the number of connections, $\sqrt{\frac{u}{b}}$, is similar to the empirical formula used in the current BitTorrent mainline, $\sqrt{0.6u}$ [CNM08]. In fact, our model could explain the formula used in BitTorrent, assuming an additive connection cost $b \approx 1.7$ KBytes/s.
- For high bandwidths, the asymptotic efficiency is $\frac{r}{R}$, which is the maximal feasible efficiency given the overhead : super seeders are nearly optimal in a live streaming system.

6 Application : dimensioning a scalable Live streaming system

In order to illustrate the interest of our work, consider a live streaming system with $r = 100$ KBytes/s, a proportional overhead of 10% ($a = 0.1$), and an additive cost that can be small ($b = 1.7$ KBytes/s) or large ($b = 25$ KBytes/s). Figure 1a and 1b indicates how many connections seeders should have as a function of their bandwidth, and the corresponding efficiency.

§. The exact computation of the gap between the actual value of η_S and the one given by (6) is complex, as it must take into account quantifications issues in the complete diffusion process (for instance, if one seeder has to redirect one of its output rate to a seeder of lesser input capacity, some additional bandwidth has to be wasted in the process). Such work could be done in a future work by adapting techniques from [LZSJ+08], but for now we argue that the loss due to quantification can be neglected compared to the loss due to limited connectivity, especially for large systems.

¶. As this may not be an integer, we should in fact take the closest integer maximizing the efficiency. For simplicity, we do not take this quantification into account, as it makes the writing more complex while not drastically changing the results presented here.



(a) Optimal number of connections (b) Maximal efficiency (c) Bandwidth required for scalability

FIGURE 1: $c = f(u)$, $\eta = f(u)$ and $u = f(\beta)$ for a system with $r = 100$ Kbytes/s, $a = 0.1$.

Many dimensioning rules can then be derived by using our proposed formulas. For instance, determining if the system is scalable would consist in checking if $\eta_L\alpha_L + \eta_S\beta\alpha_S \geq 1$. If we assume here for simplicity homogeneous bandwidth u, $\eta_S = \eta_s(u)$ (neglecting quantification issues), and optimal leechers' efficiency $\eta_L = \frac{r}{R}$ ‖, one can derive the relationship that u and β must verify for the system to be scalable :

$$\beta \geq \eta_s(u)(\frac{r}{u} - \frac{r}{R}). \tag{8}$$

If β, which indicates the ratio between idle (seeders) and active (leechers) users, is a given parameter of the system, Equation (8) can be used to derive the bandwidth u that is required for the system to be scalable. This is illustrated by Figure 1c (the performance of the ideal system, i.e. $a = b = 0$, is also plotted for comparison). Notice how even little values of β (less than 1) can give significant decrease of the required bandwidth, which is R for a seedless system with perfectly efficient leechers.

7 Conclusion

In this paper, we gave the keys to understand why seeding should be used in P2P live streaming if idle peers are present, and how efficient Live Seeding can be. After introducing ideal and limited-connectivity systems, we used a linear overhead model to give realistic formulas. Although this is a preliminary study, we believe that the results presented here are significant for the design and dimensioning of live streaming systems with seeding. In a future work, we plan on considering more carefully the quantification issues that appear in the limited-connectivity and linear-overhead models in order to refine our efficiency formulas. We also intend to apply the linear overhead approach to file-sharing and VoD systems.

Références

[BMHP08] Farid Benbadis, Fabien Mathieu, Nidhi Hegde, and Diego Perino. Playing with the bandwidth conservation law. In *IEEE P2P*, 2008.

[CNM08] Damiano Carra, Giovanni Neglia, and Pietro Michiardi. On the impact of greedy strategies in bittorrent networks : The case of bittyrant. In *IEEE P2P*, pages 311–320, 2008.

[LZSJ+08] Shao Liu, Rui Zhang-Shen, Wenjie Jiang, Jennifer Rexford, and Mung Chiang. Performance bounds for peer-assisted live streaming. *SIGMETRICS Perform.*, 36(1) :313–324, 2008.

[QS04] Dongyu Qiu and R. Srikant. Modeling and performance analysis of bittorrent-like peer-to-peer networks. *SIGCOMM Comput. Commun. Rev.*, 34(4) :367–378, October 2004.

‖. The efficiency of leechers should take into account the number of outgoing connections like we did for the seeders. However, η_L is not the main matter of this paper, so we assume without remorse perfect efficiency $\frac{r}{R}$.