OVERLAY OPTIMIZATION FOR PEER-TO-PEER SCALABLE VIDEO STREAMING

Livio Lima, Marco Dalai, Pierangelo Migliorati, Riccardo Leonardi

{*name.surname*}@ing.unibs.it
Department of Information Engineering, University of Brescia, Italy

ABSTRACT

Video streaming with Peer-to-Peer (P2P) architectures and Scalable Video Coding (SVC) appears to be an interesting solution for an efficient streaming in the heterogeneous scenario of Internet applications. A key issue in such approach is the optimization of the bandwidth capacity of the P2P system. In this paper we propose an innovative approach for the network overlay optimization based on Integer Linear Programming. The proposed approach is particularly suitable in case of push-based solutions for video streaming using SVC with prioritized content. The simulation results show the flexibility of the proposed model when used to generate an overlay following different constraints and operational requirements. The usability and the computational complexity of the proposed method is also analyzed when the overlay includes an high number of peers.

Index Terms— Video Streaming, P2P architectures, Scalable Video Coding, Overlay Optimization.

1. INTRODUCTION

Recently two technologies have gained popularity for the efficient distribution of videos through heterogeneous systems, namely: Scalable Video Coding (SVC) and Peer-to-Peer (P2P) architectures. SVC, that has been recently standardized as a scalable extension of the H.264/AVC standard [1], easily adapt to client's available bandwidth and display capability simply decoding only a subset of the full scalable video stream. P2P starts to be an interesting solution for Internet video broadcast/streaming [2], for the potential benefits as low costs for system development (no infrastructure requirements) and the easy scaling with the increasing demand (resource sharing).

The main problem in a P2P streaming system is the organization of the peers into a high quality overlay (the logical network that connects the physical nodes) for disseminating the video stream, since this can greatly affect the system performance. Many factors are involved in the overlay generation, as the number of peers, the input and output bandwidth of each node, the latency, the delay, the link packet loss rate and other system considerations. In particular, a critical aspect is the efficient bandwidth utilization, since the video requires high bandwidth, and different peers have variable upload and download capabilities. It is therefore very important to obtain an overlay with the maximum possible throughput, in which the resources of each peer (both upload and download) are fully exploited. In this paper we consider as network throughput the sum of the data rates that are delivered to all the peers in the considered network.

The existing overlay generation approaches can be grouped in two categories: tree-based and data-driven. Useful references to different overlay approaches can be found, for example, in [2]. The Tree-based approaches are typically push-based, where each node just "pushes" the received data to its children. Data-driven approaches are pull-based, where each node requests portions of the video to the neighbors, and the nodes are typically connected in a mesh. In both the mentioned approaches, the logical connections between peers are built by the exchange of messages within a subset of peers using specific algorithms for parent selection or partnership management.

This work is a part of a project that has the objective to investigate a new system for scalable video streaming adopting an hybrid approach between mesh connection and push-based data dissemination. In our approach the overlay is generated by an external entity, as a network manager, that creates it through an optimization process based on the peers properties and the estimated network features. In particular this optimization is performed through the solution of an Integer Linear Programming (ILP) problem that produces the overlay in terms of active connections between peers and the data-rate to be transmitted on each link. One of the possible critical aspects of the proposed method is the computational complexity since the ILP problem dimensionality can greatly increase in presence of an high number of peers. In this paper we have evaluated two aspects of the proposed approach: the flexibility to generate different topologies and the computational complexity, with the aim to demonstrate that the method can be applied also with an high number of peers.

The rest of the paper is structured as follow. Section 2 provides an overview of the P2P streaming system under investigation. Section 3 describes the proposed solution for the overlay generation problem. In Section 4 the preliminary simulation results will be presented, while conclusions will be drawn in Section 5.

2. P2P SYSTEM OVERVIEW

In this section we provide an overview of the P2P video streaming system currently under investigation. The considered system combines the use of SVC, a new approach for overlay optimization, and the adoption of a chunk-less approach for data dissemination.

From coding perspective, SVC is the efficient solution for video distribution in the P2P heterogeneous environment, since it enables the adaptation to different peer's features and to the variability of network bandwidth. SVC has recently been proposed in some P2P streaming systems (e.g., [3, 4, 5]). Furthermore, SVC offers native mechanisms for bitstream priority generation [6] and the possibility to increase the robustness with the combined use of SVC and Distributed Video Coding, potentially useful in P2P environment.

From network perspective, our aim is to find an approach that combines the benefits of tree-based and data-driven solutions. We identify such approach with an overlay based on a mesh topology in which each peer distributes data in a chunk-less push-based fashion. The mesh optimization has to include the description of the active links between peers and the data-rate to be sent on each link since

This work was partially supported by the PRIN 2008 project ARACHNE (prot. 2008C59JNA) founded by the Italian Ministry of University and Research (MIUR).

each peer adopts a push approach for data distribution but it receives data from multiple sources (and send to multiple peers as well). The proposed mechanism for the mesh optimization is based on the use of Integer Linear Programming (ILP) that offers several advantages, such as the high flexibility to obtain a rich problem description and the computational advantages of Linear Programming. More details will be provided in Section 3. In other previous works as [7] ILP has been proposed for overlay generation, but only to describe the active links, since a pull-based approach was adopted.

The adoption of a multi-source push-based chunk-less approach introduces the problem of robust dissemination of data in order to prevent duplication. The best identified approach to solve the problem is to use network coding based approaches (e.g., see [8, 9]), that are recently growing popularity for P2P applications. The drawback of the network coding-based approaches is the possible increased delay due the data decoding/encoding process.

3. OVERLAY GENERATION

In this section we describe how the ILP can be used for the optimal overlay generation with network throughput allocation. We adopt the ILP approach since it can offer high flexibility in problem modeling and computational advantages due to the linearity.

In this work we consider a simplified model of the overlay generation process, stated as follows. Let us consider N peers, each one defined by its upload (U) and download (D) capacity and potentially connected to M other peers, where $M < N$. The objective of the optimization is to determine which links remain active and the data-rate transmitted on each link, in order to minimize (or maximize) a particular target function. The main limitation of such model is that only peer status information, as the upload and download capacity, are involved in the overlay generation process. Information about the physical network, as delay and the link packet loss rate are not included at the moment, but ongoing investigation is currently devoted to extend the model in order to include such information.

3.1. Basic model

Since the aim of this work is to find the optimal overlay, it is important to define properly the measure of optimality. In our approach we identified two criteria: the maximization of the throughput and the minimization of peers video distortion experienced. Since the target application is the video streaming, probably the latter is preferred. Nevertheless, the main problem that concerns the peer's distortion minimization is the non-linear dependency between throughput and distortion. This non-linearity in the function to be optimized preclude the use of a pure Linear Programming approach, increasing the complexity of a resolution strategy. The adopted solution is to discretize the link capacity and to introduce a binary variable for each step that identifies the capacity:

$$c = \sum_{s=1}^{S} \delta_s \mathbf{x}^s, \qquad s = 1, ..., S, \qquad \sum_{s=1}^{S} \mathbf{x}^s \leq 1 \tag{1}$$

where c is the generic link capacity, S is the number of discrete steps, $\delta_s = sR_F/S$ is the value of the discrete step and the constraint (1) ensures that the capacity of a link is identified by a single binary variable $\mathbf{x}^s$ (possibly $c = 0$). Here R_F is assumed to be the rate of the full scalable bitstream.

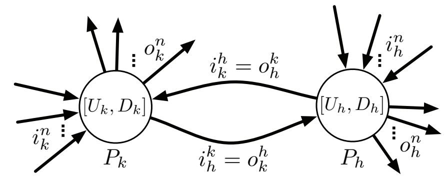

Fig. 1. Input-output connections between nodes.

The overlay generation modeling needs the introduction of further constraints. Consider the architecture of a couple of peers represented in Fig. 1. The generic peer P_k has to satisfy the following constraints on the ingoing and outgoing capacity:

$$\sum_{n \in \mathcal{I}_k} i_k^n = \sum_{n \in \mathcal{I}_k} \sum_{s=1}^{S} \delta_s \mathbf{x}_k^{n,s} \leq D_k, \qquad \forall k \in \mathcal{P} \tag{2}$$

$$\sum_{n \in \mathcal{I}_k} i_k^n = \sum_{n \in \mathcal{I}_k} \sum_{s=1}^{S} \delta_s \mathbf{x}_k^{n,s} > 0, \qquad \forall k \in \mathcal{P} \setminus \mathcal{S} \tag{3}$$

$$\sum_{n \in \mathcal{O}_k} o_k^n = \sum_{n \in \mathcal{O}_k} \sum_{s=1}^{S} \delta_s \mathbf{y}_k^{n,s} \leq U_k, \qquad \forall k \in \mathcal{P} \tag{4}$$

$$o_k^j \leq \sum_{n \in \mathcal{I}_k} i_k^n \qquad \forall j \in \mathcal{O}_k, \forall k \in \mathcal{P} \tag{5}$$

where $\mathcal{I}_k$ and $\mathcal{O}_k$ represent the set of peers with outgoing connection to P_k and the set of peers with incoming connection from P_k, respectively. D_k and U_k are the maximum download and upload bandwidth of P_k, i_k^n is the data rate of the generic incoming connection, and o_k^n is the data rate of the generic outgoing connection. $\mathcal{P}$ is the set of peers and $\mathcal{S}$ is the set of source peers that by definition do not receive data. It has to be noted that the constraint (3) ensures that all non-source peers receive a non-zero data rate. Here the binary variables associated to the incoming and outgoing connections have been differentiated for the seek of clarity, but is obvious that $\mathbf{x}_k^{n,s} = \mathbf{y}_n^{k,s}$.

Considering that the proposed overlay model is a push-based approach which extend a tree configuration, it is important to define a further constraint that avoid loops between peers. Let us assume the imposition to avoid loops of length equal to L. The expression of such constraint is:

$$\sum_{s=1}^{S} \left(\mathbf{y}_{k_1}^{k_2,s} + \mathbf{y}_{k_2}^{k_3,s} + ... + \mathbf{y}_{k_{L-1}}^{k_1,s} \right) < L \tag{6}$$

$\forall \{k_1, k_2, ...k_{L-1}\} \in \mathcal{P}$, with $k_1 \neq k_2 \neq ... \neq k_{L-1}$.
The constraints (1)-(6) represent the basic description of the overlay generation problem. In order to complete the ILP model we have to define the possible objective functions:

$$T = \sum_{k \in \{\mathcal{P} \setminus \mathcal{S}\}} \sum_{n \in \mathcal{I}_k} \sum_{s=1}^{S} \delta_s \mathbf{x}_k^{n,s} \tag{7}$$

$$\hat{D} = D_{MAX} - \sum_{k \in \{\mathcal{P} \setminus \mathcal{S}\}} \sum_{n \in \mathcal{I}_k} \sum_{s=1}^{S} d_s \mathbf{x}_k^{n,s} \tag{8}$$

with T that represent the throughput and $\hat{D}$ the estimated distortion, whereas $d_s = d(\delta_s)$ are the estimated distortion reduction contributions obtained receiving data with capacity δ_s.

Configuration	Constraints
C_1	base model
C_2	base model + (11)
C_3	base model + (11) + (9) + (10)

Table 1. Configurations used in the experiments described in Section 4.

3.2. Additional constraints

In order to improve the ILP model flexibility, additional constraints on the overlay topology could be include in the model. Basically we identify three additional constraints: limit the number of each peer incoming connections, limit the number of each peer outgoing connections, and limit the number of hops, i.e., the number of peers crossed by the same data flow. This latter constraint could be particularly useful to limit the delay. The constraints on incoming connections can be expressed as:

$$\sum_{n \in \mathcal{I}_k} \sum_{s=1}^{S} \mathbf{x}_k^{n,s} \leq I_{MAX}, \qquad \forall k \tag{9}$$

while the constraints on outgoing connections can be expressed as:

$$\sum_{n \in \mathcal{O}_k} \sum_{s=1}^{S} \mathbf{y}_k^{n,s} \leq O_{MAX}, \qquad \forall k \tag{10}$$

It is important to note that the constraints (9) and (10) can also be imposed only to a restricted sub-set of peers. The limitation on the number of hops can be imposed with a constraint similar to (6). If H is the maximum number of hops, the constraint is:

$$\sum_{s=1}^{S} \left(\mathbf{y}_{k_1}^{k_2,s} + \mathbf{y}_{k_2}^{k_3,s} + ... + \mathbf{y}_{k_{H-1}}^{k_H,s} \right) < H \tag{11}$$

$\forall \{k_1, k_2, ...k_H\} \in \mathcal{P}$, with $k_1 \neq k_2 \neq ... \neq k_H$.

4. SIMULATION RESULTS

In this section we present an experimental evaluation of the proposed approach for overlay generation. Specifically the proposed solution is evaluated from two different point of view: the flexibility to generate different overlay topologies, and the complexity of the proposed approach (in particular when the optimization is performed on a large group of peers). The complexity is evaluated as the computational time required to solve the ILP problem described in Section 3. ILP problems can be solved using solvers for mixed integer linear programming problems; in our experiments we used the CPLEX solver [10]. In both the set of experiments we have created the network to optimize by randomly extracting N peers within a set of five possible configurations: ($U = 1000, D = 1000$), ($U = 200, D = 1000$), ($U = 200, D = 500$), ($U = 600, D = 800$), ($U = 500, D = 1000$). The networks generated include also one source with upload bandwidth equal to 2000, and a sink node with download bandwidth equal to 200.

4.1. Analysis of flexibility

In order to evaluate the flexibility of the proposed approach, we present the solutions of the overlay generation problem described in Section 3 obtained under different constraints configurations, as reported in Table 1, using (7) as objective function. For simplicity of presentation, Fig. 2 shows the overlays generated for configurations C_1, C_2 and C_3 for 9 peers with different features in terms of upload and download bandwidth. Here R_F is set to 1000, $H = 4$, $I_{MAX} = O_{MAX} = 2$. Figure 2 reports the value of throughput and distortion for each configuration, evaluated as the mean PSNR over the 8 non-source peers, obtained decoding a scalable bitstream adapted to the overall incoming rate for each peer. In this experiment the rate adaptation for scalable bitstream is obtained adopting the optimized priority assignment method proposed in [6].

As it can be noted from Fig. 2, additional constraints cause a decreasing of the effective throughput. Given the bandwidth features of the peers it is easy to show that the maximum reachable throughput T is given by:

$$T = \min \left(\sum_k U_k, \sum_k D_k \right), \tag{12}$$

which, for the particular peers configuration of Fig. 2, is equal to 4600. This value is obtained only in configuration C_1. Configurations C_2 and C_3 have lower throughput since the additional constraints preclude the possibility to reach the maximum value. In general, the more constraints are included the lower is the obtained throughput.

4.2. Analysis of complexity

The experiments presented in Section 4.1 have been done configuring the overlay with a limited number of peers. Nevertheless, it is expected that, typically, P2P streaming applications involve hundreds of peers. The proposed method for overlay generation is based on the solution of an ILP problem, that generally has a complexity increasing with the problem dimensionality. In particular, analyzing the specific problem properties, the factors that strongly affect the model complexity are the number M of possibly active links for each peer and the presence of constraints (6) and (11), where the latter is, at the moment, still under investigation.

Figure 3 shows the complexity of the proposed method evaluated in term of the computational time needed to solve the overlay generation problem. Here we have evaluated the effect of the parameter M (i.e., the number of active links per node) in case of configuration C_1 with the additional constraints (9) and (10) ($I_{MAX} = O_{MAX} = 4$) for 200 peers. For each value of M it is indicated the value of the obtained throughput that, as can be noted by the values of the overall download and upload bandwidth reported in Fig. 3, is always lower than the maximum reachable value T, given by equation (12).

From the investigation of the experiments we deduce that the constraints (9) and (10) do not impact on the problem complexity. Furthermore, from the analysis of Fig. 3 it is clear that also with small values of parameter M it is possible to obtain the maximum value of reachable network throughput. This enables limited complexity for the overlay generation problem also in presence of a large number of peers. In further experiments not reported here, we have evaluated that for a values of M lower than 10, it is possible to solve the overlay generation problem in less than 10 seconds considering up to 500 peers, obtaining a throughput close to the maximum reachable value. In case of networks with more than 500 peers it is possible to define algorithms that split the peers in sub-groups and generate the overlay independently on each group.

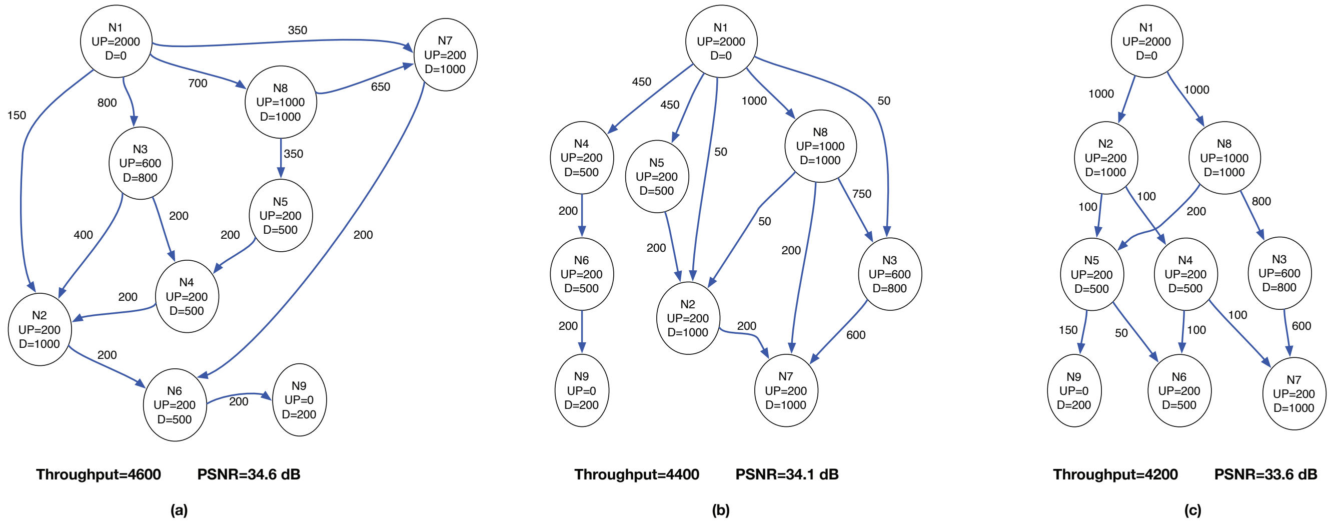

Fig. 2. Optimal mesh for the tested configuration specified in Table 1: (a) C_1, (b) C_2 and (c) C_3

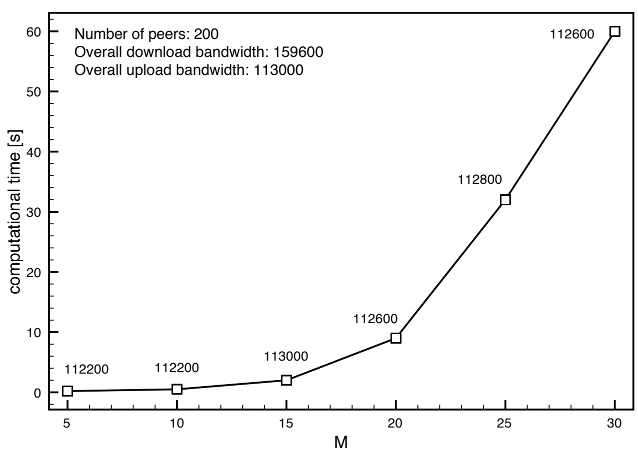

Fig. 3. Evaluation of the effect of the parameter M on the overlay generation complexity for 200 peers.

5. CONCLUSIONS

In this paper we have described a new approach for overlay generation in a P2P scalable video streaming system adopting an Integer Linear Programming approach to model the optimization problem. The ILP problem solution describes the overlay in terms of active connections between peers and data-rate to be transmitted on each link, with the aim to maximize the network throughput. The reported simulation results show that the proposed method enables high flexibility to describe different overlay topologies, such as tree, multi-tree and mesh. Furthermore, from the complexity analysis point of view, we have shown that, under an appropriate setting of the model parameters, the computational time needed to solve the problem can be maintained limited, and consequently the proposed method is suitable also to be included in a more realistic P2P streaming system.

6. REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding extension of the H.264/AVC standard," *IEEE Transaction on Circuit and System for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[2] J. Liu, S.G. Rao, B. Li, and H. Zhang, "Opportunities and challenges of peer-to-peer internet video broadcast," *Proceedings of the IEEE, Special Issue on Recent Advances in Distributed Multimedia Communications*, vol. 96, no. 1, pp. 11–24, 2008.

[3] P. Baccichet, T. Schierl, T. Wiegand, and B. Girod, "Low-delay peer-to-peer streaming using scalable video coding," in *Proc. International Packet Video Workshop*, 2007.

[4] S. Mirshokraie and M. Hefeeda, "Live peer-to-peer streaming with scalable video coding and network coding," in *Proc. of ACM Multimedia Systems*, 2010.

[5] N. Ramzan, E. Quacchio, T. Zgaljic, S. Asioli, L. Celetto, E. Izquierdo, and F. Rovati, "Peer-to-peer streaming of scalable video in future internet applications," *IEEE Communication Magazine*, vol. 49, no. 3, pp. 128–135, 2011.

[6] L. Lima, M. Mauro, T. Anselmo, D. Alfonso, and R. Leonardi, "Optimal rate adaptation with integer linear programming in the scalable extension of H.264/AVC," in *Proc. International Conference on Image Processing*, 2011.

[7] R. Iqbal and S. Shirmohammadi, "An analytical approach to model adaptive video streaming and delivery," in *Proc. Workshop on Advanced Video Streaming Techniques for Peer-To-Peer Networks and Social Networking*, 2010.

[8] C. Gkantsidis and P. Rodriguez, "Network coding for large scale content distribution," in *Proc. IEEE INFOCOM*, 2005.

[9] R. Bernardini, R. Rinaldo, and A. Vitali, "A reliable chunkless peer-to-peer architecture for multimedia streaming," in *Proc. Data Compression Conference*, 2008.

[10] ILOG INC., "ILOG CPLEX 12.1 reference manual," Incline Village: ILOG Inc., CPLEX Div., 2009.