Chunk Distribution in Mesh-Based Large-Scale P2P Streaming Systems: A Fluid Approach

Chunk Distribution in Mesh-Based Large-Scale P2P Streaming Systems: A Fluid Approach

Ana Paula Couto da Silva, Emilio Leonardi, *Senior Member*, *IEEE*,
Marco Mellia, *Senior Member*, *IEEE*, and Michela Meo, *Member*, *IEEE*

Abstract—We consider large-scale mesh-based P2P systems for the distribution of real-time video content. Our goal is to study the impact that different design choices adopted while building the overlay topology may have on the system performance. In particular, we show that the adoption of different strategies leads to overlay topologies with different macroscopic properties. Representing the possible overlay topologies with different families of random graphs, we develop simple, yet accurate, fluid models that capture the dominant dynamics of the chunk distribution process over several families of random graphs. Our fluid models allow us to compare the performance of different strategies providing a guidance for the design of new and more efficient systems. In particular, we show that system performance can be significantly improved when possibly available information about peers location and/or peer access bandwidth is carefully exploited in the overlay topology formation process.

Index Terms—P2P streaming systems, diffusion algorithms, overlay, fluid models

1 Introduction

Recently, we have been witnessing the emergence of a new class of popular P2P applications, namely, large-scale P2P video streaming. Both live and on-demand P2P streaming systems have the potential of changing the way we watch TV, providing ubiquitous access to a vast number of channels, personalizing the user's experience, and enabling roaming services.

Several commercial narrow-band (200-300 Kbit/s) P2P video streaming systems, P2P-TV for short, such as PPStream [1], PPLive [2], SOPCast [3], and TVants [4], just to mention the most popular ones, have recently been successfully exploited to broadcast particular events attracting millions of users worldwide. In [5], it was shown that the number of concurrent PPLive users reaches two million. A new and much more promising generation of high-definition commercial video P2P applications, such as Babelgum [6], Zattoo [7], and TVUnetworks [8], are at an advanced stage of prototyping and beta-testing. These systems are targeted to offer large-bandwidth video streams (1-5 Mbit/s) to a very large population of users (up to millions). The large bandwidth and the expected huge number of users' call for the development of some new tools that can help in understanding the impact of these new services on the network performance, designing new solutions that gracefully adapt to network conditions, are all driving the transition of the Internet from a network devoted to best effort data services into an integrated network providing also delay-sensitive high-bandwidth services such as TV broadcasting. The objective of this paper is to propose new analytical tools for the performance evaluation of P2P-TV systems.

This paper focuses on mesh-based, or unstructured, P2P-TV systems that adopt a swarm-like delivery of the video stream, based on organizing the information to be transmitted into small "chunks"; in particular, the paper investigates the impact of the overlay topology. In mesh-based P2P-TV systems, the overlay topology is built according to very simple rules. When joining the system, peer A contacts a central management node (often located with the source of video information) requesting a neighbor-list $\{N_A\}$, i.e., a (possibly partial) list of other peers to/from which peer A will send and receive chunks.

The neighbor-list $\{N_A\}$ assigned to a peer may be created by selecting k_0 peers. A *Random Choice* (RC) policy is adopted if this update follows a random strategy. This strategy, that has been studied in previous proposals [9], [10], [11], and [12], allows to construct random overlay topologies having good structural properties in terms of diameter (i.e., the length of the longest shortest path) and connectivity/resilience to peer churning. However, RC strategy completely ignores possible available information about peer attributes, such as peer location, peer access bandwidth, available bandwidth, etc. A natural question is, thus, whether it is possible to better exploit this information. In this paper, we answer the above question exploring alternative strategies for the generation of $\{N_A\}$. In particular, two other simple rules to build the overlay topologies are considered: in *location-aware* (LA) scheme, peers are connected to the closest peers (intended as the peers toward which exchanged packets perceive the least round-trip time (RTT)); in *hybrid* schemes (Hy), peers are connected partially to random peers and partially to the closest ones. Finally, a fourth topology construction mechanism is presented to explicitly leverage the peers' heterogeneity.

As main performance index for a P2P-TV system, we consider the *diffusion time* that is the time to distribute a chunk

- *A.P. Couto da Silva is with the Computer Science Departament, Federal University of Juiz de Fora, Instituto de Cincias Exatas (ICE), Rua José Lourenço Kelmer, s/n—Campus Universitário, Bairro São Pedro—CEP 36036-330—Juiz de Fora—Minas Gerais—Brazil. E-mail: anapaula.silva@ufjf.edu.br.*
- *E. Leonardi, M. Mellia, and M. Meo are with the Dipartimento di Elettronica, Politecnico di Torino, Duca degli Abruzzi 24, 10129 Torino, Italy. E-mail: {leonardi, mellia, meo}@polito.it.*

to all peers. A number of reasons make the diffusion time the key performance index. First, the reduction of delay translates into shorter start-up times, both when the application starts and when the user switches channel. Second, the live P2P-TV system is expected to provide the video stream to users with (roughly) the same delay, so as to avoid that geographically close users' access some time-critical content (e.g., breaking news, sport results) at different times. Finally, reducing delays implies that all the peers receive the same chunks at roughly the same time, making scheduling and redistribution of chunks easier under real-time constraint.

We propose simple fluid analytical models of the chunk diffusion time. We chose the adoption of fluid model technique since it permits to capture fairly well the main dynamics of large-scale complex systems comprising many interacting components with a model whose complexity is mainly independent from the system size. This allows us to easily analyze scenarios including millions of peers, which would be impractical with other modeling techniques including Monte Carlo approaches. With the proposed fluid models, performance can be compared under different network scenarios.

We consider *access-limited* scenarios in which the bottleneck is at the network access, as is the case of xDSL users, or *latency-limited* scenarios in which the end-to-end bandwidth is limited by the congestion control mechanisms, such as those implemented by TCP. Similarly, latency becomes the major constraint when explicit signaling is adopted to request/grant a chunk transmission among peers. Finally, we consider heterogeneous scenarios, in which, part of the peers have high-bandwidth access to the network, and part are connected by slow xDSL links.

The models are instrumental in comparing different systems so as to obtain important hints for the design of P2P streaming systems. To this purpose, we show that:

1. random topologies have excellent performance when the peer upload capacity is the transmission bottleneck;
2. location-aware topologies usually take the lead when the bottleneck is due to the congestion control mechanism;
3. hybrid scheme takes the best of the two, while it is never the best one, it always presents performance very close to those of the best scheme;
4. the creation of a cluster of large-bandwidth peers can be convenient when heterogeneous peer access capacity is considered; yet, it is important that connectivity of large-bandwidth peers with narrow-bandwidth peers is guaranteed.

2 Related Work

A few recent works have presented mesh-based P2P streaming architectures that incorporate swarm-like delivery (CoolStreaming/DONet [9], PALS [10], PULSE [11], and PRIME [12]). Mesh-based distribution systems do neither require any global form of coordination among peers, nor tight control of the overlay topology; they present undisputed potential advantages in terms of both scalability and resilience to churning with respect to other architectures and are today more promising solutions to scale up to millions of users [5].

Peers participating in mesh-based (unstructured) P2P-TV systems are organized in small "groups" that exchange information according to some scheduling algorithm. Inspired by file-swarming mechanisms (e.g., BitTorrent [18], Bullet [19]), the video stream is segmented in pieces (called chunks) that are independently distributed among peers. Each chunk is distributed along a "spanning tree," which is dynamically determined, thus enabling most peers to actively contribute their outgoing bandwidth.

Compared to traditional file sharing applications, incorporating swarm-like delivery into live P2P streaming applications is challenging due to: 1) the real-time streaming constraint requiring almost in-sequence and in-time arrival of chunks, and 2) the limited availability of future content. To this purpose, very little has been done in the direction of understanding the fundamental dynamics of unstructured P2P-TV. Only recently in [20], a fluid analytical model of unstructured P2P-TV systems has been presented. The model allows to evaluate the impact on the system performance of physical peer access bandwidth and play-out buffer, assessing at the same time the effect of peer churning. However, the model in [20] assumes a perfect mechanism according to which peers are at any moment able to completely exploit the bandwidth of the system. By doing so, the model fails to represent the chunk distribution dynamics and the possible effects on it, of both, the overlay topology structure and the chunk scheduling policy. The effect of misbehaving users has been analyzed in [25] and [26]. Effective incentive mechanisms to encourage peers to cooperate have been designed using a game-theoretical framework.

More closely related to our work, the papers [21], [22], and [23] analyze the impact of the chunk scheduling algorithms on unstructured systems performance. The analyses are carried out under the assumption that the overlay topology is a full mesh, and the system bottleneck is given by peers access links. In such cases, push-based schemes, according to which each peer transmits at most one chunk at a time to one of its neighbors, can be proved to be optimal [21]. The scheduling policy is in charge at every peer to select the peer and the chunk to be transmitted. In [21], the authors prove the rate optimality of the so-called *most deprived peer, random useful chunk* algorithm, i.e., a policy in which the peer chooses to distribute the chunk to the neighbor with the largest number of missing chunks. In [22], delay optimality (when the number of peers goes to infinite) of the *random peer, latest blind chunk* algorithm is proved assuming all peers are characterized by the same upload bandwidth (latest blind chunk refers to the fact that the latest chunk is selected regardless of whether or not the selected neighbor needs it). It turns out, however, that the delay performance of the former is poor due to the random chunk selection, while the rate performance of the latter is rather poor due to the blind nature of peer/chunk selection. More recently, in [23], it has been shown that joint optimal rate and asymptotic delay performance can be achieved using a *Random peer, latest useful chunk* scheduling algorithm; this result also applies only to the case of peers with the same upload bandwidth.

On the contrary, to the best of our knowledge, the problem of building an efficient overlay topology for unstructured systems has been addressed only in [24], for a scenario in which the stream delivery delay is mainly due to the transport network latency. Ren [24] adopts, however, a

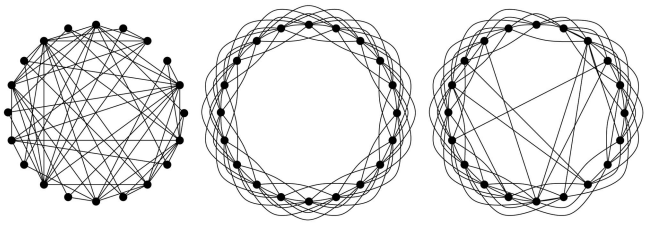

Fig. 1. Example of RC, LA, and Hy topologies, with $n = 20$ peers, $k_0 = 4$, and $p = 0.02$.

deterministic optimization approach to design efficient overlay topologies. This approach is complementary to ours, and requires a perfect knowledge of overlay link costs.

3 Strategies and Scenarios

We consider n peers with homogeneous access bandwidth. Each peer is assigned with a different location in a virtual space. Each peer receives a list of neighbors made of k_0 contacts selected according to different policies. Since links are bidirectional, a peer A contacted by another peer B becomes a neighbor of B. Thus, the average peer degree is $<k> = 2k_0$.

The list of neighbors of a peer A is denoted by $\{N_A\}$ and results from the strategies with which contacts are assigned to peers:

- **Random Choice** (RC): the k_0 contacts of peer A are randomly selected.
- **Location-Aware** (LA): $\{N_A\}$ contains k_0 peers selected based on their distance from A. The notion of distance we use is related to the RTT experienced by chunks.
- **Hybrid** (Hy): each of the k_0 contacts of A is chosen with probability $1-p$ based on its location, with probability p it is randomly chosen.

The resulting overlay topology can be described by a random graph [33].

Fig. 1 shows an example of the different topologies, in which, for simplicity, peers are located on a unidimensional ring. It can be easily observed that the adoption of different strategies generates overlay topologies with different macroscopic characteristics.

To compare the performance of RC, LA, and Hy strategies, we consider the following scenarios:

- **Access-bandwidth-limited:** in this case, the end-to-end chunk transfer rates are limited only by the access bandwidth of the peers. This is the case of chunks distributed among a peer population with narrow-band access by using an aggressive chunk transmission scheme (e.g., employing the UDP transport protocol). Moreover, the chunk propagation time results are negligible compared to the chunk transmission time, so that signaling delay (e.g., chunk acknowledgments, or request/grant messages) can be ignored.
- **Latency-limited:** in this case, chunks are exchanged among peers with moderately large access bandwidth. A window-based congestion control algorithm is adopted to support chunk transmission, such as the one implemented in TCP transport protocol. Therefore, the end-to-end chunk transfer rate is limited by the window-based congestion control mechanism, preventing peers from the full exploitation of their access bandwidth. In such conditions, the chunk download time depends on a number of parameters, including the Round-Trip Time closely related to the distance between peers. Furthermore, if the end-to-end latency is not negligible, the impact of signaling message exchange may become critical, since the RTT plays an important role also in this case. We consider two cases: connection throughput 1) inversely proportional to RTT and 2) inversely proportional to $\sqrt{\mathrm{RTT}}$ as an intermediate case.

Note that while in the former scenario (considered also in [21], [22], [23]) different chunks share the peers' bandwidth; in the latter scenario (considered in [24]), no competition for bandwidth among different chunks arises, since chunk transmission time does not depend on the peers' bandwidth.

A number of other assumptions hold. First, as in many related papers, we assume that the network is not under heavy load conditions so that it can always deliver all chunks [21], [22], [23]. Second, the effect of peer churning is neglected at first. Indeed, the chunk diffusion process lasts a few seconds, since only a marginal percentage of peers is expected to leave or join the system. Moreover, the impact of churning on system performance is expected to be similar under different strategies; thus, the relative performance is expected to be marginally affected by churning. However, in Section 7, we will explicitly assess its impact. Finally, we assume that the peers are uniformly located over the surface of a bidimensional Torus of unitary area.[1] We chose the Torus since it allows us to simplify the geometric description of the problem. Other surfaces can be taken into account as well, at the additional cost of either considering border effects, or deriving more complex expressions to define peer distances. Furthermore, the latency between two peers is assumed to be proportional to the length of the geodesic connecting two peers. While these choices may appear rather simplistic, e.g., ignoring the impact of the IP topology, they are justified in light of the several recent studies that have shown how synthetic coordinates over a low-dimensional euclidean space can be assigned to hosts in such a way that the distance between the coordinates of two hosts accurately predicts the communication latency between the hosts [27], [28], [29].

3.1 Considerations on Graphs Produced by Different Strategies

The adoption of a random choice strategy leads to an overlay topology that falls within the class of random graphs with assigned degree distribution $G(n, P(k))$, being:

$$P(k) = \begin{cases} 0, & \text{if } k < k_0, \\ \binom{n-1}{k-k_0} \left(\frac{k_0}{n-1}\right)^{k-k_0} \left(1 - \frac{k_0}{n-1}\right)^{(n-1)-k+k_0}, & \text{oth.} \end{cases}$$

1. The bidimensional (topological) Torus is the surface generated by the Cartesian product of two circles of unitary length. The Torus can be equivalently described as a quotient of the Cartesian plane under the identifications $(x, y) \simeq (x+1, y) \simeq (x, y+1)$.

TABLE 1
Summary of the Properties of Graphs Resulting from Different Strategies

Strategy	Graph	Arc length	Diameter
Random	$G(n, P(k))$	$\Theta(1)$	$\log(n)$
Loc. Aware	$G(n, d = 2)$	$\Theta(1/\sqrt{n})$	$\sqrt{n}$
Hybrid	$G_{WS}(n, k_0, p)$	$\Theta(1)$ and $\Theta(1/\sqrt{n})$	$\log(n)$

For large n, arc length (i.e., the euclidean distance between neighboring peers) is $\Theta(1)$[2] with high probability (i.e., with a probability that tends to 1 as $n \to \infty$). Note that, being the degree of each peer deterministically at least equal to k_0 by construction, the overlay graph can be proved with high probability to be at least k_0 connected, i.e., at least k_0 disjoint paths exist between any two peers, if $k_0 > 2$. This property is highly desirable for resilience to churning. In addition, the diameter of the overlay topology scales as $\Theta(\log n)$.

The adoption of the location-aware policy leads, instead, to an overlay topology with the properties of a bidimensional geometric random graph $G(n, d = 2)$. The overlay topology presents nice properties in terms of resilience to churning, since the resulting graph can be easily made k_0 connected. However, the graph diameter scales as $\Theta(\sqrt{n})$ in the bidimensional case and arc length is only $\Theta(1/\sqrt{n})$, since only close peers are connected.

Finally, the adoption of the hybrid policy generates an overlay topology with the macroscopic characteristic of a bidimensional Watts-Strogatz random graph $G_{WS}(n, k_0, p)$ in which local arcs, i.e., arcs between two close peers, have length $\Theta(1/\sqrt{n})$, while chords, i.e., arcs between distant peers, have length $\Theta(1)$. In Table 1, we summarize the properties of the generated graphs.

In the following, we relate the delay at which each chunk is received by all peers to the properties of the overlay topology. To this purpose, we devise simple fluid models. Since our focus is essentially on the overlay topology impact, we abstract from the particular swarm-like distribution mechanism, modeling the chunk distribution process as a branching process over the considered graph.

4 Chunk Distribution Models

4.1 RC Strategy in the Latency-Limited Scenario

We consider the distribution of a tagged chunk c. Let $I(t)$ denote the number of *inactive* peers, i.e., peers that at time t have already completed the download of c, and cannot serve neighboring peers, since the neighbors have already downloaded c as well. Inactive peers cannot anymore contribute to chunk distribution.

Let $S(t)$ denote the number of *seeds* or potential uploaders; seeds are peers that at time t have already completed the download of c and can upload it to some of the neighbors. Let $L(t)$ denote the number of *leechers* or potential downloaders; leechers have still to download chunk c and are neighbors of some seed. Finally, let $W(t)$ denote the number of *waiting* peers; i.e., the peers that cannot start to download the chunk at time t, since none of their neighbors is a seed.

2. Given two functions $f(n), g(n) \geq 0$, $f(n) = o(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) = 0$; $f(n) = O(g(n))$ means $\limsup_{n\to\infty} f(n)/g(n) = c < \infty$; $f(n) = \Theta(g(n))$ means $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

Denote by $r(t)$ the aggregate bit rate at which the considered chunk is transferred through the network at time t. If the bandwidth bottleneck is driven by congestion control dynamics, i.e., in the latency-limited case, the most convenient strategy for each seed is to transmit in parallel chunk c to all its leechers (i.e., no bandwidth constraints exist), so to maximize the aggregate transfer rate at which the transfer of chunks takes place. According to such a strategy, the aggregate transfer rate of the tagged chunk c is:

$$r(t) = B_c \min(K_L(t)S(t), L(t)),$$

where B_c is the average connection bandwidth, and $K_L(t)$ is the average number of neighbors that at time t are expected to be leechers of a given seed.

Now, since, by definition, no waiting peer can be neighbor of a seed:

$$K_L(t) = k_{eff}\frac{L(t)}{n - W(t)},$$

where $k_{eff} = 2k_0 - 1/2$ is the average degree of a peer reached by a randomly selected edge (see the Appendix, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2010.63, for more details).

Similarly, let $K_W(t)$ be, for a leecher, the average number of neighbors that at time t are expected to be waiting peers:

$$K_W(t) = k_{eff}\frac{W(t)}{n - I(t)}$$

since no inactive peer can be neighbor of a leecher.

The average aggregate rate at which downloads are completed is:

$$R(t) = \frac{r(t)}{q},$$

q being the chunk size.

Note that, as soon as leecher j completes the chunk download from seed i, j mutates into a seed and all the neighbors of j previously waiting are automatically mutated into leechers. In addition, i becomes inactive if no other leecher can be found among its neighbors.

The above mutation rules can be expressed by the following system of continuous-time differential equations that describe the average dynamics of $(I(t), S(t), L(t), W(t))$:

$$\frac{dI(t)}{dt} = \frac{R(t)}{K_L(t)}, \tag{1}$$

$$\frac{dS(t)}{dt} = \left[1 - \frac{1}{K_L(t)}\right]R(t), \tag{2}$$

$$\frac{dL(t)}{dt} = [K_W(t) - 1]R(t), \tag{3}$$

$$\frac{dW(t)}{dt} = -K_W(t)R(t), \tag{4}$$

$$W(t) + I(t) + S(t) + L(t) = n. \tag{5}$$

Equation (1) indeed, describes the dynamics of the inactive peers, whose number increases with time at rate $R(t)/K_L(t)$, since every seed transfers the chunk to $K_L(t)$ neighbors, on average, before becoming inactive. Equation (2) represents

the dynamics of the seeds. Leechers become seeds at an aggregate rate $R(t)$ while, at the same time, seeds become inactive at rate $R(t)/K_L(t)$. Equations (3) and (4) describe, respectively, the dynamics of leechers and waiting peers. Note that since a waiting peer is turned into a leecher as soon as the first among its neighbors gets a copy of the chunk, the rate at which waiting peers become leechers is given by $K_W(t)R(t)$. At last, (5) represents the fact that the sum of all peers is constantly equal to n at all times.

4.2 RC Strategy in the Access-Bandwidth-Limited Scenario

When the chunk transmission time is imposed by peers' access bandwidth, i.e., in the access-bandwidth-limited case, different chunks compete for the peer bandwidth. The common assumption in these cases is that the peer upload bandwidth constitutes the system bottleneck, i.e., peer download bandwidth is much larger than peer upload capacity, as in xDSL systems. In such a case, push-based schemes, according to which each peer transmits at most one chunk at a time, have been proved to be optimal [22], [23]. A scheduling policy is necessary at every seed to select which leecher and which chunk to transmit. It has also been proved that, to reduce the chunk delivery delay, seeds have to preferentially offer the "latest chunk." We, therefore, consider such a policy.

Taking into account the above considerations, we extend the previous model to consider the new constraints. First, we observe that the aggregate bit rate at which the considered chunk is transferred through the network is given by:

$$r(t) = B_u \min(L(t), S(t)),$$

B_u being the peer average upload bandwidth. Second, given the tagged chunk c, the seeds are given by those peers that at time t: 1) have already completed the download of c, 2) have leechers in their neighbors, and 3) have not completed yet the download of chunks that are more recent than c. Under stability condition, the rate at which peers are downloading chunks must be equal, on average, to the rate at which new chunks are emitted by the source; defined by λ the source chunk rate, seeds $S(t)$ receive fresher chunks with an aggregate average rate $\lambda S(t)$. As a result of previous arguments, the rate at which seeds become unavailable to serve chunk c is:

$$\gamma(t) = \max\left(\frac{R(t)}{K_L(t)}, \lambda S(t)\right).$$

The dynamics of $(I(t), S(t), L(t)W(t))$ are, therefore, driven by (3), (4), (5) and:

$$\frac{dI(t)}{dt} = \gamma(t) \tag{6}$$

$$\frac{dS(t)}{dt} = R(t) - \gamma(t) \tag{7}$$

4.3 Hy Strategy

4.3.1 Monodimensional Small-World Graphs

We first consider a unidimensional Watts-Strogatz graph. We then extend our study to the bidimensional case. Peers are deployed on a circumference as in Fig. 1. The access bandwidth case can be easily obtained extending Section 4.2 by simply considering the node degree distribution in Watts-Strogatz graph rather than the one of $G(n, P(k))$ case, i.e., by considering the correct k_{eff} factor. We treat the underlying peer lattice as a continuum rather than a discrete lattice.

We describe the latency-limited case. The chunk diffusion process originates from the source peer, and it propagates through both local arcs and possible chords. The set of peers reached exploiting only local arcs forms a symmetric interval with respect to the source on the continuous ring. It grows at rate $2k_0r_a(1-p)$, with $1/r_a$ being the average time to download a chunk through a local arc. The presence of sporadic chords acts as a shortcut that allows to reach peers that are far away on the circumference. Peers reached by means of chords become auxiliary seeds; from them, the chunk propagates through local arcs to contiguous peers, covering new intervals on the ring.

Following an approach that generalizes [34], we represent the dynamics of the chunk in the network using two auxiliary variables: $Z(t)$ and Y(t). $Z(t)$ represents the number of different intervals covered by the chunk at time t, while $Y(t)$ denotes the number of useful chords at time t, i.e., chords that may potentially support the download of the chunk at time t.

Let $1/r_c$ be the average time needed to download a chunk through a chord. The system dynamics are described by the following system of differential equations:

$$\frac{dZ(t)}{dt} = r_c Y(t)\frac{W(t)}{n} - 2r_a(1-p)k_0 Z(t)\frac{Z(t)-1}{W(t)+L(t)}$$

with $\frac{W(t)}{n}$ being the probability that a randomly selected chord reaches a waiting peer. The first term on the right represents the rate at which new intervals are generated due to chunk transmission through chords; the second term represents the rate at which intervals merge because of the endpoints coming into contact. For this purpose, observe that at time t the distribution of the size of gaps between intervals can be approximated by the distribution of the smallest number among $Z(t)-1$ uniformly distributed numbers between 0 and $W(t)+L(t)$ [34]. The last term tends to an exponential distribution with average $\frac{W(t)+L(t)}{Z(t)-1}$ for $n \to \infty$ (i.e., $Z(t)$ and $W(t)+L(t)$ jointly tending to infinity).

The dynamics of $Y(t)$ can be given by:

$$\frac{dY(t)}{dt} = 2r_a p k_0 Z(t) - r_c Y(t),$$

where the first term represents the rate at which new chords become active and the second term represents the aggregate rate at which downloads through chords terminate.

Dynamics of $(I(t), S(t), L(t), W(t))$ are easily obtained:

$$\frac{dI(t)}{dt} = 2r_a(1-p)k_0 Z(t), \tag{8}$$

$$S(t) = L(t) = 2k_0 Z(t). \tag{9}$$

The access-bandwidth-limited case can be modeled in a similar way by properly setting the chunk download rate. In particular, $r_a = r_c$, and a chunk at a time is delivered by each seed. At last, contention with other chunks for the access bandwidth can be modeled as for the RC strategy.

4.3.2 Generalization to the Bidimensional Watts-Strogatz Graphs

Similarly to the monodimensional case, the chunk propagates from the source peer reaching both local arcs and chords. The chunk propagates through local arcs between contiguous peers, forming a circular domain, centered at the source. Circle radius increases at rate

$$\theta = \sqrt{\frac{k_0}{\pi n}\ (1-p)} r_a,$$

with $1/r_a$ being the average chunk download time through a local arc. The presence of sporadic chords allows to reach far away peers on the Torus, which become then auxiliary seeds. From these peers, the chunk propagates through local arcs to contiguous peers, forming new circles.

Let $Z(\rho, t)$ be the number of circles of radius ρ at time t; let $Y(t)$ denote the number of useful chords at time t. The evolution of $Z(\rho, t)$ is driven by the following partial differential equation:

$$\frac{\partial Z(\rho,t)}{\partial t} = -\theta \frac{\partial Z(\rho,t)}{\partial \rho} + r_c Y(t) \frac{W(t)}{n} \delta(\rho) \qquad (10)$$

with $\delta(\cdot)$ being the Dirac function. The evolution of $Y(t)$ is given as before by:

$$\frac{dY(t)}{dt} = -p\frac{dW(t)}{dt} - r_c Y(t).$$

Finally, $W(t)$ can be geometrically evaluated from $Z(\rho, t)$ describing the fraction of surface that has still to be covered (the surface being a unitary torus for simplicity). Thus, we have the result:

$$W(t) = n e^{\int_0^{\theta t} Z(\rho,t)\log(1-2\pi\rho)d\rho}.$$

Notice that, being (10) a first order linear Partial Differential Equation, it exists a closed form expression for $Z(\rho, t)$ in terms of the forcing terms $Y(t)$ and $\frac{W(t)}{n}$, it results:

$$Z(\rho,t) = \begin{cases} 0, & \text{if } \rho > \theta t, \\ \frac{r_c}{n} Y\big(\rho - \frac{r}{\theta}\big) W\big(t - \frac{\rho}{\theta}\big), & \text{if } \rho \le \theta t. \end{cases}$$

4.4 LA Strategy

The evolution of chunk distribution can be obtained from the Watts-Strogatz model by setting $p = 0$, i.e., no chord exists. Therefore, chunk propagates through local arcs only between contiguous peers, forming a single circle centered at the source peer. Circle radius increases at rate

$$\theta = \sqrt{\frac{2k_0}{\pi n}} r,$$

$1/r$ being the average chunk download time through an arc. We have:

$$L(t) + W(t) = \begin{cases} n(1 - \pi(\theta t)^2), & \theta t \le \frac{1}{2}, \\ 4n\Big[\frac{\sqrt{2}}{2}\theta t \sin\big(\frac{\pi}{4} - \arccos\big(\frac{1}{2\theta t}\big)\big), \\ \quad -(\theta t)^2\big(\frac{\pi}{4} - \arccos\big(\frac{1}{2\theta t}\big)\big)\Big], & \frac{1}{2} \le \theta t \le \frac{\sqrt{2}}{2}. \end{cases}$$

Note that a rather strong degree of symmetry is induced by the fact that LA policy chooses peers in neighbor lists exploiting peer location information.

5 Experiments

5.1 Model Validation

To validate the model, we compare model predictions against results obtained with a Monte Carlo simulation (the simulator is available online [30]).

A simulation is organized in two phases. First, a random overlay topology is generated according to the RC, Hy, and LA models. Peers are placed on a square with side equal to 100 units. During the second phase, the chunk distribution dynamics are simulated until all peers have completed the content download. Chunks are generated periodically by the source at a rate $\lambda = 1$; the chunk length L is normalized to 1 so that the interchunk emission time is $T = 1$. The system comprises $n = 10{,}000$ peers. Each peer adopts a window mechanism to avoid the distribution of chunks that are not useful anymore, e.g., because their delay exceeds the play-out buffer. The size of the window has been fixed to $W = 20$ chunks. Simulation ends when 500 chunks have been distributed, and delay distribution has been obtained averaging over all chunks.

Both the bandwidth-limited and latency-limited cases are simulated. In the first case, the chunks download rate is independent from peer distance, it being limited only by the peer access bandwidth. We have assumed that 1) all peers have an infinite download and finite upload bandwidth B_u, 2) peers implement a *random peer—latest useful chunk* scheduling policy [23]. Conversely, in the latency-limited case, the chunk download rate is supposed inversely proportional to the RTT between peers, which, neglecting the packet transmission time, has been, in turn, set proportional to the physical distance between peers. Each peer implements a scheduling policy, according to which as soon as it gets a new chunk c, it pushes c to all the neighbors that still need it. Notice that, while our models scale up to millions of peers, Monte Carlo simulations are possible only on relatively small-scale systems.

We start by considering the effect of the average peer degree $<k>$ on the speed at which the chunk is propagated through the network; the RC strategy is adopted. We consider a system comprising $n = 10{,}000$ peers, and values of $k_0 = 2, 4, 6, 8$ (to which an average graph degree $<k> = 2k_0$ corresponds).

Fig. 2 shows the evolution of $W(t) + L(t)$, i.e, peers which still have to complete the chunk download at time t, considering the latency-limited scenario. In this case, peer access bandwidth is unbounded, and the average connection throughput is inversely proportional to the RTT. Note that an inversely proportional relation between the connection throughput and the associated RTT is a standard assumption when window-based congestion control protocols like TCP are considered. Similarly, RTT can be the major constraint also in case explicit signaling is required to run the scheduling algorithms, e.g., to receive an explicit grant after a chunk request message.

First, observe the evolution of the number of peers that still have to receive the chunk. At the beginning, the chunk diffusion mechanism is slow since the number of seeds $S(t)$ is limited. A sharp decrease in the number of waiting peers is then observed, due to the exponential growth of $S(t)$. The effect of seeds becoming inactive is perceived in the last part of the chunk diffusion evolution, during which few

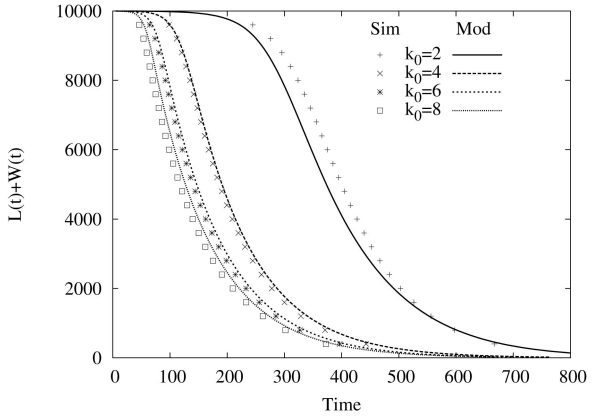

Fig. 2. RC strategy. Dynamics of $L(t)+W(t)$ for different values of k_0 when the system is latency limited; the number of peers in the system is $n = 10^4$. Time is expressed in arbitrary units.

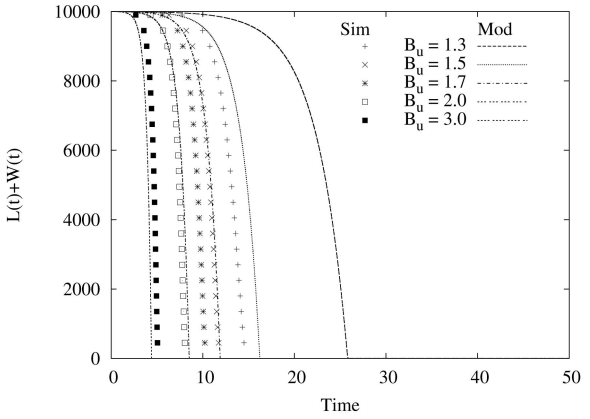

Fig. 4. RC strategy. Dynamics of $L(t)+W(t)$ when the system is access-bandwidth-limited for different values of B_u; $n = 10{,}000$ and $k_0 = 8$. Time is expressed in arbitrary units.

unlucky peers have not been served yet. In the following, we refer to these phases as "bootstrap," "explosive," and "final" phases. Second, a performance improvement is observed by increasing k_0, since seeds can exploit the increased parallelism in serving neighbors. This, in particular, affects the bootstrap phase of the propagation process, during which the number of seeds grows according to an exponential function of $\Theta(t^{k_{eff}})$. At last, observe that modeling and simulation results show an excellent match.

Fig. 3 shows, instead, the evolution of $W(t)+L(t)$, when the performance is limited by the peers' access bandwidth in the case of $B_u = 2.0$. Few considerations hold in this case. First, in general, there is also in this case a good match between the model prediction and the simulation results. Second, the connectivity degree k_0 has little impact on the delay in this scenario. Indeed, provided that the peer degree is sufficiently high to fully utilize the upload bandwidth of peers (in the considered scenario, this happens when the number of neighbors to serve is larger than B_u/λ), there is little advantage to further increase k_0.

To this extent, the results of the model are completely insensitive to k_0. On the contrary, by simulation, there is some difference between the $k_0 = 2$ and the other cases. This is due to the fact that the neighborhood size is variable in the simulator (with minimum $k_0 = 2$, and average $2k_0 = 4$), so that some peers that have only two neighbors cannot fully exploit their bandwidth, since only one neighbor can be served, and since the other peer is the seed that sent c to it.

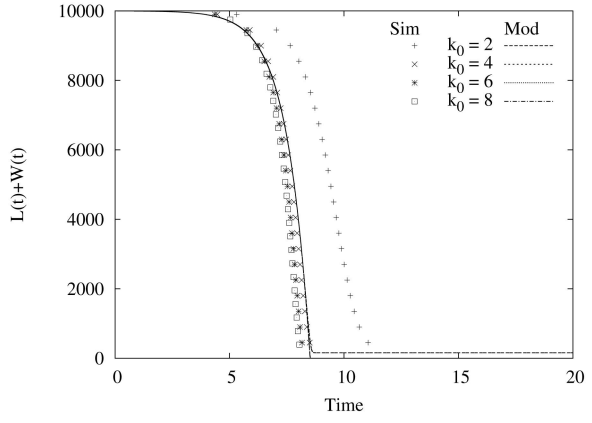

Fig. 3. RC strategy. Dynamics of $L(t)+W(t)$ for different values of k_0 when the system is access-bandwidth-limited; $B_u = 2.0$ and $n = 10{,}000$. Time is expressed in arbitrary units.

Fig. 4 shows the evolution of $W(t)+L(t)$, for different values of B_u when $k_0 = 8$. Model predictions are in good agreement with simulation results when the systems moderately loaded (i.e., for $B_u = 3.0$ and $B_u = 2.0$), while it tends to overestimate the delay when the system becomes highly loaded (i.e., for $B_u = 1.5$ and $B_u = 1.3$).

The reason why the model tends to overestimate the chunk delivery delay when the system becomes resource-constrained is related to the fact that, according to the fluid approximation, the transmission capacity of peers is partitioned among different chunks in a static fashion, i.e., every peer devotes a time exactly equal to $1/\lambda$ to the transmission of every chunk. Instead, in the real system, since chunks arrival process at peers exhibits some jitter, the service time devoted by a particular peer to the chunk transmission may be different from chunk to chunk. In other words, a chunk c that arrives at a tagged peer p with a delay smaller than average is expected to receive more service by p than average; this in light of the fact that the expected time until a newer chunk is received by p is larger than average. As a consequence, peers that are favored in the distribution process of a given chunk can devote more time to the retransmission of the considered chunk. This correlation between delay with which a chunk is received by a peer and the time devoted to its transmission is the origin of the mismatch between simulation and model results when the system becomes resource-constrained. Unfortunately, this effect can be captured only by a model that considers statistics of order greater than one, and hence, renouncing to the simplicity and scalability of the fluid approach.

In conclusion, model predictions appear very accurate in the latency-limited scenario, while some discrepancy can be observed in the more challenging access-bandwidth-limited case due to bandwidth competition among different chunks. Nevertheless, the model captures all the main phenomena as will be confirmed by results reported in Section 6. We wish also to emphasize that a similar match between model predictions and simulation results is observed when either the Hybrid strategy or the Location-Aware strategy is adopted.

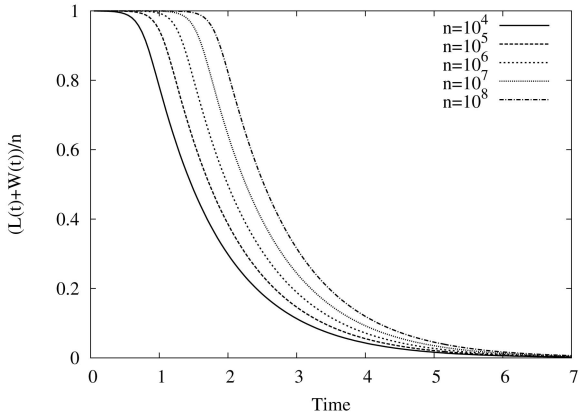

Fig. 5. RC strategy. Dynamics of $L(t) + W(t)$ when the system is latency-limited and connection throughput is inversely proportional to the RTT; for different values of n.

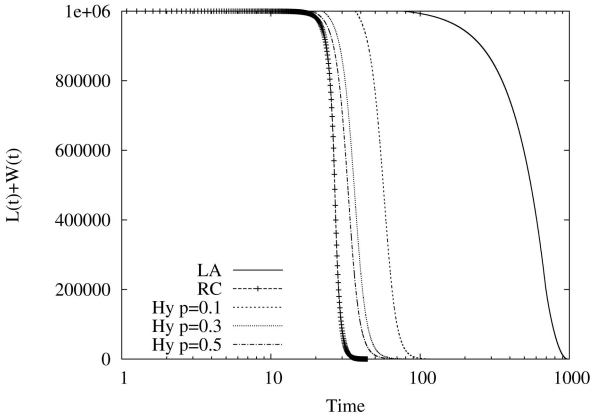

Fig. 7. Dynamics of $L(t) + W(t)$ for different strategies when the system is access bandwidth limited with $n = 10^6$. Time is expressed in arbitrary units.

5.2 Impact of the Number of Peers

Figs. 5 and 6 show the evolution of the fraction of peers that has still to receive the chunk for different values of the peer population n. The latency-limited case is considered in Fig. 5, while the access-bandwidth-limited case is considered in Fig. 6; for both cases, we have $k_0 = 6$.

As expected, the chunk delivery time increases when the number of peers in the system increases; however, observe that its dependency on n is as weak as $\log n$ for both the access-limited and latency-limited case. This confirms that mesh-based P2P streaming systems provide a highly scalable approach for the distribution of live streaming contents to large populations of users. At last, we notice that Monte Carlo simulations were possible only up to $n = 10^4$ peers.

5.3 The Effect of the Different Strategies

In this section, we compare the performance of the three different strategies. We consider a scenario with $n = 10^6$ peers randomly placed in a square of side 1,000, with $k_0 = 6$. Only results obtained solving the fluid models are presented.

Fig. 7 refers to the access-bandwidth-limited scenario. In this case, the system performance is strongly dependent on the distance and diameter properties of the overlay graph. Thus, it is not surprising that RC and Hy strategies outperform the LA strategy. The adoption of both RC and Hy policies leads to an overlay topology whose diameter increases as $\log(n)$, while the diameter increases as $\sqrt{n}$ when LA is employed. In conclusion, LA performance is significantly the worst. RC graphs outperform Hy topologies but, for moderately large values of p, performance of Hy becomes close to that of RC. In particular, the choice of p affects the initial bootstrap duration, while both the explosive and final phases are marginally affected.

The previous scenario dramatically changes when the system is latency-limited, so that chunk download throughput is inversely proportional to peer physical distance. Fig. 8 shows that the LA strategy outperforms RC by more than two orders of magnitude. This is due to the fact that links in the LA topology are chosen so as to minimize the connection RTT. This has beneficial impacts on the initial bootstrap phase, compensating the negative effect due to the larger topology diameter, which, in turn, affects the explosive phase. Note that Hy strategy performs well due the high clustering degree (i.e., the presence of many local arcs) of Watts-Strogatz graphs.

At last, we consider the intermediate scenario in which chunk transfer time is inversely proportional to the square root of RTT (we neglect the packet transmission time). In this case, as reported in Fig. 9, RC performs similarly to Hy and

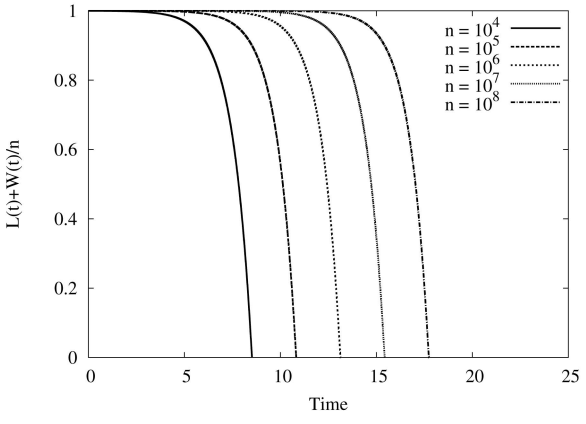

Fig. 6. RC strategy. Dynamics of $L(t) + W(t)$ when the system is bandwidth-limited with $B_u = 2.0$; for different values of n.

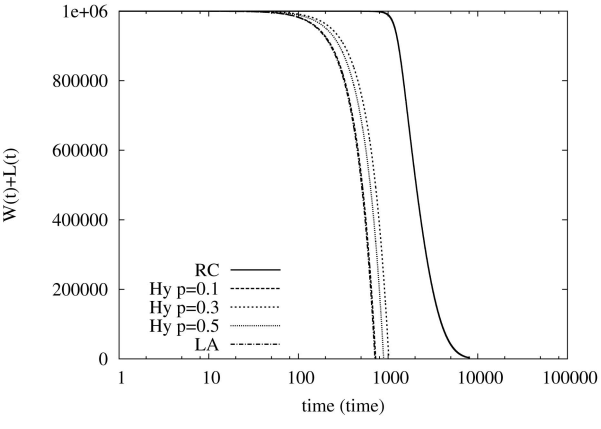

Fig. 8. Dynamics of $L(t) + W(t)$ for different strategies when the system is latency limited and connection throughput is inversely proportional to RTT; $n = 10^6$. Time is expressed in arbitrary units.

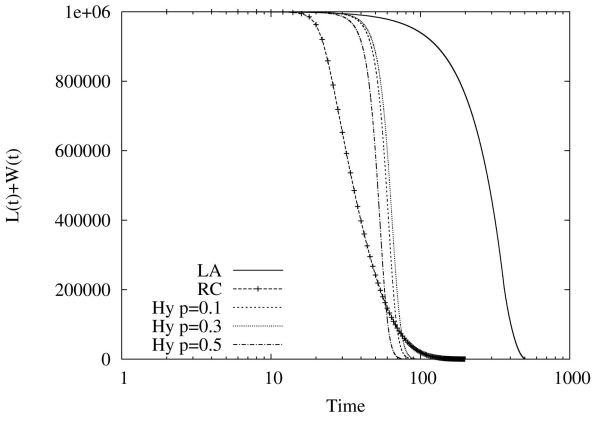

Fig. 9. Dynamics of $L(t) + W(t)$ for different strategies when the system is latency-limited and connection throughput is inversely proportional to $\sqrt{RTT}$; $n = 10^6$. Time is expressed in arbitrary units.

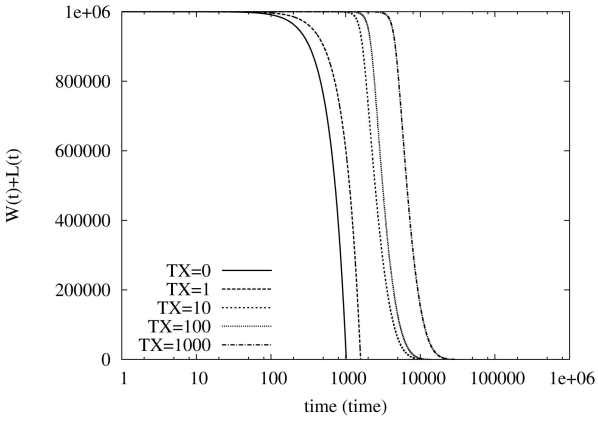

Fig. 11. Dynamics of $L(t) + W(t)$ for Hy strategy with $p = 0.3$ when the system is latency-limited and connection throughput depends on the packet transmission time TX and latency between peers; $n = 10^6$. Time is expressed in arbitrary units.

better than LA. Note that the adoption of the RC strategy, that exploits arcs with large RTT, leads to a quicker bootstrap phase, followed by a less steep explosive phase, and a longer final phase. On the contrary, the Hy strategy, exploiting arcs with both small and large values of RTT, pays a little longer bootstrap, but exhibits much steeper explosive phase.

As concluding remark, note that RC performs rather well in scenarios in which the chunk transfer time is weakly dependent on the connection RTT; ignoring location information becomes significantly penalizing when the chunk transfer time is closely related to the RTT. By contrast, simple variations of the RC strategy as those implemented within the Hybrid strategy significantly increase the system robustness.

5.4 Impact of the Transmission Time

We now investigate further the latency-limited scenario. So far, we have considered that, in this case, the connection throughput depends only on the latency. We now assume that the RTT between peers is the sum of two components: 1) the transmission time (TX) of the packets that compose the chunk, and 2) the latency, which is typically proportional to the distance between peers (in our settings, the latency exactly corresponds to the euclidean distance between peers).

Figs. 10, 11, and 12 refer to RC, Hy with $p = 0.3$, and LA, respectively, and report the dynamics of $L(t) + W(t)$ versus time for different values of the transmission time, TX, which varies from 0 to 1,000.

LA is the strategy whose performance exhibits the strongest dependency on the packet transmission time. This because the impact of the packet transmission time on the RTT, and thus, on the transfer bandwidth, is more and more significant as the latency of the path between peers reduces (we recall that LA exclusively exploits chunk transmissions on low-latency paths). On the contrary, the dependence of RC performance on the packet transmission time is much weaker, since the impact of packet transmission time on the RTT is much less significant when the physical distance between peers increases. We remind that RC does not exploit peer proximity information, while building the overlay. At last, observe that the performance of Hy policy exhibits a rather significant dependence on the packet transmission time. When the packet transmission time is negligible, Hy performs very similarly to the LA strategy, while when the packet transmission time increases the performance of Hy, it becomes closer and closer to that

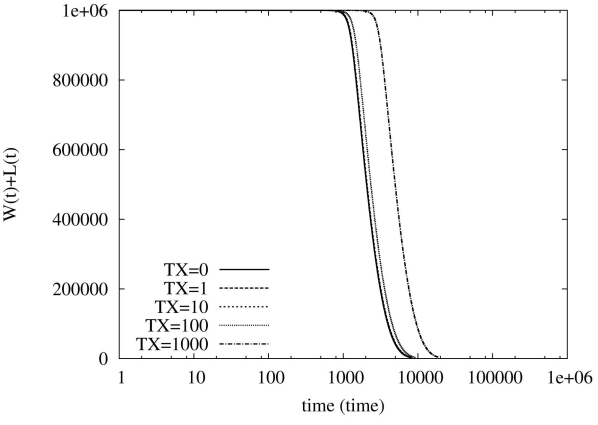

Fig. 10. Dynamics of $L(t) + W(t)$ for RC strategy when the system is latency-limited and connection throughput depends on the packet transmission time TX and latency between peers; $n = 10^6$. Time is expressed in arbitrary units.

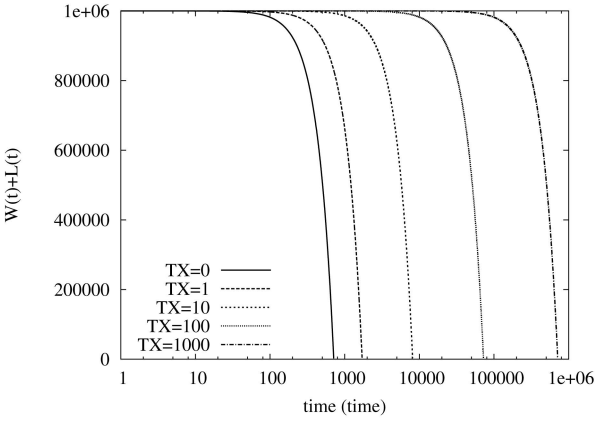

Fig. 12. Dynamics of $L(t) + W(t)$ for LA strategy when the system is latency-limited and connection throughput depends on the packet transmission time TX and latency between peers; $n = 10^6$. Time is expressed in arbitrary units.

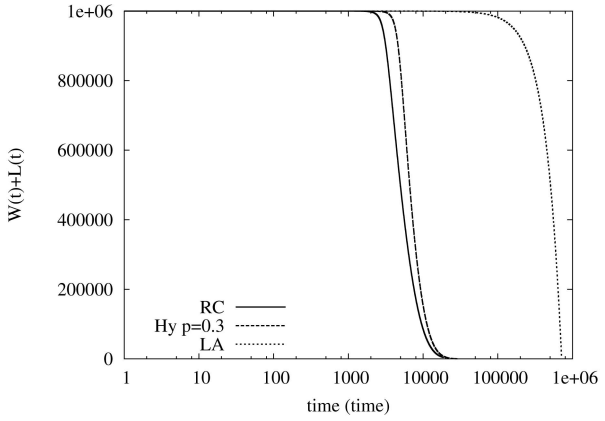

Fig. 13. Dynamics of $L(t) + W(t)$ for different strategies when the system is latency-limited and connection throughput depends on the packet transmission time $\mathrm{TX} = 1{,}000$ and latency between peers; $n = 10^6$. Time is expressed in arbitrary units.

of the RC. For all the values of the transmission range, the performance of the Hy strategy is rather close to the best performing strategy. To ease the direct comparison among different strategies, Fig. 13 shows the performance of the three schemes when the packet transmission time is 1,000.

6 The Effect of Different Peers Access Bandwidth

In this section, we account for user heterogeneity. Consider a scenario in which two classes of peers coexist: residential peers accessing the network through a xDSL connection, and business peers exploiting high bandwidth access. We consider that peers in both classes are access-bandwidth-limited, since in this scenario the model proved to be slightly less accurate in some cases (similar considerations to those drawn below can be derived from a mixed scenario in which high-bandwidth peers are latency-limited). Fluid equations can be easily obtained generalizing the approach of Section 4.2 to deal with two classes of users; for the sake of brevity, details are not reported here.

Let n_R denote the number of residential peers, while n_B represent the number of business peers, with $\frac{n_B}{n_B+n_R} = 0.1$. We assume that business peers have an upload access

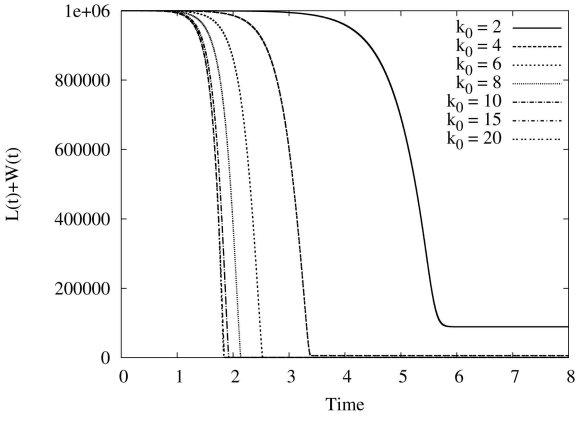

Fig. 14. Dynamics of $L(t) + W(t)$ for different values of k_0 in the heterogeneous scenario. Time is expressed in seconds.

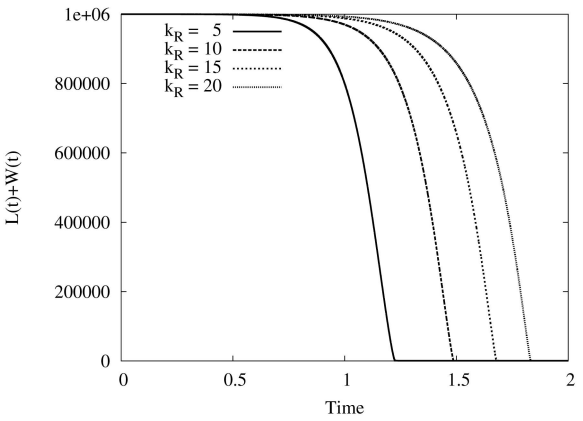

Fig. 15. Dynamics of $L(t) + W(t)$ for different values of k_R in the heterogeneous scenario. Time is expressed in seconds.

bandwidth $B_B = 10$ Mbit/s, while residential peers have an upload bandwidth limited to $B_R = 640$ Kbit/s. Let the chunk size be 100 Kbit long, and the video rate 500 Kbit/s.

Fig. 14 shows the evolution of $W(t) + L(t)$ for different choices of k_0. In this scenario, the choice of the degree appears rather critical. Small values of k_0 lead to a waste of the business users' bandwidth, with $2k_0 - 1 > B_B/\lambda$ being the condition to fully utilize the bandwidth; this condition is satisfied only by values of $k_0 > 11$. For $k_0 = 2$, there are some chunk losses.

Fig. 15 shows the the evolution of $W(t) + L(t)$ for an overlay in which the neighbor lists of business users and residential users are of different sizes. The size of the neighbor list for business users has been fixed at $k_B = 20$ while the size of neighbor list for residential users k_R is varied. Increasing k_R the results worsen. This counter-intuitive result can be explained considering the fact that decreasing k_R indirectly we increase the percentage of edges connecting two business users. Coupled with the random peer-latest useful chunk scheduler, this favors the distribution of the chunk among business users in the bootstrap phase, thereby increasing the initial chunk replication rate.

6.1 Enforcing Clustering among Classes

To further favor the distribution of chunks among business users, we propose the following *bandwidth-aware* topology construction strategy according to which every business peer receives a neighbor list comprising k_{B2B} business peers and k_{B2R} residential peers. $k_{B2B} + k_{B2R} = k_B$ is the number of neighbors of a business peer; a residential peer, instead, gets k_R neighbors randomly selected among the whole population. The rationale behind this strategy is to form a "backbone" of business peers that speed up the chunk spreading process during the bootstrap phase, since the chunk can be very quickly distributed among business peers.

Fig. 16 compares the performance of the bandwidth-aware strategy against those of the RC strategy. The neighbor list sizes have been set to $k_B = 20$ for business users and $k_R = 5$ for residential users, while we vary the value of k_{B2B}. Fig. 16 shows that performance can be significantly improved when the overlay topology is carefully designed in such a way to better exploit the available bandwidth of the business peers. In particular, the k_{B2B} has a large impact on the bootstrap phase duration, which is

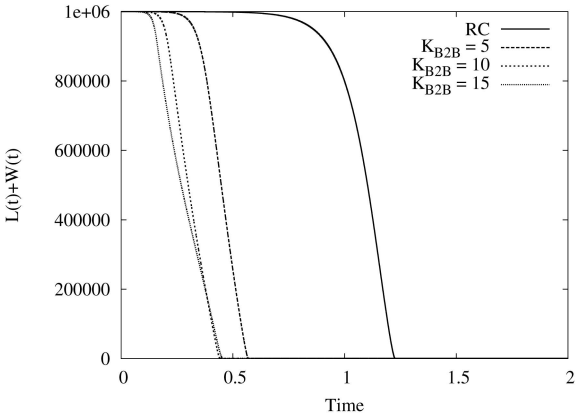

Fig. 16. Dynamics of $L(t)+W(t)$ for values of k_{BB} in the heterogeneous scenario. Time is expressed in seconds.

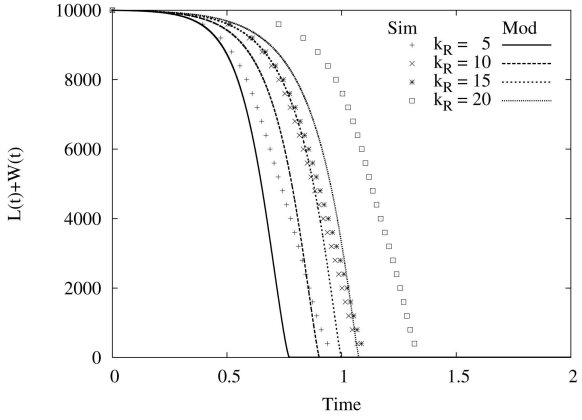

Fig. 17. Dynamics of $L(t)+W(t)$ for the heterogeneous scenario. Different curves refer to different values of K_R. Time is expressed in seconds.

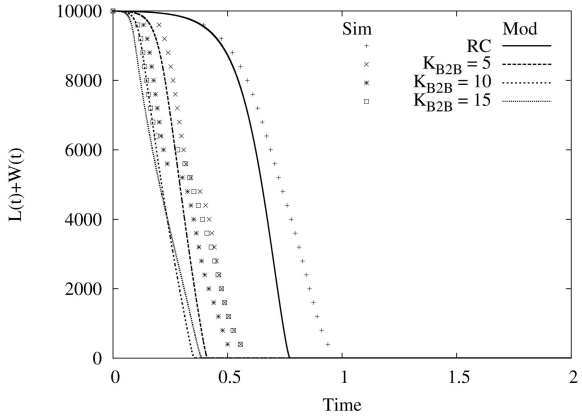

Fig. 18. Dynamics of $L(t)+W(t)$ for the heterogeneous scenario. Different curves refer to different values of K_{B2B}. Time is expressed in seconds.

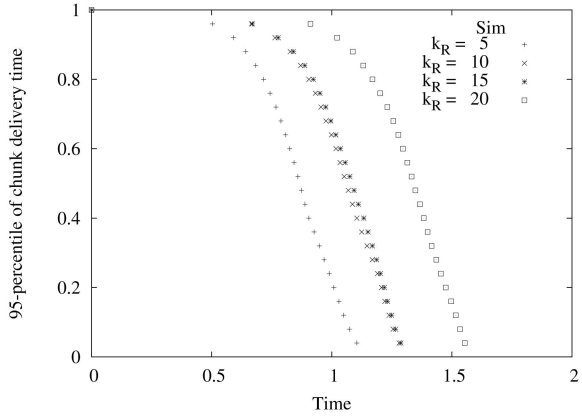

Fig. 19. CDF of the 95th percentile of chunk delivery delay for the heterogeneous scenario. Different curves refer to different values of K_R. Time is expressed in seconds.

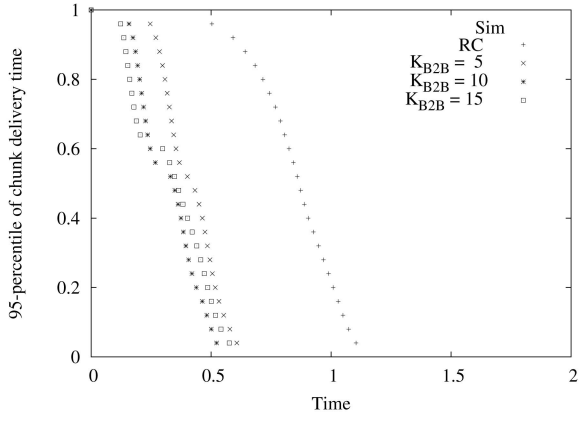

Fig. 20. CDF of the 95th percentile of the peer chunk delivery delay for the heterogeneous scenario. Different curves refer to different values of K_{B2B}. Time is expressed in seconds.

greatly reduced if large broadband peers are clustered. Notice, however, that the performance does not further improve for $k_{B2B} > 10$. This, because the advantage of favoring a quick distribution of the chunk among broadband peers, is compensated by the fact that business peers are then unable to effectively transfer the chunk to a large number of residential peers, since k_{B2R} becomes too small. As extreme case, we indeed observe that when $k_{B2B} = 20$, the overlay topology becomes disconnected (i.e., $K_{B2R} = 0$), and only the business peers receive the chunks (the source behaves as a business peer).

To validate the model in this heterogeneous scenario, we report in Figs. 17 and 18 a direct comparison between model prediction and simulation results for the case in which n has been scaled down to 10,000. In particular, we have considered the effect of varying k_R ($k_B = 20$) and the effects of varying K_{B2B}. The model correctly captures all the main phenomena, even if some discrepancy between model and simulation results can be observed. Observe, for example, as both simulation and model results agree on the fact that the delivery delay of peers worsens increasing k_R, as well as performance does not anymore improve when K_{B2B} increases above the value 10.

At last, for completeness, we report in Figs. 19 and 20 the CDF of the 95th percentile of peer delivery delay, as obtained with our simulator. Fig. 19 considers the effect of varying K_R while Fig. 20 considers the effect of varying K_{B2B}. The 95th percentiles of the chunk delivery delay follow similar trends with respect to the curves representing the average chunk distribution delay (Figs. 17 and 18). This justifies our approach according to which we use the fluid models for making design choices and comparing strategies, even if the models capture only the average dynamics.

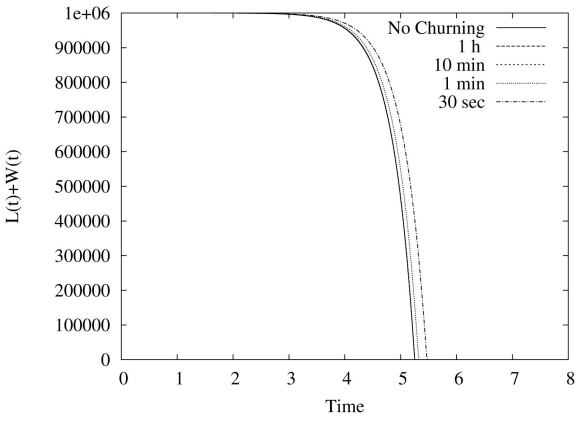

Fig. 21. Dynamics of $L(t) + W(t)$ for the case with churning, $\overline{n} = 10^6$ and mean peer holding time between 1 h and 30 s. Time is expressed in seconds.

7 The Impact of Peers Churning

In this section, we extend our model to consider the effect of peer churning. In stationary conditions, for sufficiently large values of n, the effect of churning on the global population is mainly negligible because the flow of peers leaving the system is compensated by the flow of new peers joining. Churning may, however, have an impact on the effectiveness of chunk distribution because seeds may abruptly leave the system, thus becoming unavailable to distribute the chunk. This can be easily reflected in the fluid model adding the terms describing the peer departure and arrival flows.

For example, consider the access-bandwidth-limited case and the RC strategy. Let μ_{dep} be the rate at which each participating peer leaves the system (this rate is inversely proportional to the peer holding time), and let λ_{arr} be the aggregate rate at which new peers join the system. In stationary conditions, the average number of peers in the system $\overline{n}$ can be obtained from the relation: $\lambda_{arr} = \overline{n}\mu_{dep}$. The equations become:

$$\frac{dI(t)}{dt} = \gamma(t) - \mu_{dep} I(t), \tag{11}$$

$$\frac{dS(t)}{dt} = R(t) - \gamma(t) - \mu_{dep} S(t), \tag{12}$$

$$\frac{dL(t)}{dt} = [K_W(t) - 1]R(t) - \mu_{dep} L(t), \tag{13}$$

$$\frac{dW(t)}{dt} = -K_W(t)R(t) - \mu_{dep} W(t) + \lambda_{arr}, \tag{14}$$

$$W(t) + I(t) + S(t) + L(t) = \overline{n}. \tag{15}$$

Fig. 21 reports the results with $\overline{n} = 10^6$ users whose upload bandwidth $B_u = 1$ Mbit/s. The video rate is 500 Kbit/s and the chunk size $L = 125$ Kbit. The holding times of peers in the systems vary from 30 s up to infinity. As already shown by simulation in the recent work [31], the effect of peer churning on the chunk distribution time is almost negligible. This in light of: 1) the large number of peers involved in the distribution, 2) the high resilience of the mesh-based topologies, and 3) the intrinsic characteristic of the epidemic process employed to distribute chunks.

8 Conclusions

In this paper, we have evaluated the impact of different overlay topology design strategies on the performance of large-scale mesh-based P2P streaming systems through the definition of simple, yet accurate, fluid models. Two different scenarios have been analyzed: in the first one, chunk download throughput is limited by peer access bandwidth; in the second scenario, a congestion control mechanism or a signaling delay limits the chunk download throughput. We derived a number of simple guidelines for the design of the overlay topology. First, the notion of peer location should be exploited when the performance bottleneck is the latency between peers, i.e., RTT. Strategies for the construction of the overlay based on random choices are, instead, slightly preferable when performance is dominated by the limited bandwidth of peers at the access network, such as for xDSL users. Second, source placement should favor peers with high degree of connectivity, so as to speed up the initial phase of chunk distribution. Third, connectivity of peers with large available bandwidth should be carefully implemented to create a cluster of large-bandwidth peers in which the chunk is quickly distributed and made available to others.

Acknowledgments

This work has been supported by the European Commission through NAPA-WINE Project (Network-Aware P2P-TV Application over Wise Network), ICT Call 1 FP7-ICT-2007-1.

References

[1] PPStream, http://www.PPStream.com, 2010.
[2] PPLive, http://www.pplive.com, 2009.
[3] SOPCast, http://www.sopcast.com, 2010.
[4] TVAnts, http://www.tvants.com, 2009.
[5] G. Huang, "Keynote 1: Experiences with PPLive (Slides)," *Proc. Peer-to-Peer Streaming and IP-TV Workshop (P2P-TV), Joint Event with ACM SIGCOMM '07,* http://www.sigcomm.org/sigcomm2007/p2p-tv/, Aug. 2007.
[6] Babelgum, http://www.babelgum.com, 2010.
[7] Zattoo, http://www.zattoo.com, 2010.
[8] Tvunetworks, http://www.tvunetworks.com, 2010.
[9] X. Zhang, J. Liu, B. Li, and T.-S.P. Yum, "DONet/CoolStreaming: A Data-Driven Overlay Network for Peer-to-Peer Live Media Streaming," *Proc. IEEE INFOCOM '05,* Mar. 2005.
[10] R. Rejaie and A. Ortega, "PALS: Peer-to-Peer Adaptive Layered Streaming," *Proc. Int'l Workshop Network and Operating System Support for Digital Audio and Video (NOSSDAV),* June 2003.
[11] F. Pianese, J. Keller, and E.W. Biersack, "PULSE, a Flexible P2P Live Streaming System," *Proc. IEEE Global Internet Symp.,* Apr. 2006.
[12] N. Magharei and R. Rejaie, "PRIME: Peer-to-Peer Receiver-Driven Mesh-Based Streaming," *Proc. IEEE INFOCOM,* May 2007.
[13] D.A. Tran, K.A. Hua, and T. Do, "ZIGZAG: An Efficient Peer-to-Peer Scheme for Media Streaming," *Proc. IEEE INFOCOM,* Apr. 2003.
[14] H. Deshpande, M. Bawa, and H. Garcia-Molina, "Streaming Live Media over a Peer-to-Peer Network," technical report, Stanford Univ., Aug. 2001.
[15] Y. Chu, A. Ganjam, T.S.E. Ng, S.G. Rao, K. Sripanidkulchai, J. Zhan, and H. Zhang, "Early Experience with an Internet Broadcast System Based on Overlay Multicast," *Proc. USENIX Ann. Technical Conf.,* July 2004.
[16] S. Banerjee, B. Bhattacharjee, and C. Kommareddy, "Scalable Application Layer Multicast," *Proc. ACM SIGCOMM,* 2002.
[17] Y.-H. Chu, S.G. Rao, S. Seshan, and H. Zhang, "Enabling Conferencing Applications on the Internet Using an Overlay Multicast Architecture," *Proc. ACM SIGCOMM,* Aug. 2001.
[18] B. Cohen, "Incentives Build Robustness in BitTorrent," *Proc. Workshop Economics of Peer-to-Peer Systems,* June 2003.

[19] D. Kostic, A. Rodriguez, J. Albrecht, and A. Vahdat, "Bullet: High Bandwidth Data Dissemination Using an Overlay Mesh," *Proc. ACM Symp. Operating Systems Principles (SOSP '03),* Oct. 2003.
[20] R. Kumar, Y. Liu, and K.W. Ross, "Stochastic Fluid Theory for P2P Streaming Systems," *Proc. IEEE INFOCOM,* May 2007.
[21] L. Massoulie, A. Twigg, C. Gkantsidis, and P. Rodriguez, "Randomized Decentralized Broadcasting Algorithms," *Proc. IEEE INFOCOM,* May 2007.
[22] S. Sanghavi, B. Hajek, and L. Massoulie, "Gossiping with Multiple Messages," *Proc. IEEE INFOCOM,* May 2007.
[23] T. Bonald, L. Massoulie, F. Mathieu, D. Perino, and A. Twigg, "Epidemic Live Streaming: Optimal Performance Trade-Offs," *Proc. ACM SIGMETRICS,* June 2008.
[24] D. Ren, Y.T.H. Li, and S.H.G. Chan, "On Reducing Mesh Delay for Peer-to-Peer Live Streaming," *Proc. IEEE INFOCOM,* Apr. 2008.
[25] M.K.H. Yeung and Y.-K. Kwok, "Game-Theoretic Scalable Peer-to-Peer Media Streaming," *Proc. 22nd IEEE Int'l Parallel and Distributed Processing Symp. (IPDPS '08),* Apr. 2008.
[26] M.K.H. Yeung and Y.-K. Kwok, "Game Theoretic Peer Selection for Resilient Peer-to-Peer Media Streaming Systems," *Proc. 28th Int'l Conf. Distributed Computing Systems (ICDCS '08),* June 2008.
[27] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, "Topologically-Aware Overlay Construction and Server Selection," *Proc. IEEE INFOCOM, 2002.*
[28] L. Tang and M. Crovella, "Virtual Landmarks for the Internet," *Proc. ACM Internet Measurement Conf. (IMC), 2003.*
[29] F. Dabek, R. Cox, F. Kaashoek, and R. Morris, "Vivaldi: A Decentralized Network Coordinate System," *Proc. ACM SIGCOM, 2004.*
[30] http://www.napa-wine.eu/cgi-bin/twiki/view/Public/P2PTVSim, 2010.
[31] F. Picconi and L. Massoulie, "Is There a Future for Mesh-Based Live Video Streaming?" *Proc. IEEE Int'l Conf. Peer-to-Peer Computing (P2P '08),* Sept. 2008.
[32] L. Massoulie, A. Twigg, C. Gkantsidis, and P.R. Rodriguez, "Randomized Decentralized Broadcasting Algorithms," *Proc. IEEE INFOCOM,* May 2007.
[33] B. Bollobás, *Random Graphs,* second ed. Cambridge Univ. Press, 2001.
[34] M.E.J. Newman, C. Moore, and D.J. Watts, "Mean-Field Solution of the Small-World Network Model," *Physical Rev. Letters,* vol. 84, pp. 3201-3204, Apr. 2000.

Ana Paula Couto da Silva received the BSc degree in computer science from the Federal University of Juiz de Fora, Brazil, in 1999, and the MSc and DSc degrees in computer and system engineering both from the Federal University of Rio de Janeiro, Brazil, in 2001 and 2006, respectively. She was with IRISA-Rennes, France, in 2005 and 2007, and with the Politecnico di Torino, Italy, in 2008. She is currently with the Computer Science Department at the Federal University of Juiz de Fora, Brazil, as assistant professor, since February 2009. Her areas of interest are in the field of modeling and analysis of computer systems, reliability analysis, and P2P application analysis.

Emilio Leonardi received the DrIng degree in electronics engineering in 1991 and the PhD degree in telecommunications engineering in 1995, both from the Politecnico di Torino, Italy, where he is currently an associate professor at the Dipartimento di Elettronica. In 1995, he visited the Computer Science Department of the University of California, Los Angeles (UCLA); in Summer 1999, he joined the High Speed Networks Research Group at Bell Laboratories/ Lucent Technologies, Holmdel, New Jersey; in Summer 2001, the Electrical Engineering Department of the Stanford University, and finally, in Summer 2003, the IP Group at Sprint, Advanced Technologies Laboratories, Burlingame, California. He is the scientific coordinator of the European Seventh FP STREP Project "NAPA-WINE" on P2P streaming applications, involving 11 European research institutions, operators, and manufacturers. His research interests are in the field of performance evaluation of wireless networks, P2P systems, and packet switching. He is a senior member of the IEEE.

Marco Mellia (S'08) received the PhD degree in telecommunications engineering in 2001 from the Politecnico di Torino, Italy. During 1999, he was with the Computer Science Department at Carnegie Mellon University, Pittsburgh, Pennsylvania, as a visiting scholar. During 2002, he visited the Sprint Advanced Technology Laboratories Burlingame, California. Since April 2001, he is with the Electronics Department of Politecnico di Torino as an assistant professor. He has coauthored more than 140 papers published in international journals and presented in leading international conferences, all of them in the area of telecommunication networks; also, he has participated in the program committees of several conferences including the IEEE Infocom and ACM Sigcomm. His research interests are in the fields of traffic measurement and modeling, P2P application analysis, and energy-aware network design. He is a senior member of the IEEE.

Michela Meo received the laurea degree in electronics engineering in 1993, and the PhD degree in electronic and telecommunications engineering in 1997, both from the Politecnico di Torino, Italy. Since November 1999, she is an assistant professor at the Politecnico di Torino. She has coauthored more than 120 papers, about 40 of which are in international journals. She has edited six special issues of international journals, including the *ACM Monet*, *Performance Evaluation*, and the *Journal of Computer Networks*. She was program cochair of two editions of the ACM MSWiM, general chair of another edition of the ACM MSWiM, program cochair of the IEEE QoS-IP, the IEEE MoVeNet 2007, the IEEE ISCC 2009, and was in the program committee of about 50 international conferences, including Sigmetrics, Infocom, ICC and Globecom. Her research interests are in the field of performance evaluation and modeling, traffic classification and characterization, P2P, and green networking. She is a member of the IEEE.