

# Converging an Overlay Network to a Gradient Topology

Håkan Terelius\*, Guodong Shi\*, Jim Dowling<sup>†</sup>, Amir Payberah<sup>†</sup>, Ather Gattami\* and Karl Henrik Johansson\*

\*KTH - Royal Institute of Technology, {hakante,guodongs,gattami,kallej}@kth.se

<sup>†</sup>Swedish Institute of Computer Science (SICS), {jdowling,amir}@sics.se

**Abstract**—In this paper, we investigate the topology convergence problem for the gossip-based Gradient overlay network. In an overlay network where each node has a local utility value, a Gradient overlay network is characterized by the properties that each node has a set of neighbors containing higher utility values, such that paths of increasing utilities emerge in the network topology. The Gradient overlay network is built using gossiping and a preference function that samples from nodes using a uniform random peer sampling service. We analyze it using tools from matrix analysis, and we prove both the necessary and sufficient conditions for convergence to a complete gradient structure, as well as estimating the convergence time. Finally, we show in simulations the potential of the Gradient overlay, by building a more efficient live-streaming peer-to-peer (P2P) system than one built using uniform random peer sampling.

**Keywords:** Overlay networks; topology convergence; gossiping; gradient topology

## I. INTRODUCTION

Recent years have witnessed growing interest in using randomized gossiping algorithms to build distributed systems, in particular in the areas of overlay networks, sensor networks and cloud computing storage services [1], [2]. Gossip-based, or pair-wise exchange, algorithms have primarily been used to implement aggregation algorithms, information dissemination, peer sampling (the uniform random sampling of a node from the set of all nodes in a P2P system), and to construct overlay network topologies. Much of the existing analysis of gossip-based algorithms has focused on the convergence properties of aggregation algorithms and peer sampling services, for both fixed topologies [3] and regular graphs [4], [5].

However, research in gossiping has also focused on using the Preferential Connectivity Model [6] to construct overlay network topologies, where nodes initially connected in a random graph use a preferential connection function to break the symmetry of the random graph and build a topology that contains useful global information. Barabasi first described how a preferential attachment function in a growing network can build a scale-free network topology from a random graph [7]. In particular, he showed how the power-law distribution of links in the the World Wide Web can emerge when arriving nodes preferentially attach to existing nodes with higher edge degree. Information about the structure of the Web's topology is currently used, among other things, to build more efficient search algorithms. Barabasi's preferential attachment functions are based on global state (the in-degree of nodes). However, in overlay networks, nodes have only a relatively small

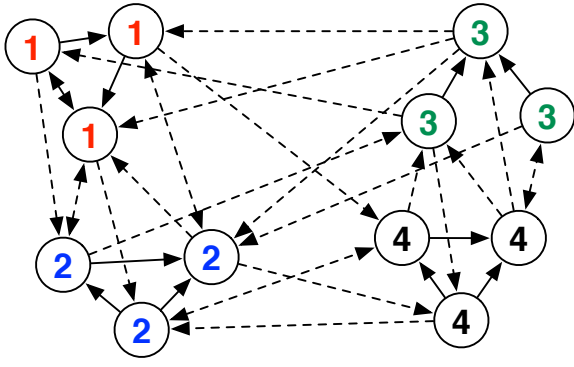
partial view of the system, so preference functions are based only on local state and the state of the node's neighbors. Examples of existing overlay networks that construct their topologies using gossiping and preference functions include Spotify, that preferentially connects nodes with similar music play-lists [8], Sepidar, that preferentially connects P2P live-streaming nodes with similar upload bandwidth capacity [9], and T-Man, a framework that provides a generic preference function for building such overlays [10].

To the best of our knowledge, there has been no analysis of the convergence properties of such information-carrying gossip-generated topologies built using preference functions. These systems, however, do not require the growth of a network to construct a new topology, as systems are constantly updated using a peer sampling service. In this paper, we introduce an analysis of the convergence properties for the Gradient overlay network. The Gradient topology belongs to this class of gossip-generated overlay networks that are built from a random overlay by symmetry breaking using a preference function. Formally, a Gradient topology is defined as an overlay network where, for any two nodes  $p$  and  $q$  that have local utility values  $U(p)$  and  $U(q)$ , if  $U(p) \geq U(q)$  then  $dist(p, r) \leq dist(q, r)$ , where  $r$  is a (or the) node with highest utility in the system and  $dist(x, y)$  is the shortest path length between nodes  $x$  and  $y$  [11]. In the Gradient overlay, nodes have two preference functions that build two sets of neighbors: a similar view and a random view. For the similar view, nodes prefer neighbors with closer but slightly higher utility values, while for the random view, nodes are selected with uniform probability. Together these preference functions build a topology where gradient paths of increasing utilities emerge in the system [12], see figure 1.

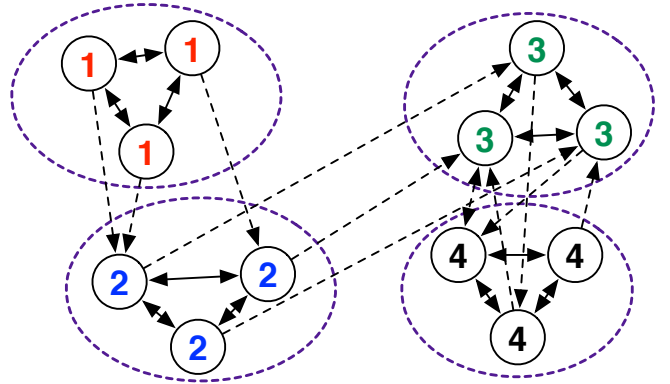
Our analysis of the Gradient overlay, involves proving that the preference functions cause the system topology to converge to a gradient structure. We also establish bounds on the worst-case convergence rate for a given initial graph. Finally, we show in simulations how the Gradient structure is used to build a more efficient live-streaming system than one built using uniform random peer sampling.

## II. PROBLEM SETUP

Consider a network whose topology can be described by a directed graph  $\mathcal{G}(\mathcal{N}, \mathcal{E})$ . Each node in the network is represented by a vertex in the graph, and each link is represented by a directed edge (see figure 1(a)). We denote the vertex set by  $\mathcal{N} = \{1, \dots, N\}$ , where each node  $i$  is given a utility value  $U(i)$ .



(a) The initial graph.



(b) The graph after converging to a gradient topology.

Fig. 1. The network is described as a directed graph. The nodes are labeled with their respective utility value, and the edges from the similar neighbor set are shown. Solid edges are used between nodes with equal utility value, and dashed edges between nodes with different utility value.

The neighbor set  $N_i(t)$  of node  $i$  at time  $t$  consists of two parts, the similar set  $N_i^s(t)$  and the random set  $N_i^r(t)$ . Nodes in the similar set are supposed to be the neighbors whose utility values are close to  $U(i)$ , while nodes in the random set are a random sample of the nodes in the network.

Each node  $i$  defines a *preference function*  $>_i$  over its neighbors, where node  $i$  is said to prefer node  $a$  over node  $b$  ( $a >_i b$ ) if

$$U(a) \geq U(i) > U(b) \quad \text{or if} \\ |U(a) - U(i)| < |U(b) - U(i)|$$

when  $U(a), U(b) > U(i)$  or  $U(a), U(b) < U(i)$ . Further,  $\min N_i^s$  denotes node  $i$ 's least preferred neighbor in its similar neighbor set.

For any given initial graph, each node  $i$  at each time  $t$  updates its neighbor set  $N_i(t)$  independently of the other nodes according to Algorithm 1.

---

**Algorithm 1:** Topology Dynamics

---

```

for  $t = 1, 2, 3, \dots$  do
  // Choose a random node  $j \in \mathcal{N}$  with
  uniform probability  $p_t$ ,  $0 < Np_t < 1$ 
   $N_i^r(t) = \{j\}$ 
  if  $N_i^r(t) \neq \emptyset$  then
    if  $j \notin N_i^s(t-1)$  and  $j >_i \min N_i^s(t-1)$  then
       $N_i^s(t) = N_i^s(t-1) \cup \{j\} \setminus \{\min N_i^s(t-1)\}$ 
    end
  end
end
end

```

---

In summary, if the random node is preferred over the least preferred node in the similar set, then those two nodes are exchanged. Notice that the probabilities  $p_t$  are time varying, and that the random neighbor set  $N_i^r(t)$  is empty with probability  $1 - Np_t$ .

*Remark 2.1:* The node degree  $d_i(t) = |N_i^s(t)| = d_i$  stays constant throughout the algorithm.

This paper considers the problem of whether the system topology will converge to a gradient structure with the proposed algorithm, and the convergence rate for a given initial graph.

Let  $\Lambda_i$  denote the set of optimal similar neighbor sets for node  $i$ , i.e.,  $\forall \hat{N} \in \Lambda_i$  there are no  $j \in \hat{N}$  and  $k \in \mathcal{N} \setminus \hat{N}$  such that  $k >_i j$ .

For every node  $i \in \mathcal{N}$ , we define

$$X_t^{(i)} = d_i - \max_{\hat{N} \in \Lambda_i} |N_i^s(t) \cap \hat{N}|$$

Thus,  $X_t^{(i)}$  counts the number non-optimal neighbors in  $i$ 's similar neighbor set. Notice that  $X_t^{(i)}$  is monotonically decreasing since an optimal neighbor will never be removed from the similar neighbor set  $N_i^s(t)$ .

Let  $\mathcal{G}(t)$  be the graphs generated by Algorithm 1. Then we give the definition of gradient convergence as follows (see also figure 1).

*Definition 2.1:*  $\mathcal{G}(t)$  is said to converge to a gradient topology if  $\lim_{t \rightarrow \infty} X_t^{(i)} = 0$  for all  $i \in \mathcal{N}$ .

### III. CONVERGENCE ANALYSIS

In this section, we propose a convergence analysis of Algorithm 1 to a gradient topology. Since each node updates its neighbor set independently, the analysis on respective  $X_t^{(i)}$  can be carried out separately. Therefore, to simplify the notations, we let  $X_t$  represents  $X_t^{(i)}$ ,  $i \in \mathcal{N}$ , in the following discussion.

Denote the maximum degree  $D = \max_i \{d_i\}$ , then it is not hard to see that  $X_0 = D$  is the worst initial condition. Furthermore,  $X_t$  decreases whenever the random node is a new optimal neighbor, and the probability of this event occurring is minimal when the optimal solution is unique, and then equal to

$$\mathbf{P}[X_{t+1} = k - 1 \mid X_t = k] = kp_t, \quad (1)$$

where  $k$  are the number of non-optimal neighbors.

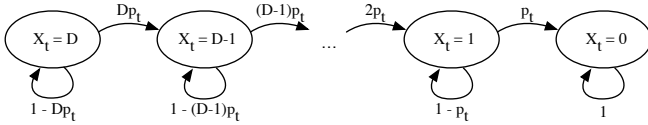
### A. Almost Sure Convergence

We propose a both necessary and sufficient condition on the probabilities  $p_t$  for the convergence of Algorithm 1.

**Theorem 3.1:** The graph generated by Algorithm 1 converges to a gradient topology ( $X_t = 0$ ) with probability 1 if and only if

$$\lim_{T \rightarrow \infty} \prod_{t=0}^T (1 - p_t) = 0. \quad (2)$$

Before proving Theorem 3.1, let us take a closer look at Algorithm 1, and notice especially that the stochastic process (1) for  $X_t$  has the Markov property, hence we can describe it as a Markov chain.



Let  $\pi(t)$  denote the (row vector) probability distribution for the states  $X_t$ , i.e.,

$$\pi_i(t) = \mathbf{P}[X_t = i]. \quad (3)$$

The evolution of  $\pi(t)$  can be written in matrix form as

$$\pi(t+1) = \pi(t)P_t, \quad (4)$$

where  $P_t$  is the transition matrix at time  $t$ ,

$$P_t = \begin{bmatrix} 1-Dp_t & Dp_t & 0 & \dots & 0 & 0 \\ 0 & 1-(D-1)p_t & (D-1)p_t & \dots & 0 & 0 \\ 0 & 0 & 1-(D-2)p_t & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \dots & 1-p_t & p_t \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Since  $P_t$  is a triangular matrix, the eigenvalues are given by the diagonal elements, i.e., the eigenvalues of  $P_t$  are  $\lambda_i(t) = 1 - (D - i)p_t$ ,  $i = 0, \dots, D$ . Notice that  $\lambda_D(t) = 1$ , and all other eigenvalues are strictly less than one. Furthermore, all eigenvalues are distinct, hence the eigenvectors form a basis for  $\mathbb{R}^D$ . In the following lemma, we characterize the eigenvectors.

**Lemma 3.1:** The eigenvector  $\xi^i(t)$  corresponding to eigenvalue  $\lambda_i(t)$  is independent of  $p_t \neq 0$ ,  $i = 0, \dots, D$ .

*Proof:* The (left-)eigenvectors of  $P_t$  satisfy  $\lambda_i(t)\xi^i(t) = \xi^i(t)P_t$ . Let  $\xi_j^i(t)$  denote the  $j$ :th component of  $\xi^i(t)$ , then

$$\begin{cases} (1 - (D - i)p_t)\xi_0^i(t) = (1 - Dp_t)\xi_0^i(t) \\ (1 - (D - i)p_t)\xi_j^i(t) = (1 - (D - j)p_t)\xi_j^i(t) + \\ (D - j + 1)p_t\xi_{j-1}^i(t) \quad j = 1, \dots, D \end{cases} \quad (5)$$

$$\Downarrow$$

$$\begin{cases} i\xi_0^i(t) = 0 \\ (i - j)\xi_j^i(t) = (D - j + 1)\xi_{j-1}^i(t) \quad j = 1, \dots, D \end{cases}$$

$$\Downarrow$$

$$\begin{cases} \xi_j^i(t) = 0 & \text{if } j < i \\ \xi_j^i(t) \frac{i-j}{D-j+1} = \xi_{j-1}^i(t) & \text{if } j > i \end{cases}$$

while  $\xi_0^i(t)$  can be chosen as an arbitrary non-zero value.  $\blacksquare$

Lemma 3.1 implies especially that all  $P_t$  are simultaneously diagonalizable, hence we can drop the parameter  $t$  from  $\xi^i$ .

Let us now return to the initial probability distribution  $\pi(0)$ , and let us express it in the eigenvector basis as

$$\pi(0) = \sum_{i=0}^D \alpha_i \xi^i, \quad (6)$$

for some real numbers  $\alpha_i$ .

**Lemma 3.2:**  $\alpha_D \xi^D = \mathbf{e}_D$ , where  $\mathbf{e}_i$  is the Cartesian unit vector  $[0, \dots, 0, 1, 0, \dots, 0]^T$  with 1 in position  $i$ .

*Proof:* Let us consider  $\xi^i \mathbf{1}$  for  $i = 0, \dots, D - 1$ . By equation (5),

$$\xi^i \mathbf{1} = \sum_{j=0}^D \xi_j^i = \sum_{j=i}^D \xi_j^i = \sum_{j=0}^{D-i} \xi_{i+j}^i$$

We will show by induction that

$$\sum_{j=0}^k \xi_{i+j}^i = \frac{D - i - k}{D - i} \xi_{i+k}^i. \quad (7)$$

The case when  $k = 0$  is clearly true, thus, assume (7) holds for  $k$  and consider  $k + 1$ ,

$$\begin{aligned} \sum_{j=0}^{k+1} \xi_{i+j}^i &= \sum_{j=0}^k \xi_{i+j}^i + \xi_{i+k+1}^i = \frac{D - i - k}{D - i} \xi_{i+k}^i + \xi_{i+k+1}^i \\ &= \frac{D - i - k}{D - i} \frac{D - i - k - (k + 1)}{D - i - k} \xi_{i+k+1}^i + \xi_{i+k+1}^i \\ &= \frac{D - i - (k + 1)}{D - i} \xi_{i+k+1}^i \end{aligned}$$

Using (7) implies that  $\xi^i \mathbf{1} = 0$ ,  $i = 0, \dots, D - 1$ , and further,  $\pi(0) \mathbf{1} = \alpha_D \xi^D \mathbf{1}$ . Since  $\pi(0)$  is a probability distribution, we know that  $\pi(0) \mathbf{1} = 1$ , but (5) tells us that only the last component of  $\xi^D$  is non-zero, hence the lemma follows.  $\blacksquare$

We are now ready to prove the main theorem.

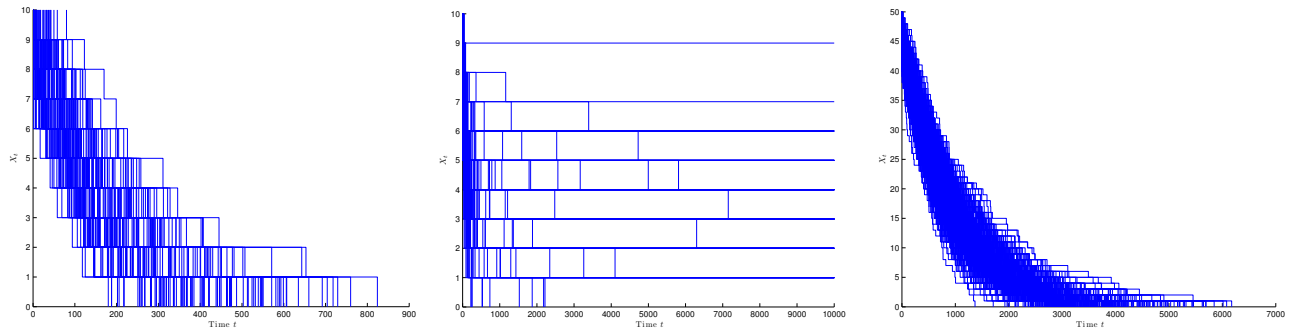
*Proof:* (Theorem 3.1) The convergence condition is equivalent to  $\lim_{T \rightarrow \infty} \pi(T) = \mathbf{e}_D$ . Using (4) and (6) gives us

$$\begin{aligned} \pi(T) &= \pi(0) \prod_{t=0}^{T-1} P_t = \sum_{i=0}^D \alpha_i \xi^i \prod_{t=0}^{T-1} P_t = \\ &= \sum_{i=0}^D \alpha_i \xi^i \prod_{t=0}^{T-1} \lambda_i(t) = \sum_{i=0}^{D-1} \alpha_i \xi^i \prod_{t=0}^{T-1} \lambda_i(t) + \mathbf{e}_D \quad (8) \end{aligned}$$

Consider the limit  $\lim_{T \rightarrow \infty} \pi(T)$ ,

$$\begin{aligned} \lim_{T \rightarrow \infty} |\pi(T) - \mathbf{e}_D| &= \lim_{T \rightarrow \infty} \left| \sum_{i=0}^{D-1} \alpha_i \xi^i \prod_{t=0}^{T-1} \lambda_i(t) \right| \leq \\ &= \sum_{i=0}^{D-1} |\alpha_i \xi^i| \cdot \lim_{T \rightarrow \infty} \prod_{t=0}^{T-1} (1 - p_t). \end{aligned}$$

Clearly, this converges to zero if  $\lim_{T \rightarrow \infty} \prod_{t=0}^T (1 - p_t) = 0$ .



(a) Convergence in a network with 100 nodes,  $d_i = 10$ , and constant probability  $Np_t = \frac{1}{2}$  (b) Convergence is missing in a network with 100 nodes,  $d_i = 10$ , and decaying probability  $Np_t = \frac{1}{(1+t/100)^2}$  (c) Convergence in a network with 500 nodes,  $d_i = 50$ , and constant probability  $Np_t = \frac{1}{2}$

Fig. 2. Convergence rate simulations. The neighbor set measurement  $X_t$ , for each node in the network, is shown as a function of the iteration number  $t$ .

Also, the set of initial probability distributions spawns  $\mathbb{R}^D$ , thus, there exists an initial probability distribution  $\pi(0)$  such that  $\alpha_{D-1} \neq 0$ . Assume  $\lim_{T \rightarrow \infty} \prod_{t=0}^T (1 - p_t) = c > 0$  (the limit exists, since it is a monotone bounded sequence), then

$$\lim_{T \rightarrow \infty} |\pi(T) - \mathbf{e}_D| = \left| \sum_{i=0}^{D-2} \alpha_i \xi^i \left( \lim_{T \rightarrow \infty} \prod_{t=0}^{T-1} \lambda_i(t) \right) + c \alpha_{D-1} \xi^{D-1} \right| \quad (9)$$

Since the eigenvectors are linearly independent, the RHS of (9) is non-zero. Thus, we have proved the theorem. ■

*Corollary 3.1:* The graph generated by Algorithm 1 converges to a gradient topology with probability 1 if and only if

$$\lim_{T \rightarrow \infty} \sum_{t=0}^T p_t = \infty. \quad (10)$$

*Proof:* This follows from Theorem 3.1, and the relation

$$\lim_{T \rightarrow \infty} \prod_{t=0}^T (1 - p_t) = 0 \Leftrightarrow \lim_{T \rightarrow \infty} \sum_{t=0}^T p_t = \infty$$

for  $0 < p_t < 1$ . ■

### B. Convergence Rate Estimation

In this subsection, we investigate the convergence rate of  $X_t$ , with a constant sampling probability  $p_t = p$ . Define

$$T_i = \inf_t [X_t = 0 \mid X_0 = i]$$

as the first time when  $X_t$  reaches 0, when starting with  $X_0 = i$ . Further, let  $M_i = \mathbf{E}[T_i]$  denote the expected time of convergence. Clearly  $M_0 = 0$ , and for  $i = 1, \dots, D$  we have

$$\begin{aligned} M_i &= 1 + \mathbf{P}[X_{t+1} = i - 1 \mid X_t = i] \cdot M_{i-1} \\ &\quad + \mathbf{P}[X_{t+1} = i \mid X_t = i] \cdot M_i \\ &= 1 + ipM_{i-1} + (1 - ip)M_i \\ &\Rightarrow \\ M_i &= \frac{1 + ipM_{i-1}}{ip} = \frac{1}{ip} + M_{i-1} \end{aligned}$$

Continuing by induction yields

$$M_i = \frac{1}{p} \sum_{n=1}^i \frac{1}{n}.$$

The worst initial case is when  $X_0 = D$ , where the expected convergence time is

$$M_D = \frac{1}{p} \sum_{n=1}^D \frac{1}{n} \leq \frac{1 + \ln(D)}{p}. \quad (11)$$

*Remark 3.1:* Notice that  $M_D$  is the expected time for an individual node to converge, and not the expected time for all the nodes in the network to converge to a gradient topology.

## IV. CONVERGENCE SIMULATION

In this section, we examine the convergence of Algorithm 1 with numerical examples. In the first two simulation (figure 2(a) and figure 2(b)) the degree of each node is  $d_i = 10$  and the total number of nodes in the network is  $N = 100$ . For the third simulation (figure 2(c)) the degree is  $d_i = 50$ , and the total number of nodes in the network is  $N = 500$ .

The similar view  $N_i^s(0)$  is initialized with  $d_i$  nodes uniformly chosen among all nodes in the network. In the first and third simulation the sampling probability  $p_t$  is held at a constant value of  $\frac{1}{2N}$ . Hence, for each node, and at each iteration of the algorithm, the random view is empty with probability  $\frac{1}{2}$ . Theorem 3.1 guarantees the convergence of the algorithm for these examples, which is also confirmed by the simulations. These two simulations should also be compared to the expected convergence rate given by equation (11), 566 and 4479 iterations respectively.

In the second simulation (figure 2(b)), we also analyze a decaying probability  $p_t = \frac{1}{N(1+t/100)^2}$ . Notice that  $\sum_{t=0}^{\infty} Np_t < 101$ , hence, by Corollary 3.1, there is a positive probability that the algorithm does not converge to a gradient topology. This is also confirmed by the simulation, in which the gradient topology is missing.

## V. LIVE-STREAMING USING THE GRADIENT - EXPERIMENTS

Here, we evaluate the effect of sampling nodes from the Gradient overlay compared to a random overlay when building a P2P live-streaming application called GLive. GLive is based on nodes cooperating to share a media stream supplied by a source node. GLive uses an approximate auction algorithm to match nodes that are willing and able to share the stream with one another. GLive extends our previous work on tree-based live-streaming, gradienTv [13] and Sepidar [9], to mesh-based live-streaming.

Nodes want to establish connections to other nodes that are as close as possible to the source. They bid for connections to the best neighbors using their own upload bandwidth, and nodes share their bounded number of connections with nodes who bid the highest (contribute the most upload bandwidth). Auctions are continuous and restarted on failures or free-riding. The desired affect of our auction algorithm is that the source will upload to nodes who contribute the most upload bandwidth, who will, in turn, upload to nodes who contribute the next highest amount of bandwidth, and so on until the topology is fully constructed. More details on our approximate assignment algorithm can be found in [9].

One of the main problems with the lack of global information about nodes' upload bandwidths is that it affects the rate of convergence of the auction algorithm. Nodes would ideally like to bid for connections to other nodes who they can afford to connect to, rather than win a connection to a better node and later be removed because a better bid was received. The traditional way to discover nodes (to bid on) is using a uniform random peer-sampling service [5]. Instead, we use the Gradient overlay to sample nodes, where a node's utility value is the upload bandwidth it contributes to the system. As such, the Gradient should provide other nodes with references to nodes who have well-matched upload bandwidths. In [9], we showed that using the Gradient overlay reduced the rate of parent switching for tree-based live-streaming by 20% compared to random peer sampling. Here, we show for GLive the effect of sampling neighbors using random peer sampling (GLive/Random) versus sampling from the Gradient overlay (GLive/Gradient).

We implemented GLive using Kompics' discrete event simulator that provides different bandwidth, latency and churn models. In our experimental setup, we set the streaming rate to  $512Kbps$ , which is divided into blocks of  $16Kb$ . Nodes start playing the media after buffering it for 5 seconds. The size of similar-view in GLive is 15 nodes. In the auction algorithm, nodes have 8 download connections. To model upload bandwidth, we assume that each upload connection has available bandwidth of  $64Kbps$  and that the number of upload connections for nodes is set to  $2i$ , where  $i$  is picked randomly from the range 1 to 10. This means that nodes have upload bandwidth between  $128Kbps$  and  $1.25Mbps$ . As the average upload bandwidth of  $704Kbps$  is not much higher than the streaming rate of  $512Kbps$ , nodes have to find good matches as parents to achieve good

streaming performance. The media source is a single node with 40 upload connections, providing five times the upload bandwidth of the stream rate. We assume 11 utility levels, such that nodes contributing the same amount of upload bandwidth are located at the same utility level. Latencies between nodes are modeled using a latency map based on the King data-set [14]. We assume the size of sliding window for downloading is 32 blocks, such that the first 16 blocks are considered as the in-order set and the next 16 blocks are the blocks in the rare set. A block is chosen for download from the in-order set with 90% probability, and from the rare set with 10% probability. In the experiments, we measure the following metrics:

- 1) *Playback continuity*: the percentage of blocks that a node received before their playback time. We consider the case where nodes have a playback continuity of greater than 99%;
- 2) *Playback latency*: the difference in seconds between the playback point of a node and the playback point at the media source.

We compare the playback continuity and playback latency of GLive/Gradient and GLive/Random in the following scenarios:

- 1) *Churn*: 500 nodes join the system following a Poisson distribution with an average inter-arrival time of 100 milliseconds, and then until the end of the simulations nodes join and fail continuously following the same distribution with an average inter-arrival time of 1000 milliseconds;
- 2) *Flash crowd*: first, 100 nodes join the system following a Poisson distribution with an average inter-arrival time of 100 milliseconds. Then, 1000 nodes join following the same distribution with a shortened average inter-arrival time of 10 milliseconds;
- 3) *Catastrophic failure*: 1000 nodes join the system following a Poisson distribution with an average inter-arrival time of 100 milliseconds. Then, 500 existing nodes fail following a Poisson distribution with an average inter-arrival time 10 milliseconds;

Figure 3 shows the percentage of the nodes that have playback continuity of at least 99%. We see that all the nodes in GLive/Gradient receive at least 99% of all the blocks very quickly in all scenarios, while it takes slightly more time for GLive/Random. That is because nodes in GLive/Random randomly sample nodes to run the auction algorithm against, while GLive/Gradient runs the auction algorithm against nodes that contribute similar amounts of upload bandwidth. Random sampling takes longer time to find good matches for delivering the stream. One point to note is that the 5 seconds of buffering cause the spike in playback continuity at the start, which then drops off as nodes are joining the system. To summarize, using the Gradient overlay instead of random sampling produces better performance when the system is undergoing large changes - such as large numbers of nodes joining or failing over a short period of time. Figure 4 shows the playback latency of the systems in the different scenarios.

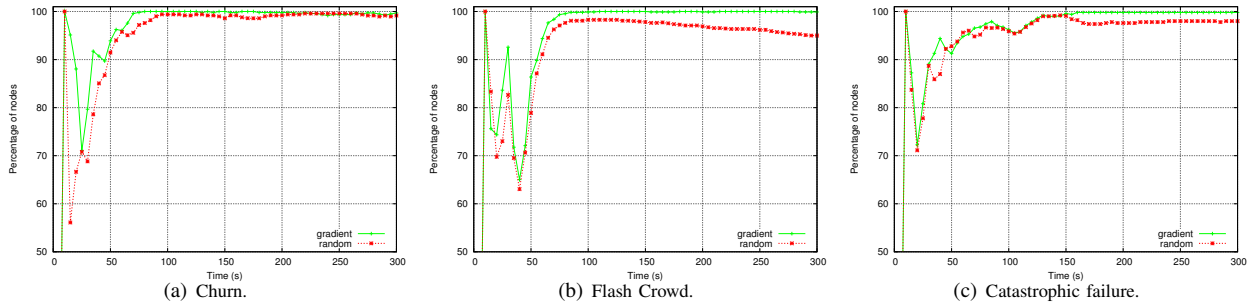


Fig. 3. Playback continuity of the systems in different scenarios.

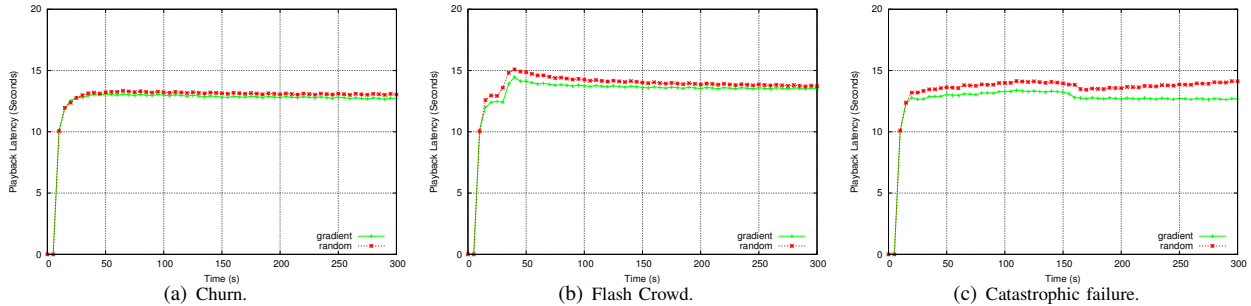


Fig. 4. Playback latency of the Gradient versus Random sampling in different scenarios.

As we can see, although there is only a small difference between the systems, GLive/Gradient consistently maintains relatively shorter playback latency than GLive/Random for all experiments. The playback latency includes both the 5 seconds buffering time and the time required to pull the blocks over the live-streaming overlay constructed using the auction algorithm.

## VI. CONCLUSIONS

In this paper, we introduced the topology convergence problem for the gossip-generated Gradient overlay network. We showed the necessary and sufficient conditions for convergence to a complete gradient structure, and we also characterized the expected convergence time. Our experiments show the potential advantages of topologies built using preference functions. We showed how nodes can use implicit information captured in the Gradient topology to more efficiently find suitable neighbors compared to random sampling. As such, our work on proving convergence properties of the Gradient topology should have significance for other future information-carrying topologies. In future work, we will examine modifications to the topology construction algorithm that improve convergence time, as well as further applications of the topology in building P2P applications.

## REFERENCES

- [1] A.-M. Kermarrec and M. van Steen, "Gossiping in distributed systems," *Operating Systems Review*, vol. 41, no. 5, pp. 2–7, 2007.
- [2] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Transactions on Networking*, vol. 14, pp. 2508–2530, June 2006.
- [3] A. Olshevsky and J. N. Tsitsiklis, "Convergence rates in distributed consensus and averaging," in *IEEE Conference on Decision and Control*, 2006, pp. 3387–3392.
- [4] J. Liu, B. D. O. Anderson, M. Cao, and A. S. Morse, "Analysis of accelerated gossip algorithms," in *CDC*, 2009, pp. 871–876.
- [5] M. Jelasity, S. Voulgaris, R. Guerraoui, A.-M. Kermarrec, and M. van Steen, "Gossip-based peer sampling," *ACM Trans. Comput. Syst.*, vol. 25, no. 3, 2007.
- [6] M. Mihail, C. Papadimitriou, and A. Saberi, "On certain connectivity properties of the internet topology," in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, ser. FOCS '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 28–. [Online]. Available: <http://portal.acm.org/citation.cfm?id=946243.946323>
- [7] A.-L. Barabási, "Linked: The new science of networks," *J. Artificial Societies and Social Simulation*, vol. 6, no. 2, 2003.
- [8] G. Kreitz and F. Niemela, "Spotify—Large Scale, Low Latency, P2P Music-on-Demand Streaming," in *IEEE International Conference on Peer-to-Peer Computing (P2P)*. IEEE, 2010, pp. 1–10.
- [9] A. H. Payberah, J. Dowling, F. Rahimian, and S. Haridi, "Sepidar: Incentivized market-based p2p live-streaming on the gradient overlay network," *International Symposium on Multimedia*, pp. 1–8, 2010.
- [10] M. Jelasity, A. Montresor, and Ö. Babaoglu, "T-man: Gossip-based fast overlay topology construction," *Computer Networks*, vol. 53, no. 13, pp. 2321–2339, 2009.
- [11] J. Sacha, "Exploiting heterogeneity in peer-to-peer systems using gradient topologies," Ph.D. dissertation, Trinity College Dublin, 2009.
- [12] J. Sacha, B. Biskupski, D. Dahlem, R. Cunningham, R. Meier, J. Dowling, and M. Haahr, "Decentralising a service-oriented architecture," *Peer-to-Peer Networking and Applications*, vol. 3, no. 4, pp. 323–350, 2010.
- [13] A. H. Payberah, J. Dowling, F. Rahimian, and S. Haridi, "Gradientv: Market-based p2p live media streaming on the gradient overlay," in *DAIS*, 2010, pp. 212–225.
- [14] K. P. Gummadi, S. Saroiu, and S. D. Gribble, "King: Estimating latency between arbitrary internet end hosts," in *SIGCOMM Internet Measurement Workshop*, 2002.