

# Peer counting and sampling in overlay networks based on random walks

A. J. Ganesh · A.-M. Kermarrec · E. Le Merrer ·  
L. Massoulié

Received: 28 July 2006 / Accepted: 19 March 2007 / Published online: 5 June 2007  
© Springer-Verlag 2007

**Abstract** In this article, we address the problem of counting the number of peers in a peer-to-peer system. This functionality has proven useful in the design of several peer-to-peer applications. However, it is delicate to achieve when nodes are organised in an overlay network, and each node has only a limited, local knowledge of the whole system. In this paper, we propose a generic technique, called the *Sample&Collide* method, to solve this problem. It relies on a sampling sub-routine which returns randomly chosen peers. Such a sampling sub-routine is of independent interest. It can be used for instance for neighbour selection by new nodes joining the system. We use a continuous time random walk to obtain such samples. The core of the method consists in gathering random samples until a target number of redundant samples are obtained. This method is inspired by the “birthday paradox” technique of Bawa et al. (Estimating aggregates on a peer-to-peer network, Technical Report, Department of Computer Science, Stanford University), upon which it improves by achieving a target variance with fewer samples. We analyse the complexity and accuracy of the proposed method. We illustrate in particular how *expansion* properties of the overlay affect its performance. We use simulations to

evaluate its performance in both static and dynamic environments with sudden changes in peer populations, and verify that it tracks varying system sizes accurately.

## 1 Introduction

Peer-to-peer systems have achieved fast and widespread adoption for both legal and illegal applications, ranging from file sharing (e.g. Kazaa or eDonkey) to VoIP (e.g. Skype). It is reasonable to expect novel applications to appear, and the scale of such systems to increase beyond millions of interacting peers.

A key feature of peer-to-peer systems is their *distributed* nature. Effectively, popular architectures organise the peers in an *overlay* network, typically layered over the Internet, and let peers communicate solely with their overlay neighbours. In such architectures, a peer’s knowledge of the system is limited to its collection of neighbours. These architectures have good scalability properties; more specifically, they do not suffer from central servers becoming performance bottlenecks, or single points of control and failure.

On the flip side, however, overlay architectures make it delicate to monitor system characteristics of interest that would be straightforward to observe in centralised systems. One such characteristic that will concern us in this paper is the system size, namely the number of peers. The need to perform peer counting arises in the following contexts. Recently proposed overlay maintenance protocols such as Viceroy [28] rely on approximate knowledge of the overlay size to incorporate a newly arrived peer in the system. Several gossip-based information dissemination protocols (e.g. [14,16]) rely on system size to determine the number of gossip targets per peer. We expect that peer counting could find other applications. For instance, in a Live Media

---

A. J. Ganesh (✉)  
Microsoft Research, Cambridge, UK  
e-mail: ajg@microsoft.com

A.-M. Kermarrec  
INRIA, Rennes, France  
e-mail: Anne-Marie.Kermarrec@irisa.fr

E. Le Merrer  
IRISA/INRIA and France Telecom R&D, Lannion, France  
e-mail: elemerre@irisa.fr

L. Massoulié  
Thomson Research, Paris, France  
e-mail: laurent.massoulie@thomson.net

streaming system such as that of [38], it may be of interest to measure the number of peers using a Broadband connection or a dialup connection, in order to decide whether new dialup users can be accepted without compromising performance.

For particular overlay architectures, specific overlay properties may be exploited to efficiently measure system size. In contrast, our aim here is to design measurement techniques that are generic, in that they are applicable to arbitrary overlay networks.

We propose one such technique, which we call the *Sample&Collide* (S&C) method. One of its building blocks is a sampling function, which aims to provide a requesting overlay node with another node chosen uniformly at random from the overlay. Previous proposals have relied on a discrete time random walk (DTRW) stopped after a large constant time; clearly, this yields samples biased towards high-degree nodes. We describe a sampling algorithm based on a continuous time random walk (CTRW), which yields unbiased samples. We characterise the sampling quality/complexity trade-off and show that it is critically affected by the expansion properties of the overlay graph.

The S&C method produces an estimate of the system size based on the number of uniform random samples it takes before a target number of redundant samples are obtained. This method is inspired by the ‘‘Inverted Birthday Paradox’’ method of [6]. We provide a detailed analysis of the accuracy and complexity of the S&C method, and show how it improves upon the original proposal of [6] by achieving a target accuracy with fewer samples.

We also provide experimental results, matching the theoretical expectations. While we do not study the impact of churn analytically, we evaluate it through simulations. These show that the proposed techniques are robust to both gradual and sudden changes in system size.

The paper is structured as follows: in the next section we survey related work. We present the S&C method and its theoretical analysis in Sect. 3. An evaluation by simulation is provided in Sect. 4. We conclude in Sect. 5.

## 2 Related work

Random-walk based peer sampling techniques such as [5, 13] are relevant to our work to some extent. However, we focus here on techniques for system size estimation and distinguish two classes. Techniques of the first type are tailored to a specific overlay architecture, while those of the second type are generic and applicable to any overlay.

### 2.1 Architecture-specific techniques

In structured peer to peer overlay networks, peers are assigned identifiers drawn uniformly at random. The approach taken

in [10] deduces the network size in distributed hash tables (DHT) by measuring the density of identifiers around a node initiating a size estimate. The communication cost for getting a relative error of order  $\epsilon$  is  $O(1/\epsilon^2)$  message exchanges, irrespective of the number of nodes  $N$ . A similar approach is also considered in [19, 23, 29, 36].

In [12], an estimate of the system size is constructed based on observations of node degrees, and relies on prior knowledge of a power law structure for the distribution of node degrees. A conceptually similar approach is described in [6]. It produces an estimate of system size based on node degree observations, assuming a specific topology (namely, the Erdős-Rényi random graph model). No error estimates are provided in these papers; the cost of the latter is  $O(\log N)$ . Jelasy and Preuß [21] estimate the network size by observing the renewal of contacts in peers’ views in a gossip based overlay.

Ntarmos et al. [34] propose a fully decentralized implementation of *hash sketches*, previously introduced by Flajolet and Martin, working over DHTs. The standard deviation is  $1.05/\sqrt{m}$  and  $0.78/\sqrt{m}$  (where  $m$  is the number of *bitmaps* used) for the two presented techniques.

Another approach involves building a spanning tree on top of the overlay and using it to estimate the system size [8, 25, 33] by aggregating estimates along the tree. The obtained estimates are then exact in the absence of failures, and the cost is  $\Theta(N)$ .

### 2.2 Generic techniques

Jelasy and Montresor [20] have considered the following gossip-based method. Initially one distinguished node sets a counter to 1 while all other nodes set their counter to 0. Nodes communicate asynchronously; when a pair of nodes communicates, they both reset their individual counters to the mean of the two previous values. In the long run, all counter values coincide with the reciprocal of the system size. This approach is suitable in stable environments. As all users eventually share the same size estimate, its cost is amortized over all nodes when they are all interested in obtaining such an estimate. A theoretical evaluation of the cost of such schemes can be found in [9]. The cost, evaluated in number of messages, is  $\tilde{O}(N^{1+2/d})$  for  $d$ -dimensional random geometric graphs.<sup>1</sup> It is  $O(N \log(N))$  for expander graphs.

Another generic approach [4, 15, 23, 36], sometimes referred to as probabilistic polling, consists of a querying node requesting all nodes to report their presence probabilistically, the probability of responding being a function of node characteristics, such as distance (in number of hops) from the initial requestor. This produces unbiased estimates. One

<sup>1</sup> See [9] for a definition of such graphs. Here, we write  $f(N) = \tilde{O}(g(N))$  when  $f(N) = O(g(N) \log(N)^\beta)$  for some  $\beta$ .

drawback of the method is that the initial querier is potentially faced with “ACK implosion”. The cost of this method scales linearly with system size.

Mosk-Aoyama and Shah [32] propose a gossip-based technique for distributed computation of symmetric functions, a special case of which is the number of nodes in the system. In that case, their algorithm corresponds to each node generating an independent, unit mean exponential random variable. Nodes exchange their values using gossip in order to compute the system-wide minimum of these random variables. The minimum is an exponential random variable with mean equal to the reciprocal of the number of nodes, and thus provides a way to estimate this number. Repeating the procedure with multiple independent samples can reduce the estimation error. The spreading time of the gossip is determined by the expansion properties of the overlay, which also play an important role in the algorithm that we propose.

Horowitz and Malkhi [19] propose to maintain a logical ring between overlay nodes. Each joining node is given a contact node assigned at random, in order to join the ring; it then simply increments its successor’s estimation of the logarithm of the system size. This technique has an expected accuracy within the range  $n/2-n^2$ .

Finally, Bawa et al. [6] propose a method which assumes one can sample peers uniformly at random. They form an estimate of system size based on the number of samples required before the same peer is sampled twice. The cost, measured in number of samples, scales like  $\ell\sqrt{N}$  where  $N$  is the system size, and for a target relative error of  $1/\sqrt{\ell}$ . The technique we shall describe builds on this work. It improves upon it by proposing a scheme to generate approximately uniform random samples, and also reduces the number of samples required to achieve the same target accuracy to  $\sqrt{\ell N}$ , hence a reduction by a factor of  $\sqrt{\ell}$ .

We can also mention the *Random Tour* approach that we presented in an earlier version of this paper [30]. This approach is based on the return time of a continuous time random walk to the node originating the query.

### 3 The S&C method

In this section we present an algorithm which is based on, and improves upon, a technique proposed in [6]. This technique essentially relies on sampling uniformly at random from the peer population. It then produces an estimate of system size based on how many random samples are required before two samples return the same peer.

We improve the proposal of [6] in two ways. First, we propose a uniform peer sampling technique which produces unbiased samples by emulating a CTRW. Many existing proposals rely on DTRW stopped at a *fixed* time and consequently suffer from a bias whenever peers have unequal degrees.

Second, we refine the way those samples are used, and effectively obtain estimates with a given variance with fewer sampling steps.

#### 3.1 Peer sampling with CTRW

The probing peer’s label is denoted by  $i$ , and the overlay is modelled as an undirected graph  $G$ . We use the standard CTRW, namely the random walk where each visit to a node  $j$  lasts for an exponentially distributed random time with expected duration  $1/d_j$ , where  $d_j$  is the degree of node  $j$ . The stationary distribution of the standard DTRW puts mass  $d_j/\sum_k d_k$  on each node  $j$ , and is thus biased towards high degree nodes. In contrast, the CTRW we described has a uniform stationary distribution. Our peer sampling algorithm proceeds as follows.

1. A timer is set at some predefined value  $T > 0$ , by the initiator, node  $i$ , in a sampling message.
2. Any node  $j$ , either after receiving the sampling message, or (if it is the initiator) after having initialised the timer, does the following. It picks a random number  $U$ , uniformly distributed on  $[0, 1]$ . It decrements  $T$  by  $\log(1/U)/d_i$  (i.e.,  $T \leftarrow T - \log(1/U)/d_i$ ). If  $T \leq 0$ , then this node  $j$  is the sampled node; it returns its ID to the initiator, and the procedure stops. Otherwise, it forwards the message with the updated timer to one of its  $d_j$  neighbors, chosen uniformly at random.

This procedure returns a random node sample, the distribution of which is exactly that of the state of the standard CTRW at time  $T$ , started from node  $i$ . This follows from the well-known fact that  $\log(1/U)$  has a unit mean exponential distribution [37].

The CTRW on  $G$  is in fact a continuous time Markov chain with state space  $G$  and infinitesimal generator given by the negative of the Laplacian of  $G$ , defined below.

**Definition 1** The Laplacian matrix of a graph  $G$  is the matrix  $L$  such that  $L_{ij} = -1$  if  $i \neq j$ , and  $(i, j)$  is an edge of the graph  $G$ ,  $L_{ij} = d_i$  if  $j = i$ , and  $L_{ij} = 0$  otherwise. Its eigenvalues  $\lambda_1, \dots, \lambda_N$  are real and non-negative. Assuming they are sorted in non-decreasing order ( $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ ), then  $\lambda_1 = 0$ , and  $\lambda_2$  is called the spectral gap of the graph.

The mixing time of the CTRW on  $G$  is related to the spectral gap of the Laplacian  $L$ , as Lemma 1 below shows.

A convenient measure of accuracy of the proposed sampling technique is provided by the *variation distance* between the probability distribution of the returned sample, and the target uniform probability distribution. Recall that the variation distance between two measures  $p, q$  on  $\mathcal{N}$  is defined as

$$d(p, q) := \frac{1}{2} \sum_{i \in \mathcal{N}} |p_i - q_i|.$$

It admits the following useful interpretation [26, Theorem 5.2]: random variables  $X$  and  $Y$  with marginal distributions  $p$  and  $q$  can be defined on the same probability space such that  $X = Y$  with probability  $1 - d(p, q)$ . Moreover, this is the largest probability of coincidence achievable among all joint distributions with the given marginals  $p$  and  $q$ .

We can now relate the quality of our sampling method to the choice of  $T$ , and the spectral gap of the graph:

**Lemma 1** *Let  $\{X_t\}_{t \geq 0}$  be a continuous time, reversible Markov process on a finite state space  $\mathcal{N}$ , with spectral gap  $\lambda_2$ , and stationary distribution  $\pi$ . Denote by  $p_{i \cdot}(t)$  the distribution of  $X_t$  when the process is started at  $X_0 = i$ . Then it holds that*

$$d(p_{i \cdot}(t), \pi) \leq \frac{1}{2\sqrt{\pi_i}} e^{-\lambda_2 t}.$$

*Proof* By Lemma 8, p. 10, Chap. 4 in [2], it holds that

$$d(p_{i \cdot}(t), \pi) \leq \frac{1}{2\sqrt{\pi_i}} \sqrt{p_{ii}(2t) - \pi_i}. \tag{1}$$

Besides, as mentioned on Eq. (46), p. 20, Chap. 3 in [2], the function  $t \rightarrow p_{ii}(t) - \pi_i$  is *completely monotone* (see p. 19, Chap. 3 in [2] for a definition), and thus, by Lemma 13, p. 20, Chap. 3 [2], it verifies

$$p_{ii}(t) - \pi_i \leq [p_{ii}(0) - \pi_i] e^{-\lambda_2 t}.$$

Combined with (1), this yields the claim of the lemma.  $\square$

When specialised to the standard CTRW, for which  $\pi_i = 1/N$ , taking  $T = c \log(N)/\lambda_2$ , we obtain

$$d(p_{i \cdot}(T), \pi) \leq \frac{1}{2} N^{(1/2)-c}.$$

If for instance  $c \geq 3/2$ , then this variation distance is of order  $N^{-1}$ .

In view of the above-mentioned interpretation of variation distance, for such a choice of  $T$ ,  $X_T$  coincides with a uniform random sample from  $\mathcal{N}$  with probability  $1 - O(N^{-1})$ . Equivalently, it takes on average of the order of  $N$  samples before retrieving an improperly selected node.

The reason this is important is the following: we shall see that our algorithm requires of the order of  $\sqrt{N}$  uniform random samples in order to estimate  $N$ . But since we cannot sample uniformly at random exactly, we use the CTRW procedure described above to obtain *approximately uniform* random samples. The error estimate on the approximation tells us that for  $T$  chosen as above, it is *as if* we sampled exactly from the uniform distribution; with high probability, our sampling procedure will yield the same samples (on runs of length  $O(N)$ ) as the exact procedure.

*Remark 1* Instead of the standard CTRW, we could alternatively base sampling on the CTRW with deterministic sojourn times. This suppresses the need to generate uniform random numbers  $U$  at nodes traversed by the random walk. However, in general there is no analogue of Lemma 1 for the deterministic sojourn time CTRW, as the following counter-example shows.

Consider a bipartite, regular graph with common node degrees  $d$ , nodes being partitioned into  $\mathcal{N}_1$  and  $\mathcal{N}_2$ , with  $|\mathcal{N}_1| = |\mathcal{N}_2|$ . Then with the latter CTRW, whenever  $\lfloor T/d \rfloor$  is even, the returned node belongs to the same bipartition as the node  $i$  from which sampling started, no matter how large  $T$  is. Assume say that node  $i$  belongs to  $\mathcal{N}_1$ . Then the variation distance between the sampled distribution and the uniform distribution is at least  $\sum_{j \in \mathcal{N}_2} 1/N = 1/2$ , and does not go to zero.

The problem arises because this is effectively a discrete time Markov chain which is periodic due to the graph being bipartite. Aperiodicity can be guaranteed by introducing self-loops into the graph. A related approach, which involves adding enough self-loops to make the graph regular, is considered in [5].

We should mention that, as both  $N$  and  $\lambda_2$  are a priori unknown, it is in practice not feasible to set  $T$  to precisely say,  $2 \log(N)/\lambda_2$ . One possibility is to use sampling with a first value of  $T$ , get back from the S&C procedure (described in Sect. 3.2) an estimate  $\hat{N}_1$  of  $N$ , then re-run the whole procedure with  $2\hat{N}_1$  instead of  $T$ , get a new estimate  $\hat{N}_2$  of  $T$ , and repeat until estimates  $\hat{N}_i$  appear to stabilise; they should increase with  $T$  until  $T$  is sufficiently large. While there are pathological graphs on which this method would fail (e.g., a “dumbbell”, which consists of two large densely connected components joined by a long, narrow bridge), we believe that it would work on most commonly used overlays. Evaluating this approach on real-world overlays is left to future work.

The approach we take in this paper is to assume that suitable lower bounds on  $\lambda_2$ , and upper bounds on  $\log(N)$  are known, so that a conservative value of  $T$  can be used. To see how such lower bounds on  $\lambda_2$  could be obtained, we now relate the spectral gap  $\lambda_2$  to the expansion properties of the graph  $G$ .

We introduce the notation

$$I(G) := \inf_{S: |S| \leq N/2} \frac{E(S, \bar{S})}{|S|},$$

where  $E(S, \bar{S})$  denotes the number of edges in the graph  $G$  between the set of nodes  $S$ , and the complementary set  $\bar{S}$ . The constant  $I(G)$  is known as the isoperimetric constant of the graph, or also as its conductance; see e.g. Mohar [31] for further discussion. The so-called Cheeger inequality (see [31]) states that the spectral gap  $\lambda_2$  of the graph verifies

$$\lambda_2 \geq \frac{I(G)^2}{2\Delta(G)},$$

where  $\Delta(G)$  is the maximal degree of nodes in the graph. The conductance parameter  $I$  is sometimes called the expansion parameter of a graph, and graphs with large  $I$  are referred to as expanders. The reader can find additional material on expanders in [3], or [27]. Several overlay architectures proposed in the literature ensure good expansion properties by design: the expansion parameter  $I$  is bounded away from 0.

In particular, overlays comprising sufficiently many “random” edges have large expansion parameter. It is shown for instance in [17], Theorem 5.4, that Erdős-Rényi graphs on  $N$  nodes with average degree  $d$  such that  $d \gg \log(N)$  have an expansion of  $d/2$ . It is also shown in [18] that, if each node chooses  $m \geq 2$  other nodes uniformly at random as its neighbours in the overlay, then the resulting graph has expansion at least  $m/5$ .

### 3.2 Estimation procedure

The technique we use is as follows. We pick an integer  $\ell > 0$ , which will determine the accuracy of our estimates. We then obtain node samples  $X(1), \dots, X(n)$ . Denote by  $C_1$  the first time  $n$  when a sample  $X(n)$  is obtained which has already been seen (the first collision), i.e., for some  $m < n$ ,  $X(m) = X(n)$ . Likewise, denote by  $C_2$  the second time  $n$  when the corresponding sample  $X(n)$  has previously been observed, and define  $C_i$  similarly for  $i \geq 1$ . We shall stop sampling at  $n = C_\ell$ , that is, when exactly  $\ell$  newly obtained samples have previously been observed, where  $\ell$  is a fixed control parameter.

For a given  $N$ , we denote by  $L_N(n_1, \dots, n_\ell)$  the probability that  $C_1 = n_1, \dots, C_\ell = n_\ell$ . Elementary combinatorics show that

$$L_N(n_1, \dots, n_\ell) = N^{-n_\ell} [N(N-1) \cdots (N-n_\ell + \ell + 1)] \times [(n_1 - 1)(n_2 - 2) \cdots (n_\ell - \ell)].$$

This formula can be written as the product of two terms, one that is a function of  $(n_1, \dots, n_\ell)$  only, and another that is a function of  $N$  and  $n_\ell$  only. This implies that  $C_\ell$  is a *sufficient statistic* for the estimation of  $N$ : all the information about the unknown parameter  $N$  that is carried by the observations  $C_1, \dots, C_\ell$  is contained in the variable  $C_\ell$ . Or to put it another way, given  $C_\ell$ , the other  $C_i$ s do not contain any additional information about  $N$ . Hence, the best estimator (for any performance measure) based on  $C_1, \dots, C_\ell$  is a function of  $C_\ell$  only.

Our approach to estimating  $N$  will be to use the Maximum Likelihood (ML) method. Note that

$$\begin{aligned} \frac{\partial}{\partial N} \log L_N(C_1, \dots, C_\ell) &= -\frac{C_\ell}{N} + \sum_{i=0}^{C_\ell - \ell - 1} \frac{1}{N - i} \\ &= \frac{1}{N} \left[ -\ell + \sum_{i=0}^{C_\ell - \ell - 1} \frac{i}{N - i} \right]. \end{aligned} \quad (2)$$

The derivative of the log-likelihood function is called the score, and the expectation of its square is called the Fisher information. These quantities play a role in determining the variance of the optimal estimator, as we shall see later.

It is clear from the formula above that the likelihood is well defined for  $N$  in the interval  $N \in (C_\ell - \ell - 1, +\infty)$ , is increasing on the interval  $(C_\ell - \ell - 1, \widehat{N}]$ , and decreasing on  $[\widehat{N}, +\infty)$ , where  $\widehat{N}$  is the ML estimate. (It is also clear that  $C_\ell$  cannot be bigger than  $N + \ell$ .) Thus the ML estimate  $\widehat{N}$  can be readily computed by solving the equation

$$F(N) := \sum_{i=0}^{C_\ell - \ell - 1} \frac{i}{N - i} - \ell = 0, \quad (3)$$

using standard bisection search. Before detailing the procedure, we note the following. The monotonic decreasing function  $F(N)$  satisfies

$$\begin{aligned} \frac{(C_\ell - \ell - 1)(C_\ell - \ell)}{2N} - \ell &\leq F(N) \\ &\leq \frac{(C_\ell - \ell - 1)(C_\ell - \ell)}{2(N - (C_\ell - \ell - 1))} - \ell. \end{aligned}$$

Each of the bounding functions is also monotonic decreasing in  $N$ . This readily implies that the ML estimate  $\widehat{N}$  lies in the interval  $[N^-, N^+]$ , where

$$\begin{cases} N^- = \frac{(C_\ell - \ell - 1)(C_\ell - \ell)}{2\ell}, \\ N^+ = \frac{(C_\ell - \ell - 1)(C_\ell - \ell)}{2\ell} + C_\ell - \ell - 1. \end{cases} \quad (4)$$

The bisection search determination of  $\widehat{N}$  then proceeds as follows. Initialize the search range  $[N^-, N^+]$  with the values given in (4). Then repeat the following step until  $N^+ - N^- \leq 1$ . Set  $N = (N^+ + N^-)/2$ ; if  $F(N) > 0$ , set  $N^- = N$ ; otherwise set  $N^+ = N$ .

### 3.3 Accuracy/complexity trade-off

We now provide an asymptotic analysis of the quality of the proposed ML estimate when  $N$  is large. This will then be used to analyse the accuracy/complexity trade-off of the S&C procedure, under the assumption that samples returned by the CTRW module are indeed uniformly distributed.

**Proposition 1** *Let  $\ell > 0$  be fixed. As  $N$  tends to infinity,  $C_\ell^2/(2N)$  converges in distribution to a gamma random variable:*

$$\frac{C_\ell^2}{2N} \rightarrow E_1 + \dots + E_\ell, \tag{5}$$

where  $E_1, \dots, E_\ell$  are i.i.d. random variables that are exponentially distributed with parameter 1.

Furthermore, for any positive  $p$ , we also have convergence of the  $p$ -th moments:

$$\mathbf{E}_N \left[ \left( \frac{C_\ell^2}{2N} \right)^p \right] \rightarrow \mathbf{E} [(E_1 + \dots + E_\ell)^p]. \tag{6}$$

*Proof* We prove the weak convergence property by induction on  $\ell$ . We shall evaluate the following conditional probability:

$$\mathbf{P}_N(C_\ell - C_{\ell-1} > b\sqrt{N} | C_{\ell-1} = a_N\sqrt{N}),$$

where  $a_N \sim a$ ,  $a$  is a fixed positive number, as  $N \rightarrow \infty$ . Let  $m = a_N\sqrt{N} - (\ell - 1)$ , and  $k = b\sqrt{N}$ . Elementary combinatorics show that this conditional probability equals

$$\frac{1}{N^k} (N - m)(N - m - 1) \dots (N - m - k + 1),$$

and hence:

$$\begin{aligned} & \log \left( \mathbf{P}_N \left( C_\ell - C_{\ell-1} > b\sqrt{N} \mid C_{\ell-1} = a_N\sqrt{N} \right) \right) \\ &= \sum_{i=0}^{k-1} \log \left( \frac{N - m - i}{N} \right) \sim - \sum_{i=0}^{k-1} \frac{i + m}{N} \\ &\sim -ab - \frac{b^2}{2}. \end{aligned}$$

We thus have the following equivalent as  $N \rightarrow \infty$ :

$$\mathbf{P}_N \left( C_\ell - C_{\ell-1} > b\sqrt{N} \mid C_{\ell-1} = a_N\sqrt{N} \right) \sim e^{-ab - b^2/2}.$$

In turn, setting  $a = \sqrt{2y}$  and  $b = \sqrt{2(x + y)} - \sqrt{2y}$ , we get

$$\mathbf{P} \left( \frac{C_\ell^2}{2N} > x + y \mid \frac{C_{\ell-1}^2}{2N} = y \right) \sim e^{-ab - b^2/2} = e^{-x}.$$

This establishes the claimed weak convergence property.

In order to deduce convergence of moments from weak convergence, it is enough to show that for all  $p > 0$ , the distributions of variables  $[C_\ell/\sqrt{N}]^p$  for varying  $N$  are uniformly integrable (for a definition see e.g. [7]). By a standard criterion for uniform integrability, this will follow if for some positive  $\theta$ , it holds that

$$\sup_{N>0} \mathbf{E}_N \left[ e^{\theta C_\ell/\sqrt{N}} \right] < \infty. \tag{7}$$

By a simple coupling argument (see [26] for background on coupling), it can be shown that the distribution of  $C_\ell$  is stochastically dominated by that of the sum of  $\ell$  independent copies of  $C_1$ . Thus,

$$\mathbf{E}_N \left[ e^{\theta C_\ell/\sqrt{N}} \right] \leq \left[ \mathbf{E}_N \left( e^{\theta C_1/\sqrt{N}} \right) \right]^\ell.$$

It is therefore enough to prove (7) in the special case where  $\ell = 1$ . Write

$$\begin{aligned} \mathbf{E}_N \left[ e^{\theta C_1/\sqrt{N}} \right] &= \int_0^\infty \mathbf{P}_N \left( e^{\theta C_1/\sqrt{N}} > y \right) dy \\ &= \int_0^\infty \mathbf{P}_N \left( C_1 > \frac{\sqrt{N}}{\theta} \log(y) \right) dy. \end{aligned}$$

We bound the integrand in the last expression as follows. Set  $k = \sqrt{N} \log(y)/\theta$ . Then,

$$\begin{aligned} \mathbf{P}_N(C_1 > k) &= \exp \left( \sum_{i=0}^{k-1} \log(1 - i/n) \right) \\ &\leq \exp \left( -(k - 1)^2/n \right). \end{aligned}$$

Combined with the previous expression, this yields

$$\begin{aligned} & \mathbf{E}_N \left[ e^{\theta C_1/\sqrt{N}} \right] \\ &\leq 1 + \int_1^\infty \exp \left( - \frac{(\sqrt{N} \log(y)/\theta - 1)^2}{N} \right) dy \\ &= 1 + \int_{-1/\sqrt{N}}^\infty e^{-v^2} e^{\theta(v+1/\sqrt{N})} \theta dv \\ &\leq 1 + \int_{-1}^\infty e^{-v^2 + \theta v + \theta} \theta dv, \end{aligned}$$

where the equality is obtained by the change of variables  $v = \log(y)/\theta - 1/\sqrt{N}$ . The final term is finite and independent of  $N$ , which implies the announced uniform integrability.  $\square$

This proposition allows us to evaluate the asymptotic mean square error of the ML estimate:

**Corollary 1** *The ML estimate  $\widehat{N}$  is such that*

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \mathbf{E}_N \left( \widehat{N} - N \right)^2 = \frac{1}{\ell}. \tag{8}$$

*Proof* We have by the Cauchy–Schwarz inequality that

$$\begin{aligned} & \mathbf{E}_N \left[ \left( \widehat{N} - N \right)^2 \right] \\ &= \mathbf{E}_N \left[ \left( N^- - N \right)^2 \right] + 2\mathbf{E}_N \left[ \left( \widehat{N} - N^- \right) \left( N^- - N \right) \right] \\ &\quad + \mathbf{E}_N \left[ \left( \widehat{N} - N^- \right)^2 \right] \\ &\leq \mathbf{E}_N \left[ \left( N^- - N \right)^2 \right] \\ &\quad + 2\sqrt{\mathbf{E}_N \left( \widehat{N} - N^- \right)^2} \sqrt{\mathbf{E}_N \left( N^- - N \right)^2} \\ &\quad + \mathbf{E}_N \left[ \left( \widehat{N} - N^- \right)^2 \right]. \end{aligned} \tag{9}$$

In view of the bounds (4),  $\widehat{N} - N^-$  is bounded in absolute value by  $N^+ - N^- \leq C_\ell$ . Hence,

$$\mathbf{E}_N \left[ (\widehat{N} - N^-)^2 \right] \leq \mathbf{E}_N (C_\ell^2) \sim 2\ell N, \quad (10)$$

by (6). Convergence of moments (6) also guarantees, in view of the expression for  $N^-$ , that

$$\mathbf{E}_N [N^-] \sim \mathbf{E}_N \left( \frac{C_\ell^2}{2\ell} \right) \sim N, \quad (11)$$

while

$$\begin{aligned} \mathbf{E}_N \left[ (N^-)^2 \right] &\sim \mathbf{E}_N \left( \frac{C_\ell^4}{4\ell^2} \right) \\ &\sim \frac{N^2}{\ell^2} \mathbf{E}[(E_1 + \dots + E_\ell)^2] \\ &= \frac{N^2}{\ell^2} (\ell^2 + \ell). \end{aligned} \quad (12)$$

Here, we have used the fact the exponential distribution with parameter 1 has mean 1 and variance 1. Now, by (11) and (12),

$$\mathbf{E}_N [(N^- - N)^2] \sim \mathbf{E}_N [(N^-)^2] - N^2 \sim \frac{N^2}{\ell}. \quad (13)$$

Substituting (10) and (13) in (9), we get

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \mathbf{E}_N (\widehat{N} - N)^2 = \frac{1}{\ell},$$

as claimed.  $\square$

The corollary shows that, for a variance of the order of  $N^2/\ell$ , we use  $C_\ell$  samples, hence on average a number of samples of the order of  $\sqrt{N\ell}$ . The average number of messages exchanged in a single sampling step is, assuming the originator is randomly selected, equal to  $T\bar{d}$ , where  $\bar{d} = N^{-1} \sum_j d_j$  is the average node degree. Thus, assuming as in Sect. 3.1 that  $T$  equals  $2 \log(N)/\lambda_2$ , the average number of messages used by the S&C method is of order  $(\bar{d} \log(N) \sqrt{N\ell}/\lambda_2)$ , for an estimate with relative variance of  $1/\ell$ . This presents an improvement on the cost of the ‘‘inverted birthday paradox method’’ of [6] by a factor  $\sqrt{\ell}$ .

Next, we show that no other unbiased estimator can do substantially better.

**Lemma 2** *Let  $\tilde{N} = f(C_\ell)$  be an arbitrary estimator with the property that  $\mathbf{E}_N[f(C_\ell)] = N$ . Then,*

$$\liminf_{N \rightarrow \infty} \frac{1}{N^2} \text{Var}_N(\tilde{N}) \geq \frac{1}{\ell}.$$

It is a well-known result in statistics that the maximum likelihood estimator is asymptotically efficient (has asymptotic mean square error no larger than that of the best unbiased estimator), but this is in the context where the parameter to

be estimated is fixed, while the number of samples increases to infinity. In the setting we are studying, both the parameter and the number of observations go to infinity; hence, we have included a proof of optimality.

*Proof* Let  $\tilde{N} = f(C_1, \dots, C_\ell)$  be any unbiased estimator of  $N$ , i.e.,  $\mathbf{E}_N[f(C_1, \dots, C_\ell)] = N$ . We shall use the Cramér–Rao inequality (see, e.g., [11, Theorem 12.11.1]) to obtain a lower bound on the variance of this estimator.

To use this inequality, we need to compute the Fisher information (a measure of the ‘‘information’’ that the random vector  $(C_1, \dots, C_\ell)$  contains about the parameter  $N$ ). The Fisher information  $I(N)$  is defined as the variance of the score function,

$$s(N) = \frac{\partial}{\partial N} \log L_N(C_1, \dots, C_\ell).$$

Therefore, it follows from (2) that

$$\begin{aligned} I(N) &= \frac{1}{N^2} \mathbf{E}_N \left[ \left( \sum_{i=0}^{C_\ell-1} \frac{i}{N-i} \right)^2 \right] + \frac{\ell^2}{N^2} \\ &\quad - \frac{2\ell}{N^2} \mathbf{E}_N \left[ \sum_{i=0}^{C_\ell-1} \frac{i}{N-i} \right]. \end{aligned} \quad (14)$$

Recall that the mean of the score function is zero (see, e.g., [11, Sect. 12.11]). Thus, by (2),

$$\mathbf{E}_N \left[ \sum_{i=0}^{C_\ell-1} \frac{i}{N-i} \right] = \ell.$$

Substituting this in (14) yields

$$I(N) = \frac{1}{N^2} \mathbf{E}_N \left[ \left( \sum_{i=0}^{C_\ell-1} \frac{i}{N-i} \right)^2 \right] - \frac{\ell^2}{N^2}. \quad (15)$$

Now observe using (6) and Markov’s inequality that, for any fixed  $\epsilon > 0$

$$\mathbf{P}_N(C_\ell > \epsilon N) \leq (\epsilon N)^{-6} \mathbf{E}_N[C_\ell^6] \leq cN^{-3}, \quad (16)$$

for some constant  $c > 0$  that depends on  $\epsilon$  but not on  $N$ . Now, on the event that  $C_\ell > \epsilon N$ ,

$$\sum_{i=0}^{C_\ell-1} \frac{i}{N-i} \leq \sum_{i=0}^{N-1} \frac{i}{N-i} \leq N \log N,$$

whereas, on the event that  $C_\ell \leq \epsilon N$ ,

$$\sum_{i=0}^{C_\ell-1} \frac{i}{N-i} \leq \frac{1}{1-\epsilon} \frac{C_\ell^2}{2N}.$$

Hence,

$$\begin{aligned} \mathbf{E}_N \left[ \left( \sum_{i=0}^{C_\ell-1} \frac{i}{N-i} \right)^2 \right] &\leq (N \log N)^2 \mathbf{P}_N(C_\ell > \epsilon N) \\ &\quad + \frac{1}{(1-\epsilon)^2} \mathbf{E}_N \left[ \left( \frac{C_\ell^2}{2N} \right)^2 \right]. \end{aligned} \quad (17)$$

Now, the first term in the last expression above goes to zero as  $N \rightarrow \infty$  by (16), while the second term goes to  $(\ell^2 + \ell)/(1 - \epsilon)^2$  by (6), and well-known properties of the exponential distribution. Hence, we have from (15) and (17) that

$$I(N) \leq \frac{1}{N^2} \left( \frac{\ell^2 + \ell}{(1 - \epsilon)^2} - \ell^2 \right).$$

Hence, by the Cramér–Rao bound [11, Theorem 12.11.1], for any unbiased estimator  $\tilde{N} = f(C_1, \dots, C_\ell)$ , we have

$$\text{Var}(\tilde{N}) \geq \frac{1}{I(N)} \geq \frac{(1 - \epsilon)^2 N^2}{\ell + (2\epsilon - \epsilon^2)\ell^2}.$$

Since  $\ell$  is fixed, letting  $\epsilon$  decrease to zero yields the claim of the lemma.  $\square$

*Remark 2* Observe that, since  $C_\ell^2/(2\ell) \sim N$  (in probability and expectation), the bounds  $N^-$  and  $N^+$  in (4) are both asymptotic to  $N$ , and differ only by a term of order  $\sqrt{N}$ . Hence, instead of computing the maximum likelihood estimator by bisection search, we could equally well use either  $N^-$  or  $N^+$ , or the asymptotically unbiased estimator  $\tilde{N} = C_\ell^2/(2\ell)$ . All three estimators are within  $\sqrt{N}$  of the ML-estimator, and hence, all three are asymptotically efficient. In fact, for ease of computation, we use the estimator  $\tilde{N} = C_\ell^2/2\ell$  in the next section, where we evaluate the algorithm.

## 4 Experimental results

We evaluated the S&C approach through simulations. In this section, we report the evaluation results. First, we describe the experimental set-up. We then report results on the accuracy and cost of the proposed approach in a static and dynamic environments with both gradual and abrupt changes in system sizes.

### 4.1 Setup and evaluation criteria

Our experiments are simulation-based; we used PeerSim [35], a discrete event/cycle based simulator, well-suited for large-scale networks. We consider overlay networks of exactly 100,000 nodes in the static case, and of a size ranging from 50,000 to 150,000 nodes in the dynamic setting.

In order to assess the impact of the underlying network topology and connectivity, we consider two types of topologies in the evaluation, which we refer to as balanced and scale-free random graphs.

Overlays of the first type (balanced random graphs) are generated so as to guarantee node degrees lying between 1 and 10, in the following manner. Sequentially, each node  $i$  selects a random number  $d_i^{out}$  between 1 and 10. This ensures that the degrees are not homogeneous with a reasonable spread of the distribution around the average. We believe

that this is a realistic assumption for the considered peer-to-peer systems targeted. It then selects  $d_i^{out}$  target nodes at random, among target nodes with a current degree  $< 10$ . Then  $d_i^{out}$  undirected edges are created between node  $i$  and its  $d_i^{out}$  targets, whose degree is increased by 1 at this stage. A node that selects a  $d_i^{out}$  less than or equal to its current number of neighbors (previous wiring) obviously immediately stops its wiring process. The resulting average degree is between 7 and 8. From the results of [18], we expect such graphs to have large expansion, hence a favourable situation for our approach. However, it should be noted that existing overlay maintenance protocols aim to maintain graphs with similar statistical properties; see e.g. [22] and [16]. Therefore, we believe that this is a realistic setting for practical peer-to-peer systems.

In scale-free networks on the other hand, the node degree distribution follows a power law; the Internet and the World-Wide-Web have this property. We generate random scale-free graphs using the preferential attachment scheme of Barabási and Albert [1]. Here, each new node added to the network chooses its links preferentially targeting high-degree nodes. The result is a random graph in which the probability that a node has  $k$  neighbors decays like  $k^{-3}$ . Thus, node degrees are much more varied than in the balanced random graph model.

In the dynamic scenarios, newly incorporated nodes are connected via their own set of random targets, chosen according to the rule for the corresponding model. Nodes to be removed are selected uniformly at random, and the remaining nodes that lose neighbors do not search for new ones. The actual system size we report is always that of the connected component to which the probing node belongs.

We evaluate our algorithm for estimating system size along the following metrics. *Accuracy* relates to the relative error in the system size estimate and is clearly a basic criterion. It can be improved by taking more measurements. So there is a tradeoff with the *Overhead*, specified as the number of messages required to obtain the system size estimate. Depending on the application, a quick approximate estimate could be preferable to a more accurate one which would take much longer to compute, and create more overhead. This could also be the case when churn is high, causing the system size to change rapidly. In that case, *Reactivity to changes* is an important characteristic of the algorithms. To evaluate this, we compute the time to react to a growth or increase in the number of peers in the system.

### 4.2 Results in static settings

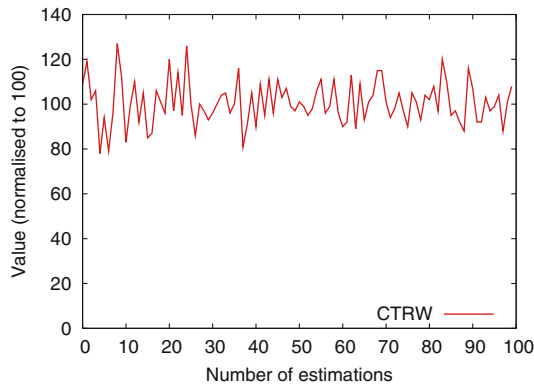
#### 4.2.1 Balanced random graphs

We performed repeated runs on a 100,000 node overlay of S&C with  $\ell = 10$ , and with  $\ell = 100$ . For both instances, the timer value used in the sampling module was fixed to

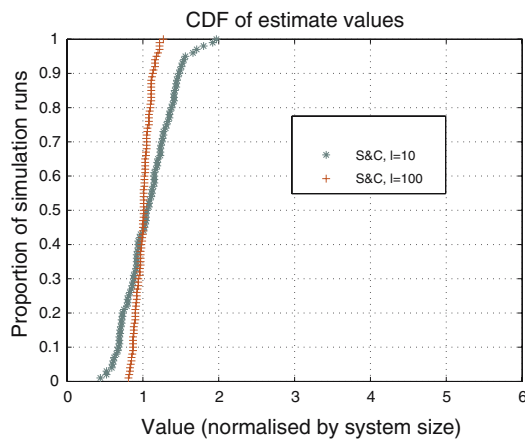
$T = 10$ . In view of our suggestion to take  $T \approx 2 \log(N)/\lambda_2$ , this is consistent with a spectral gap  $\lambda_2 > 2.3$ .

Figure 1 plots a run of S&C method with  $\ell = 100$ . It shows that S&C with  $\ell = 100$  provides an accuracy of  $\pm 20\%$ .

The cumulative distribution functions (cdfs) of the normalised estimate values for S&C ( $\ell=10$  and  $100$ ) are displayed on Fig. 2. The steeper the curve, the less dispersed the sample values. This is further illustrated by the summary statistics reported in Table 1. Our method provides samples with the correct mean value; The variances of S&C configurations match the theoretical prediction, and coincide with  $1/\ell$ .



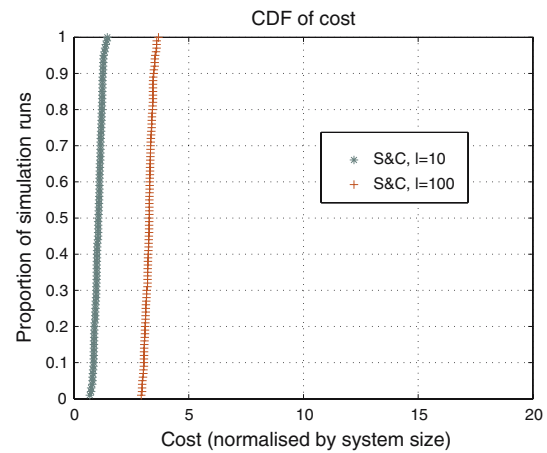
**Fig. 1** Sample&collide with  $l = 100$ , on a 100,000 node random graph.



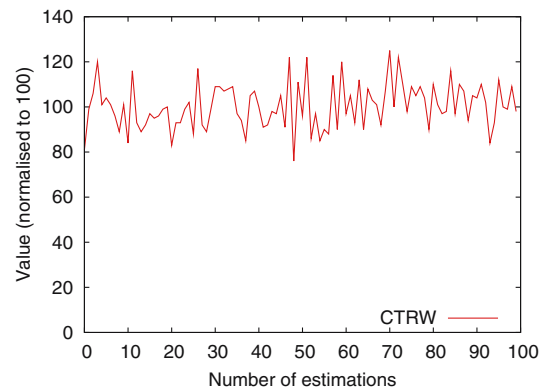
**Fig. 2** CDF of estimate values, normalised by system size, on a 100,000 overlay random graph, with  $\ell = 10$  and  $\ell = 100$ .

**Table 1** Summary statistics of sampling strategies: mean and variance of normalised estimate values, and mean and variance of normalised estimate costs

Algorithm	S&C, $\ell = 10$	S&C, $\ell = 100$
Average value	1.08	1.01
Variance(value)	0.1	0.01
Average cost	1.08	3.27
Variance(cost)	0.1	0.02



**Fig. 3** CDF of estimation cost in messages, normalised by system size, on a 100,000 overlay random graph, with  $\ell = 10$  and  $100$ .



**Fig. 4** Sample&collide with  $l = 100$ , on a 100,000 node scale-free graph.

We next report on the costs incurred in a single run of the approach. The cdfs of costs, normalised by system size, are shown in Fig. 3.

Note that S&C ( $\ell=100$ ) incurs a cost per run that is larger than that of S&C ( $\ell=10$ ) by only a factor of 3.27 (consistent with the ratio of  $\sqrt{100}/\sqrt{10} \approx 3.16$  predicted by the analysis), for a variance reduction by a factor of 10.

#### 4.2.2 Scale-free graphs

Figure 4 depicts the system size estimates as a percentage of the actual system size, on scale-free graphs in the static scenario. The plots show that S&C methods achieve accuracy comparable to what they achieved in the balanced random graph setting. This suggests that they are capable of dealing with considerable node heterogeneity in providing unbiased estimates of system size.

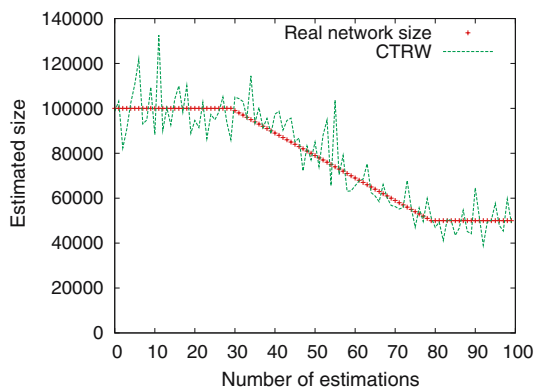
### 4.3 Results in dynamic settings

In order to assess the reactivity to changes of the proposed approach, we performed repeated runs on random graphs with varying numbers of nodes for the procedure  $\ell = 100$ .

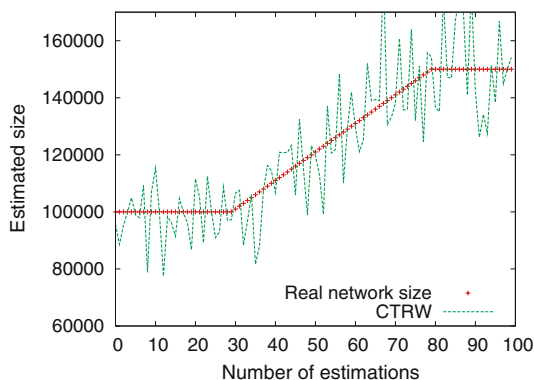
We considered three distinct dynamic scenarios:

- Shrinking network: gradual decrease, where the node population steadily decreases from 100,000 to 50,000;
- Growing network: gradual increase, where the node population grows regularly from 100,000 to 150,000;
- Catastrophic changes, where the initial node population of 100,000 is suddenly decreased to 75,000 and then to 50,000, and finally faces a flash crowd with a sudden arrival of 25,000 nodes.

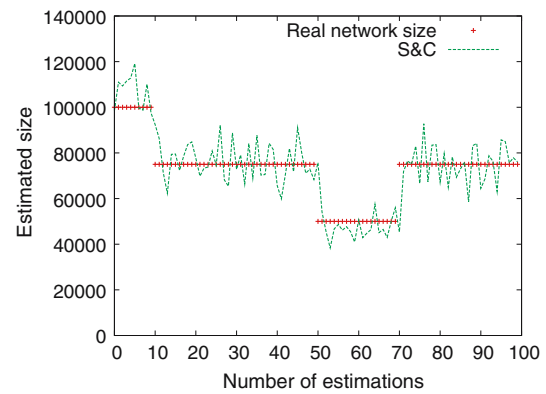
Performance of S&C is illustrated on a shrinking network in Fig. 5, on a growing network (gradual increase scenario) in Fig. 6, and on a network with catastrophic failures and flash crowd arrivals in Fig. 7.



**Fig. 5** *Sample&Collide* with  $\ell = 100$ , on shrinking network; 100,000 nodes at beginning, 50% nodes removal from run 30 to 80.



**Fig. 6** *Sample&Collide* with  $\ell = 100$ , on growing network; 100,000 nodes at beginning, 50% nodes join from run 30 to 80.



**Fig. 7** *Sample&Collide* with  $\ell = 100$ , under catastrophic failures:  $-25,000$  nodes at runs 10 and 50,  $+25,000$  nodes at run 70.

In general, changes in system size could affect the performance of S&C, because the quality of the random samples returned by the sampling module deteriorates if the expansion of the overlay is reduced. In the current settings, the results show that even if half of the nodes are removed in a random fashion, it is not sufficient to affect significantly the expansion properties of the graphs and that our algorithm still provides accurate results. However, as we observe on the three figures, the S&C estimator maintains a consistent level of accuracy. The analysis predicts a relative variance of  $1/\ell = 1/100$  for individual estimates, under the assumption of perfect random sampling. Hence theory predicts a relative standard deviation of 10% for the estimates plotted on these figures, provided sampling works well. Fluctuations on Figs. 5, 6 and 7 are consistent with a 10% magnitude.

We should mention at this stage that in the simulations reported here, we did not allow a departing node to leave the system with the probing message. Such a situation may however arise in practice, for instance due to node crash or improper departure from the overlay. To protect the sampling procedure used in S&C, one way of handling such message loss at the node initiating the measurement is to declare a probing message to be lost if it has not notified the sampling node (when it stops on the sampled node) in a given duration since its launch. The corresponding time-out parameter needs to be sufficiently large so that only few sampling trips time-out while the corresponding message is not lost, and still traveling through the system. One could for instance set this time-out to the average trip time, plus a few multiples of the trip time standard deviation. Here trip time refers to real-world time, as measured by the initiator's clock. Both standard deviation and average of trip time can be estimated adaptively from past measurements.

From the experimental results, we conclude that S&C and its sampling module are robust to changes in system size, both gradual and sudden.

## 5 Conclusion

In this paper, we addressed the issue of estimating the size of large-scale peer-to-peer overlay networks and proposed a peer counting approach that uses random walks for peer sampling. This is useful for basic overlay maintenance, and we expect it to be useful as well for applications such as live media streaming.

Our peer counting method requires random samples of peers. We proposed a peer sampling algorithm based on a CTRW and showed that it produces asymptotically uniform samples, in contrast to previous proposals which were biased towards high degree nodes. We showed that its cost for a specified accuracy is characterised by the expansion parameter of the overlay. We constructed a system size estimate based on the number of samples required to observe duplicated samples. We analysed in detail the asymptotic properties of this estimate, and showed that it makes the most efficient use of the information in the samples, by achieving the smallest possible variance. To our knowledge this achieves the best cost/accuracy trade-off of proposals to date [24], with a cost scaling like the square root of the system size, and the square root of the required accuracy (measured in reciprocal of relative variance). It is therefore a suitable candidate for large scale environments.

Finally we evaluated our scheme via simulations, in both static and dynamic environments. The simulation results confirmed the theoretical analysis and showed, furthermore, that the scheme was robust to system changes, both gradual and sudden.

## References

1. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47 (2002)
2. Aldous, D., Fill, J.: Reversible markov chains and random walks on graphs. Monograph in preparation, <http://stat-www.berkeley.edu/users/aldous/RWG/book.html>
3. Alon, N., Spencer, J.: *The Probabilistic Method*. Wiley, London (2002)
4. Alouf S., Altman E., Nain, P.: Optimal on-line estimation of the size of a dynamic multicast group. *IEEE INFOCOM '02* (2002)
5. Bar-Yossef, Z., Friedman, R., Kliot, G.: Rawms—random walk based lightweight membership service for wireless ad-hoc networks. In: *Mobihoc*. Florence, Italy (2006)
6. Bawa, M., Garcia-Molina, H., Gionis, A., Motwani, R.: Estimating aggregates on a peer-to-peer network. Technical Report, Dept. of computer science, Stanford University (2003)
7. Billingsley, P.: *Convergence of Probability Measures*. Wiley, London (1999)
8. Bolot, J.-C., Turetli, T., Wakeman, I.: Scalable feedback control for multicast video distribution in the internet. *ACM SIGCOMM* (1994)
9. Boyd, S., Ghosh, A., Prabhakar, B., Shah, D.: Gossip algorithms: design, analysis and applications. *IEEE INFOCOM* (2005)
10. Castro, M., Druschel, P., Ganesh, A., Rowstron, A., wallach, D.: Security for structured peer-to-peer overlay networks. *OSDI* (2002)
11. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Wiley, London (1991)
12. Dolev, D., Mokryn, O., Shavitt, Y.: On multicast trees: structure and size estimation. *IEEE INFOCOM* (2003)
13. Dolev, S., Schiller, E., Welsh, J.: Random walk for self-stabilizing group communication in ad-hoc networks. *PODC* (2002)
14. Eugster, P., Handurukande, S., Guerraoui, R., Kermarrec, A.-M., Kouznetsov, P.: Lightweight probabilistic broadcast. *ACM Trans. Comput. Syst.* **21**(4), (2003)
15. Friedman, T., Towsley, D.: Multicast session membership size estimation. *IEEE INFOCOM* (1999)
16. Ganesh, A.J., Kermarrec, A.-M., Massoulié, L.: Peer-to-peer membership management for gossip-based protocols. *IEEE Trans Comput.* **52**(2), 2003
17. Ganesh, A.J., Massoulié, L., Towsley, D.: The effect of network topology on the spread of epidemics. *IEEE INFOCOM* (2005)
18. Ganesh, A.J., Xue, F.: Expansion properties of  $k$ -out random graphs. Preprint (2004)
19. Horowitz, K., Malkhi, D.: Estimating network size from local information. *Inf. Process. Lett.* **88**(5), 237–243 (2003)
20. Jelasity, M., Montresor, A.: Epidemic-style proactive aggregation in large overlay networks. *ICDCS* (2004)
21. Jelasity M., Preuß, M.: On obtaining global information in a peer-to-peer fully distributed environment, vol. 2400. LNCS. Springer, Heidelberg, pp. 573–577 (2002)
22. Jelasity, M., Voulgaris, S., Guerraoui, R., Kermarrec, A.-M., van Steen, M.: Gossip-based peer sampling (submitted) (2007)
23. Kostoulas, D., Psaltoulis, D., Gupta, I., Birman, K., Demers, A.: Decentralized schemes for size estimation in large and dynamic groups. *IEEE NCA '05* (2005)
24. Le Merrer, E., Kermarrec, A.-M., Massoulié, L.: Peer to peer size estimation in large and dynamic networks: a comparative study. In: *The proceedings of HPDC-15*, June (2006)
25. Li, J., Lim, D.-Y.: A robust aggregation tree on distributed hash tables. MIT student oxygen workshop 2004, (2004)
26. Lindvall, T.: *Lectures on the Coupling Method*. Dover, New York (2002)
27. Linial, N., Wigderson, A.: Expander graphs and their applications. Lecture Notes, <http://www.cs.huji.ac.il/~nati/> (2002)
28. Malkhi, D., Naor, M., Ratajczak, D.: Viceroy: a scalable and dynamic emulation of the butterfly. In: *Proceedings of principles of distributed computing (PODC 2002)*, (2002)
29. Manku, G.S.: Routing networks for distributed hash tables. *PODC* (2003)
30. Massoulié, L., Le Merrer, E., Kermarrec, A.-M., Ganesh, A.J.: Peer counting and sampling in overlay networks: random walk methods. *PODC* (2006)
31. Mohar, B.: Some applications of laplace eigenvalues of graphs, in graph symmetry: algebraic methods and applications. In: Hahn, G., Sabidussi, G.(eds.) *NATO ASI Series C*, vol. 497. Kluwer, Dordrecht, pp. 225–275 (1997)
32. Mosk-Aoyama, D., Shah, D.: Computing separable functions via gossip. *ACM PODC* (2006)
33. Nonnenmacher, J., Biersack, E.W.: Optimal multicast feedback. *IEEE INFOCOMM '98* (1998)
34. Ntarmos, N., Triantafyllou, P., Weikum, G.: Counting at large: efficient cardinality estimation in internet-scale data networks. *Proc. ICDE '06* (2006)
35. <http://www.peersim.sourceforge.net>

36. Psaltoulis, D., Kostoulas, D., Gupta, I., Birman, K., Demers, A.: Practical algorithms for size estimation in large and dynamic groups. PODC (2004)
37. Ross, S.: Simulation. Elsevier, Amsterdam (2001)
38. Zhang, X., Liu, J., Li, B., Yum, T.-S.P.: Donet/coolstreaming: A data-driven overlay network for live media streaming. IEEE INFOCOM (2005)