

Mean-Field Analysis for the Evaluation of Gossip Protocols

Rena Bakhshi¹

Lucia Cloth²

Wan Fokkink¹

Boudewijn R. Haverkort²

¹Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam, Netherlands

[rbakhshi,wanf]@few.vu.nl

²Centre for Telematics & Information Technology
University of Twente
Enschede, Netherlands

[lucia, brh]@ewi.utwente.nl

ABSTRACT

Gossip protocols are designed to operate in very large, decentralised networks. A node in such a network bases its decision to interact (gossip) with another node on its partial view of the global system. Because of the size of these networks, analysis of gossip protocols is mostly done using simulation, which tend to be expensive in computation time and memory consumption.

We introduce mean-field analysis as an analytical method to evaluate gossip protocols. Nodes in the network are represented by small identical stochastic models. Joining all nodes would result in an enormous stochastic process. If the number of nodes goes to infinity, however, mean-field analysis allows us to replace this intractably large stochastic process by a small deterministic process. This process approximates the behaviour of very large gossip networks, and can be evaluated using simple matrix-vector multiplications.

1. INTRODUCTION

We consider large-scale networks where a large number of nodes interacts. In such networks, gossip protocols have shown to be a sensible paradigm for developing scalable and reliable communication mechanisms. For instance, information can be spread in a large-scale network if nodes periodically contact each other in a random fashion, and exchange their local information. When a large number of nodes interacts in a connected environment, various phenomena emerge that cannot be explained in terms of the behaviour of a single node. Therefore, we are interested in going from a detailed local model of the system at the node level to an abstract global model of the system.

To study the emergent behaviour of gossip protocols demands the consideration of large-scale networks [17]. Thus, the analysis of gossip protocols with automated tools is hard – it is, for example, beyond the capabilities of current probabilistic model-checking tools. In this paper, we show that mean-field analysis is well suited for a formal evaluation of gossip protocols. The stochastic process representing the modelled system converges to a deterministic process if the number of nodes goes to infinity, providing an approximation for large numbers of nodes.

This paper is further organized as follows. Sec. 2 gives a brief overview of the gossip paradigm, and explains an instance of such a protocol, i.e., the gossiping time protocol (GTP). In Sec. 3, we describe the necessary mean-field theory, and devise a simple analytical model for gossip-based information dissemination as an illustrative example. In Sec. 4

we present an analysis of the GTP using the mean-field convergence result from Sec. 3. Sec. 5 discusses related work. Finally, Sec. 6 concludes our paper.

2. GOSSIP PROTOCOLS

Gossip-based protocols (sometimes referred to as epidemic protocols) are appealing in large-scale decentralized systems. In these protocols, nodes exchange data similar to the way a contagious disease spreads. That is, a node can choose with some probability a peer to exchange information with. The gossip concept has originally been proposed for database replication [13].

2.1 A Generic Gossip Protocol

Figure 1 illustrates the skeleton of a generic gossip-based protocol. Each node has a local state s and executes two different threads, an active and a passive one. The active thread periodically initiates a state exchange with a random peer p by sending it a message containing the local state s , after which it waits for a response. The passive thread waits for a message sent by an initiator and replies to it with its local state. The random peer selection is based on the set of neighbours as determined by a membership protocol (e.g., [17]).

```
while true do  
wait ( $\Delta t$  time units)  
 $p \leftarrow \text{RandomPeer}()$ ;  
prepare( $s$ );  
send  $s$  to  $p$ ;  
 $s_p \leftarrow \text{receive}(p)$ ;  
 $s \leftarrow \text{Update}(s, s_p)$ ;
```

(a) active thread (push)

```
while true do  
 $s_p \leftarrow \text{receive}()$ ;  
prepare( $s$ );  
send  $s$  to  $\text{sender}(s_p)$ ;  
 $s \leftarrow \text{Update}(s, s_p)$ ;
```

(b) passive thread (pull)

Figure 1: The skeleton of a gossip protocol

For a pair of nodes A and B , where A is the active node and B is the passive one, we describe the protocol from the point of view of each participating node. In particular, node A picks a neighbouring node B at random (method `RandomPeer()`) after a not necessarily constant time span of length Δt , and initiates the state exchange (gossip) with it. It does so by sending (a part of) its local state s to B , and waits for B 's response. Upon receipt of the response, node A updates its local state (according to the method `Update(s, s_p)`). In response to being contacted by A , node

B sends (part of) its local state to A and updates its local state accordingly (method `Update(s, s_p)`).

Method `Update` is protocol specific. It updates the local state of a node based on the previous local state, and the state information received from the random gossiping partner. In gossip-based information dissemination protocols (as in, e.g., distributed news service protocols [18, 14]), a finite list of data items (e.g., news items), called the cache, composes the local state of a node. The generic operation `prepare(s)` in Figure 1 is replaced by an operation $s \leftarrow \text{RandomItems}()$. The method `Update` merges the list of old items with the list of received items. In gossip-based membership management protocols, a finite set of peer addresses, called the partial view, comprises the local state of a node. The method `Update` (as in [29, 2]) creates a new state through a sample of the union of the old and the received views. In probabilistic broadcasting (e.g., [31]), the state of a node is a flag that records whether the node is infected. Method `Update` sets the state to infected if the received state is infected. In gossip-based distributed aggregation (e.g. [19]), the state of a node is a numeric value, which can be any parameter of the environment, such as a temperature or the current load. All values at nodes contribute to an aggregate value, computed using some aggregation function, for instance, average, sum, etc. The method `Update` simply returns the result of the aggregation function. We refer to [23] for a thorough survey on gossiping applications.

The state exchange between nodes can be implemented as one of the following policies: only the node that initiates a gossip sends (part of) its local state to its partner (push), a node-initiator requests state information from its gossip partner (pull), both nodes send their state information to each other (push-pull).

2.2 Gossiping Time Protocol

Protocols based on epidemic and gossip concepts have found various practical applications [23], including non traditional gossip applications [11], such as gossip-based clock synchronization. The Gossiping Time Protocol (GTP) [16] is a self-managing gossip time synchronization protocol for peer-to-peer networks.

The protocol operates in a network of nodes, each of which equipped with a local clock, and assumes the presence of at least one node with accurate and robust time in the network. Time is disseminated throughout the network by letting nodes periodically gossip their clock settings. That is, each node periodically selects (initiates a gossip with) a random peer from the network to exchange time information with. The nodes subsequently exchange their local settings such that afterwards the node with the worse-quality time has adopted the higher-quality time of the other node. The protocol assumes a presence of a peer-sampling service [17], which allows a node to contact a uniformly randomly selected alive node.

The quality of the time at a node is based on an appropriate metric. In this paper, we consider the hop-count metric that is based on the distance from the time source to the node. The quality of a time sample is given by the number of nodes on the synchronization path from the node to the time source. Therefore, the time source has hop count equal to 0. Node A decides to adopt a clock setting after a timestamp exchange with node B only if its hop count h_A is larger than the hop count of node B plus one, $h_B + 1$.

Then, A sets its hop count to $h_B + 1$. For simplicity, we here assume that nodes only rely on the hop count metric to judge about the quality of clock samples.

Furthermore, each node may decide to adapt a rate (gossiping frequency), at which it initiates a timestamp exchange, based on its local settings. For instance, the better synchronized the node is, the lower the gossiping frequency it may assume. In doing so, the gossiping frequency gradually decreases when the network is synchronized and stable.

Our goal is to show how a mean-field framework can be applied to gossiping protocols. Therefore, we use an instance of the GTP. That is, nodes execute basic GTP based on an immediate clock adjustment model, and change gossiping frequencies, depending on the hop count. However, the version of the protocol that we use for modeling deviates from the original GTP as follows. Firstly, we use the hop count metric to model an abstract notion of time, instead of computing a clock offset (i.e., time difference between two gossiping nodes). Secondly, a node may only be involved in exactly one gossip interaction. Otherwise, a collision takes place and no node participating in the interaction, will update its state. This prevents a node to enhance its hop count by several hops within one gossip interaction.

For the original protocol and its design details, we refer to [15, 16].

3. MEAN-FIELD MODELLING AND CONVERGENCE

This section introduces the theory needed to apply mean field results to gossip protocols. We stay close to the presentation in [10] but change notation when appropriate and simplify things if possible in the gossip context.

3.1 Modelling and Convergence result

A *discrete-time Markov chain* (DTMC) is a stochastic process $\{Y(t) \mid t \in \mathbb{N}\}$ that takes values in a countable state space S . A DTMC obeys the Markov property, that is, the next state is independent of the past, given the present state:

$$\begin{aligned} \Pr\{Y(t+1) = j \mid Y(0) = i_1, \dots, Y(t) = i_t\} &= \\ \Pr\{Y(t+1) = j \mid Y(t) = i_t\}, \quad i_t, j \in S. \end{aligned}$$

We consider a system of $N \in \mathbb{N}$ *interacting objects* that are identically defined. The object with index $n \in \{1, \dots, N\}$ is represented by the discrete-time stochastic process $\{X_n^N(t) \mid t \in \mathbb{N}\}$ which takes values in the set $S = \{0, \dots, K-1\}$ where $K = |S|$ is the number of different states.

Example 1 *In a gossip network, a node is represented by an interacting object. As a running example we consider a simple information dissemination protocol. A piece of information, e.g., the current time, is forwarded through the net. A node can be in one of two states: either it already has the information (state 0) or it is not yet informed (state 1). Hence, the state space for a node is $S = \{0, 1\}$ with $|S| = K = 2$. Let m_0 be a fraction of informed nodes, and $p^N(m_0)$, the probability of moving from state 1 to state 0. Figure 2 shows a graphical representation of the state-transition diagram describing such a node, the possible transitions and their probabilities will be explained later in the text.*

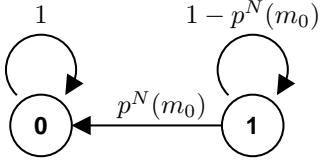


Figure 2: Single node for the information dissemination example

The complete system is composed of the N objects and is, consequently, also described by a discrete-time stochastic process:

$$Y^N(t) = (X_1^N(t), \dots, X_N^N(t)).$$

Its state space is S^N which has $|S|^N$ elements. For the mean-field convergence result we assume that we can not distinguish objects that are in the same state. It then suffices to keep track of the fraction of objects in each state. These fractions are collected in another stochastic process $M^N(t) = (M_0(t), \dots, M_{K-1}(t))$ called the *occupancy measure*. Its elements are defined as

$$M_i^N(t) = \frac{1}{N} \sum_{n=1}^N 1_{\{X_n^N(t)=i\}}, \quad i \in S,$$

where $1_{\{X_n^N(t)=i\}}$ is 1 if $X_n^N(t) = i$ and 0 otherwise. Its state space $S_M^N \subset \mathbb{R}^K$ has

$$|S_M^N| = \binom{K+N-1}{K-1}$$

elements (the number of possibilities to distribute N objects over the K states they can be in). One state from this state space is denoted $\mathbf{m} = (m_0, m_1, \dots, m_{K-1}) \in S_M^N$, where m_i is a fraction of nodes in the state i .

Example 2 For the information dissemination example, the state space of the occupancy measure is

$$S_M^N = \left\{ \left(\frac{k}{N}, 1 - \frac{k}{N} \right) \mid k \in \{0, \dots, N\} \right\}.$$

Its size is

$$|S_M^N| = \binom{2+N-1}{2-1} = N+1.$$

The evolution of the system of interacting objects is described by the *local* transition probabilities of each object. The next state of any object does not only depend on the current state of the object *but also* on the current occupancy measure \mathbf{m} :

$$P_{i,j}^N(\mathbf{m}) = \Pr\{X_n^N(t+1) = j \mid X_n^N(t) = i, M^N(t) = \mathbf{m}\}, \\ i, j \in S, \mathbf{m} \in S_M^N.$$

These probabilities are the same for all objects. They are gathered into the transition probability matrix $P^N(\mathbf{m})$. These local transition probabilities determine the unique transition probability matrix for the global system $Y^N(t)$, which is a DTMC because its next state (occupancy measure) only depends on the current state (occupancy measure).

Example 3 A node can only move from being uninformed (state 1) to being informed (state 0). Afterwards it stays in state 0 forever, that is, it never forgets. Suppose that in each time step a node A initiates a gossip interaction with probability g . It randomly chooses a partner node B among the $N-1$ other nodes. If B is already informed and A is not, A moves to state 0, so that we model a simple pull protocol. Note that m_0 is the fraction of informed nodes in the system and $m_1 = 1 - m_0$ the fraction of uninformed nodes. The total probability for moving from state 1 to state 0 equals

$$p^N(m_0) = P_{1,0}^N((m_0, m_1)) = g \cdot \frac{m_0 \cdot N}{N-1}.$$

Here, $m_0 \cdot N$ is the number of informed nodes and $m_0 \cdot N / (N-1)$ is the probability that a node chooses an informed node out of the $N-1$ possible nodes (it does not pick itself) as gossip partner. The complete probability matrix is then given by

$$P^N((m_0, m_1)) = \begin{pmatrix} 1 & 0 \\ p^N(m_0) & 1 - p^N(m_0) \end{pmatrix}.$$

For the global system, the probability to move from a fraction of m_0 informed nodes to m'_0 informed nodes, for $m'_0 \geq m_0$, equals

$$\binom{m_1 \cdot N}{(m'_0 - m_0) \cdot N} (p^N(m_0))^{(m'_0 - m_0)N} (1 - p^N(m_0))^{m'_1 N},$$

where $m_1 = 1 - m_0$, $m'_1 = 1 - m'_0$. This binomial expression is composed of the number of possibilities to choose exactly the “missing” $(m'_0 - m_0) \cdot N$ objects out of the $m_1 \cdot N$ uninformed nodes, these then all have to take the transition to state 0, and all other $m'_1 \cdot N$ nodes remain in state 1.

Consider now the occupancy measure $M^N(t)$ of the system at a given finite time $t \in \mathbb{N}$. Recall that $M^N(t)$ is a random variable. For a given initial occupancy measure \mathbf{m}_0^N , there are two ways to determine the distribution of $M^N(t)$: first, we can calculate the transient distribution analytically at time t , requiring t vector-matrix multiplications with a vector of size $|S_M^N|$. Second, we can employ discrete-event simulation to estimate the distribution. Often only discrete-event simulation is possible since, for large N , the size of the state space makes the analytical computation of the transient probabilities practically infeasible. But even discrete-event simulation of this large DTMC is expensive.

Example 4 Figure 3 shows the analytically (using mean-field analysis) computed distribution of the fraction of informed nodes at time $t = 10$, for the initial occupancy measure $M^N(0) = (0.01, 0.99)$. Note that the distribution is “more deterministic” for larger N .

We also simulated this simple dissemination protocol in a round-based fashion similar to simulations in PeerSim [20]. Using 1000 independent runs for each curve, the resulting distributions for $M_0^N(10)$ are shown in Figure 4, together with the corresponding analytical distributions. The curves roughly coincide, however, the simulated distributions lie always below the analytical ones, that means, the assumed values are, in fact, higher.

In the simulations all nodes proceed in a lock-step fashion within a round, whereas the DTMC always envisions a transition to the next discrete time-step as an atomic step. In round-based simulations, for instance, node A can “inform” node B , which in turn can “inform” node C . This

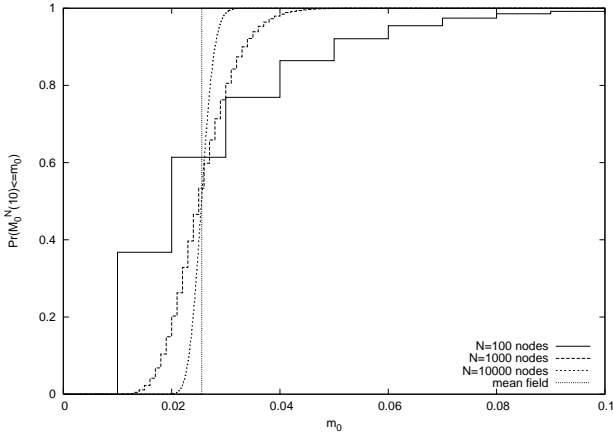


Figure 3: Distribution of $M_0^N(10)$ for $M^N(0) = (0.01, 0.99)$

scenario is not possible in the DTMC, because if A “informs” B, C cannot be “informed” by B, since at the beginning of the step, B does not have that information yet. This decreases the chance of C to get the information. Consequently, the probability to have a high number of informed nodes is smaller for the DTMC than for simulation. Hence, the simulated distributions in Figure 4 are always a bit lower than the ones of the DTMC.

At this point, the so-called mean-field convergence result applies. It captures the limiting behaviour of the complete system if the number of objects N goes to infinity and so provides an approximation for the occupancy measure for large N . The requirement is that for all local states $i, j \in S$, all $\mathbf{m} \in \mathbb{R}^K$ and for $N \rightarrow \infty$

$P_{i,j}^N(\mathbf{m})$ converges uniformly¹ in \mathbf{m} to some $P_{i,j}(\mathbf{m})$, which is a continuous function of \mathbf{m} .

If this requirement is satisfied, the occupancy measure converges almost surely to a deterministic limit. This means that for each local state i the fraction $M_i^N(t)$ of objects in state i at time t is known with probability one.

Theorem 1 (cf. [10]) Fix the initial occupancy measure to be identical for all $N \in \mathbb{N}$:

$$M^N(0) = \mu(0).$$

Define the limit of the local probability matrix:

$$P(\mathbf{m}) = \lim_{N \rightarrow \infty} P^N(\mathbf{m}), \quad \mathbf{m} \in \mathbb{R}^K.$$

Define the deterministic process

$$\mu(t+1) = \mu(t) \cdot P(\mu(t)).$$

Then for any $t \in \mathbb{N}$,

$$\lim_{N \rightarrow \infty} M^N(t) = \mu(t), \quad \text{with probability 1,}$$

that is, $\mu(t)$ is the limit occupancy measure for $N \rightarrow \infty$.

¹A sequence f_N of real valued functions converges uniformly with limit f if for every $\varepsilon > 0$ there exists a natural number n such that for all x and all $N \geq n$ we have $|f_N(x) - f(x)| < \varepsilon$.

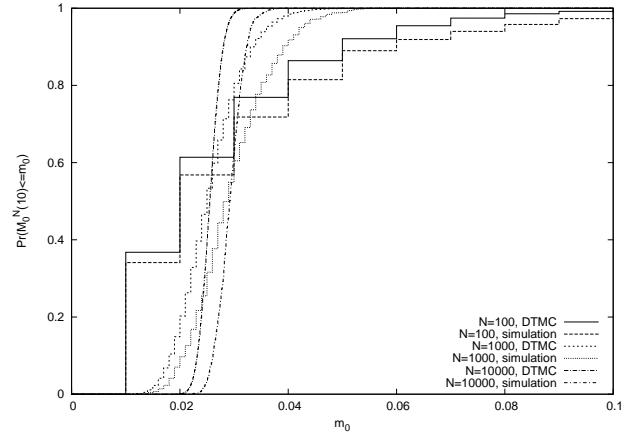


Figure 4: Distribution of $M_0^N(10)$ computed using simulation

For large N we can now approximate the stochastic process for the occupancy measure by a deterministic process.

Example 5 The limit of the probability to move from state 1 to state 0 is

$$p(m_0) = \lim_{N \rightarrow \infty} g \cdot \frac{m_0 \cdot N}{N-1} = g \cdot m_0,$$

which is continuous in m_0 . The requirement for the application of the mean-field convergence result is thus satisfied. If we set $\mu(0) = (0.01, 0.99)$, the deterministic limit for time $t = 10$ is

$$\mu(10) = (0.0256, 0.9744)$$

computed by ten matrix-vector multiplications. It is indicated by the vertical line for m_0 in Figure 3.

3.2 A Methodology for the Mean-field Analysis of Gossip Protocols

We summarise how mean-field analysis can be used for the performance evaluation of gossiping protocols. Our methodology consists of the following steps:

Step 1: Formal description.

The formal specification of a system helps to obtain not only a better (more modular) description, but also a clear understanding and an abstract view of the system. In general, it is hard to give a full specification of a system or protocol under study. Such a study is usually done on a simplified system model of the actual protocol: one has to decide which characteristics of the protocol should be studied, and which parameters of the protocol should be modelled in order to study these characteristics. In order to simplify the system model, assumptions should be made. These assumptions should be supported by experimental study.

Step 2: Identification of local states and transitions.

This step requires to identify the set S of local states of a node. The states should reflect all relevant situations a node can be in. Transitions between local states usually occur because of gossip interactions.

Step 3: Transition probabilities.

The (local) transition probabilities depend on the global state of the gossip network model. These probabilities have to be investigated thoroughly. A node might also behave intrinsically in a probabilistic way. At the end of this step stands a directive of how to calculate the transition probability matrix depending on the current global state.

Step 4: Mean-field convergence requirements.

Only if the local transition probabilities converge appropriately for $N \rightarrow \infty$ we can successfully apply the mean-field convergence theorem.

Step 5: Mean-field limit.

Finally, we can compute the mean-field limit for our model using the computation from Theorem 1. With the obtained results we can test and compare different designs.

4. APPLICATION

In this section we present a mean-field model for the GTP, where samples are evaluated using the hop count metric. The aim of our presentation is to illustrate the usefulness of mean-field analysis for the formal evaluation of gossip protocols. The shown model is not intended to be accurate image of reality, but merely to illustrate the mean-field method.

4.1 State Space

The state of a node in a GTP network is given by the number of hops its timing information has travelled from the time source. There is an additional state indicator for nodes that have not yet received any time sample. If there are N nodes the maximum possible hop count is $N - 1$. We therefore set the state space of a node to be $\{0, \dots, N - 1, N\}$. While states 0 to $N - 1$ denote actual hop counts, state N reflects the completely unsynchronised situation.

Note that the state space of a node is of size $N + 1$. However, assuming that at the beginning ($t = 0$) all nodes are either time sources (hop count 0) or unsynchronised (hop count N), at time $t > 0$ the hop count is also limited by t . As long as $t < N$, which typically will be the case as we are interested in large N , we can take as state space

$$S = \{0, \dots, t, N\}.$$

The number of considered states of S is then $K = t + 2$.

4.2 Local Transition Probabilities

In each discrete time step a node initiates a gossip with a probability depending on the hop count of its current timing information. We want the gossip probability to increase with the hop count to accelerate the receipt of a “good” time stamp. Nodes with an already low hop count contribute less to network load by initiating few gossip interactions. If a node has hop count $i > 0$, we set the gossip probability to

$$g_i = g + (1 - g) \cdot \left(1 - e^{-a \cdot (i-1)}\right),$$

and $g_0 = g$. Here, g is the basic gossip probability and $a \geq 0$ is a parameter ruling the influence of the second term. If $a = 0$, all nodes gossip with the same probability g . If $a > 0$, a node with hop count 1 gossips with probability g and for $i \rightarrow \infty$ the gossip probability goes to 1. Thus,

the gossip probability increases with the current hop count. The function g_i has been chosen for convenience giving a nice limit for $i \rightarrow \infty$. The function increases with hop count i and does not depend on the network size N , because nodes do not have a global knowledge, such as the network size.

In our model, the hop count of a node A can only decrease, never increase. If A has hop count i , the probability to get hop count $1 \leq j < i$ in the next step is composed of several parts.

- First, A itself can initiate a gossip interaction. To get new hop count j , it has to choose a node B with hop count $j - 1$. Node B is not gossiping itself, otherwise there would be a collision of gossip interactions and no valid time stamp exchange can take place. For the current occupancy measure \mathbf{m} , the probability for this type of event is

$$\text{pull}_{i,j}^N(\mathbf{m}) = m_{j-1} \cdot N \cdot g_i \cdot \frac{1}{N-1} \cdot (1 - g_{j-1})$$

where m_{j-1} is the fraction of nodes with hop count $j - 1$ and $(m_{j-1}N)/(N - 1)$ is the probability to pick a $j - 1$ node.

- Second, node A can be contacted by B , one of the $m_{j-1} \cdot N$ nodes with hop count $j - 1$, without gossiping itself. The probability for this case is

$$\text{push}_{i,j}^N(\mathbf{m}) = m_{j-1} \cdot N \cdot g_{j-1} \cdot \frac{1}{N-1} \cdot (1 - g_i)$$

In both cases, to avoid collisions we require that no other node is allowed to initiate a gossip interaction with either node A or node B . Assume that node A has hop count i (i.e. $m_i > 0$) and node B has hop count $j - 1$ (i.e. $m_{j-1} > 0$). With $\text{noc}_{i,j}^N(\mathbf{m})$ (no collision) we denote the probability that there is no collision when nodes A and B interact. For the probability that there is no interaction with either A or B , we multiply the probability of each node different from A and B not to contact either node A or B .

$$\begin{aligned} \text{noc}_{i,j}^N(\mathbf{m}) &= \prod_{\substack{k \in S \\ k \neq i, \\ k \neq j-1}} \left[\underbrace{(1 - g_k)}_{\text{no gossip}} + \underbrace{g_k \cdot \frac{N-3}{N-1}}_{\text{gossip with other node than } A \text{ or } B} \right]^{m_k N} \\ &\quad \cdot \left[(1 - g_i) + g_i \cdot \frac{N-3}{N-1} \right]^{m_i N-1} \\ &\quad \cdot \left[(1 - g_{j-1}) + g_{j-1} \cdot \frac{N-3}{N-1} \right]^{m_{j-1} N-1} \\ &= \prod_{\substack{k \in S \\ k \neq i, \\ k \neq j-1}} \left[1 - \frac{2g_k}{N-1} \right]^{m_k N} \cdot \left[1 - \frac{2g_i}{N-1} \right]^{m_i N-1} \\ &\quad \cdot \left[1 - \frac{2g_{j-1}}{N-1} \right]^{m_{j-1} N-1} \end{aligned}$$

That is, there are three cases: (1) all $m_k \cdot N$ nodes with hop count k ($k \neq i$ and $k \neq j$) either do not gossip or gossip with other nodes but A and B ; (2) all nodes with hop count i but node A (i.e., $m_i N - 1$ nodes) either do not gossip or gossip with other nodes but A and B ; (3) all nodes with hop

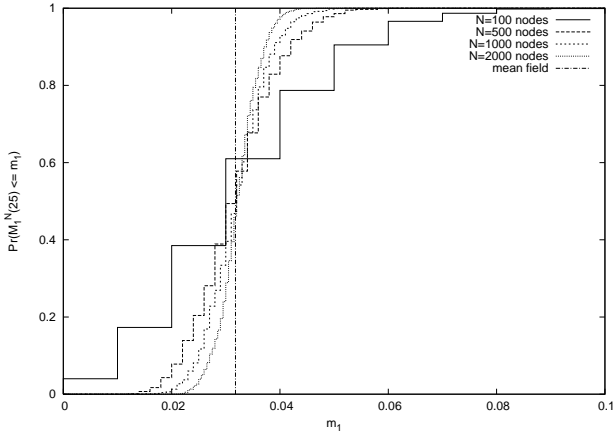


Figure 5: Distribution of $M_1^N(25)$ for different N computed using DTMC simulation, as well as mean-field analysis

count $j - 1$ but node B (i.e., $m_{j-1}N - 1$ nodes) either do not gossip or gossip with other nodes but A and B .

Combining the above terms, the complete probability for a node with hop count i to become a node with hop count $j < i$ in the next step equals

$$P_{i,j}^N(\mathbf{m}) = \left(\text{push}_{i,j}^N(\mathbf{m}) + \text{pull}_{i,j}^N(\mathbf{m}) \right) \cdot \text{noc}_{i,j}^N(\mathbf{m}).$$

The probability not to change the hop count in the next step equals

$$P_{i,i}^N(\mathbf{m}) = 1 - \sum_{j=1}^{i-1} P_{i,j}^N(\mathbf{m}).$$

If $j = 0$ or $j > i$, $P_{i,j}^N(\mathbf{m}) = 0$.

4.3 Global DTMC

The probability matrix of the DTMC of the complete system with all nodes is not easily derived. However, a discrete-event simulation of the DTMC is possible even when only knowing the local probabilities. In each step, for all nodes the new state has to be determined individually.

In Figure 5 we depict the distribution of $M_1^N(25)$, i.e., the fraction of states that have hop count one at time $t = 25$ for different N . In the initial state, 1% of the nodes are time sources (hop count 0) and 99% are unsynchronised (hop count N). The base gossip probability is chosen to be $g = 0.1$, and we select $a = 0.1$. The distributions are derived experimentally using 1000 simulation runs for each curve. Note that the distribution show less variance for larger N .

We have implemented the simulation of a single time step in two stages: first we decide for each node whether it initiates a gossip interaction or not, based on its gossip probability and, if so, choose the partner node. Second, we update all nodes that move to a lower hop count because they gossip exclusively with an appropriate partner.

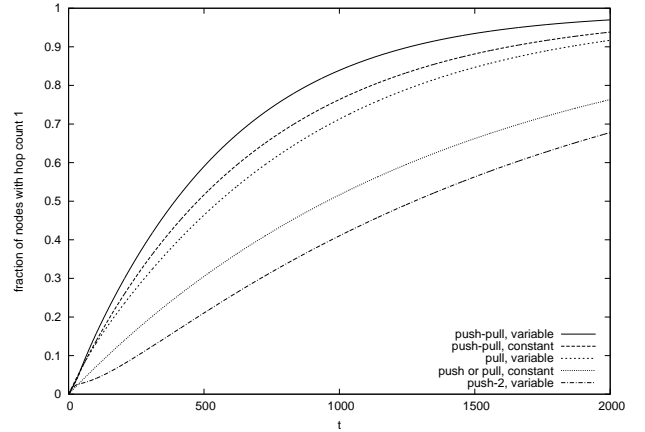


Figure 6: Fraction of nodes with hop count 1 over time computed using mean field analysis and protocol simulation

4.4 Mean-Field Limits

For $N \rightarrow \infty$ we have the following limiting probabilities:

$$\begin{aligned} \text{pull}_{i,j}(\mathbf{m}) &= \lim_{N \rightarrow \infty} \text{pull}_{i,j}^N(\mathbf{m}) = g_i \cdot (1 - g_{j-1}) \cdot m_{j-1} \\ \text{push}_{i,j}(\mathbf{m}) &= \lim_{N \rightarrow \infty} \text{push}_{i,j}^N(\mathbf{m}) = g_{j-1} \cdot (1 - g_i) \cdot m_{j-1} \\ \text{noc}_{i,j}(\mathbf{m}) &= \lim_{N \rightarrow \infty} \text{noc}_{i,j}^N(\mathbf{m}) = e^{-\sum_{k=0}^{\infty} m_k \cdot g_k}. \end{aligned}$$

The sum in the exponent in the last expression always converges since all $m_k, g_k \in [0, 1]$, and $\sum_k m_k = 1$. Using the mean-field convergence result (Theorem 1) we can compute the limiting fraction of nodes with hop count one at time $t = 25$. For the initial occupancy measure $\mu(0) = (0.01, \dots, 0.99)$, we have $\mu(25) = (0.01, 0.0318, \dots, 0.8336)$.

In contrast to the expensive simulations over 25 time steps, the mean-field result only requires 25 vector matrix multiplications, where the vector has 27 entries and the matrix of size 27×27 is recalculated in each time step. Measures for much higher time horizons can so be computed.

As an example, Figure 6 shows the evolution of the fraction of nodes with hop count one up to time $t = 2000$ for various protocol settings. Protocols (\cdot ,variable) use variable gossip probability ($a = 0.1 > 0$), protocols (\cdot ,constant) use constant gossip probability ($a = 0$). The basic gossip probability is $g = 0.1$ for all variants. Initial occupancy measure is again $\mu(0) = (0.01, \dots, 0.99)$. The protocols (push-pull, \cdot) implement the push-pull version of the GTP protocol: in a successful gossip interaction, both nodes send their respective time samples and the one with the higher hop count updates its local timer. The protocols (pull, \cdot) implement the pull version of the GTP protocol. Here only the initiator of a gossip interaction might update its timer, which is reflected by the fact that for the calculation of the transition probabilities, the term $\text{push}(\mathbf{m})$ is omitted. In contrast, in the protocols (push, \cdot) only the passive partner of a gossip interaction can enhance its timing information.

All curves show an increase of the fraction of nodes with hop count one that is steeper at the beginning and levels out when the maximal value is approached. The maximum is 99% since 1% of the nodes are time sources with hop count zero. The push-pull variant with variable hop count

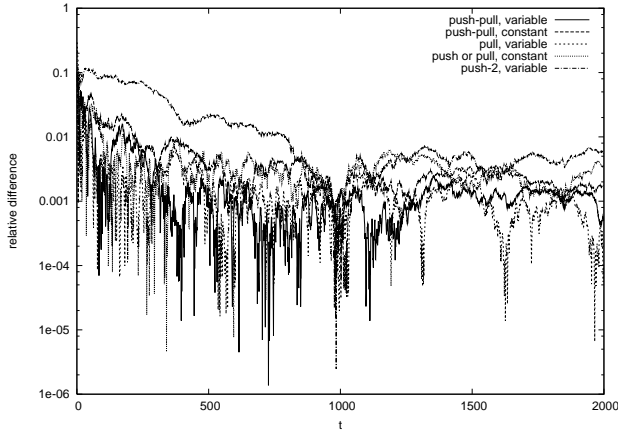


Figure 7: Relative difference between mean-field and simulation results.

exhibits the fastest growth. For constant gossip probability the fraction of nodes grows a bit slower, because all nodes are bound to the relatively small gossip probability $g = 0.1$. When restricting the protocol to the pull-only variant, the fraction of nodes is even smaller since fewer successful interactions occur. When using constant gossip probabilities, the curves of the push and the pull variant coincide because the two expressions $push(\mathbf{m})$ and $pull(\mathbf{m})$ are identical if $g_i = g_{j-1} = g$. When using the push variant with constant gossip probability, the expression

$$push_{\infty,j}(\mathbf{m}) = g_{j-1}(1 - g_{\infty})m_{j-1} = g_{j-1}(1 - 1)m_{j-1} = 0$$

equals zero and so no unsynchronised node can ever get a finite hop count. Thus, the fraction of nodes with hop count one remains zero forever. Actually, any gossip initiation of unsynchronised nodes is useless in this setting, since it will never lead to a time update. If we enhance this protocol version by setting $g_{\infty} = 0$, the curve labelled (push-2, variable) results. However, increasing the gossip probability with the hop count makes in general no sense for the push variant. A fair comparison would require that the gossip probability *decreases* with the hop count.

We also simulated the different variants of the protocol in a round-based fashion using $N = 10000$ nodes. The curves match so closely that we do not include them in the graph. In Figure 7 we show the relative difference between mean-field and simulation results w.r.t. the mean-field values. The relative difference is in most cases below 1% (note the log y -scale). For the protocols with constant gossip probability, the mean-field results are in general closer to the simulation results than for variable gossip probabilities.

4.5 Optimal Gossip Probability

The gossip probability $g = 0.1$ used for the examples so far has been chosen arbitrarily. An obvious question is to ask for the optimal gossip probability. Small probabilities imply fewer gossip interaction attempts but also lead to only a few collisions. Higher values for g result in more attempts but also in more collisions. Figure 8 shows the fraction of nodes with hop count one at time $t = 25$, for the gossip probability ranging over $[0, 1]$. Obviously, if no node gossips at all ($g = 0$) or all nodes always gossip ($g = 1$), no node can ever have hop count one because there are no successful

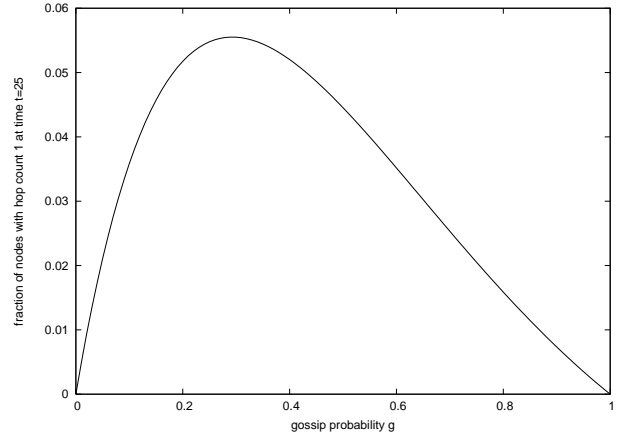


Figure 8: Performance for different constant gossip probabilities g

gossip interactions. The optimal value lies at approximately $g = 0.3$. Further mean field computations suggest that this is the optimal value for all times t .

4.6 Enhancing the Model

The GTP also incorporates the evaluation of the clock sample quality based on the last sample used by a node for synchronization (dispersion metric). It is possible to model the GTP with its detailed time synchronization method, but this requires the use of a so-called global memory in mean-field analysis [10], which we have not presented in this paper. This extension affects the state space, as well as the transition probabilities.

5. RELATED WORK

The notion of “mean-field” is often used in the literature, with different meanings. The mean-field concept was first introduced in physics. It has been used in the context of Markov chain models of systems like plasma and dense gases where the strength of the interaction between particles is inversely proportional to the size of the system. A particle is seen as under a collective force generated by the other particles in a continuous time and space setting. In the area of communication networks, mean-field convergence results have been applied in various forms to a variety of case studies, including TCP connections [6, 4, 3, 28], HTTP flows [5], bandwidth sharing [24], transportation networks [1], swarm robotic systems [25], reputation determination [10], queueing networks [12, 22, 30], and Internet congestion control [21].

We are not familiar with prior work dealing with mean-field theory for the evaluation of gossip protocols. Previous work on gossip protocols has used a notion of mean-value and infinite limit (when the number of nodes $N \rightarrow \infty$) to simplify computation for their analysis. Notably in [9], Bonnet studied the evolution of the in-degree distribution of nodes executing the Cyclon protocol [29]. The states of the associated Markov chain represent the fraction of nodes with a specific in-degree distribution. From the designed Markov chain he determined the distribution to which the protocol converges. The author showed that the system converges by constructing a generating function, a series whose co-

efficients encode the in-degree distribution. The generating function then enabled algebraic means to compute the mean value and the standard deviation of the stationary distribution.

Allavena et al. [2] proposed a gossip-based membership management protocol and analysed the evolution of the number of links between two nodes executing the protocol. The states of the associated Markov chain are given by the numbers of links between a pair of nodes. From the designed Markov chain they calculated the expected time until a network partition occurs. Their goal is to show an effect of the mixing of both pull and push approaches. Citing [2, Sect. 4.1.1]: “The model is obtained from a slightly modified version of the completely unsynchronised protocol further simplified by some sort of mean-field approximation.” However, there are no further details related to mean-field approximation in the paper.

Stojanovic et al. [27] analysed and compared delay performance of network coding and cooperative diversity in a single-hop wireless network. The authors performed an asymptotic analysis (for the number of nodes $N \rightarrow \infty$) of the expected delay associated with the broadcasting of a file consisting of a certain amount of packets.

6. CONCLUSION

The main motivation for developing a modelling methodology for gossip protocols is that, although these protocols are appealing with respect to scalability, robustness, and individual simplicity, they do not provide us with a way to quantitatively predict the performance according to a particular metric or analyse further possible optimizations and limitations analytically.

We have demonstrated that mean-field analysis is suitable for gossip protocols. The following premises enable mean-field analysis:

- there is a very large number of identically behaving nodes (symmetry property [7]);
- there are no central servers or global resources;
- the behaviour of a single node can be described in a local way;
- the number of states a node can be in is small in comparison to the number of nodes;
- transient measures (“at time t ”) are to be computed.

Extensions of the theory presented here would also allow for the incorporation of a global memory, the failure or entering/leaving of nodes [10], the employment of continuous-time models, and steady-state measures [8]. However, the mean-field approach does not allow for the evaluation of a centrally managed network, the distinguished modelling of one single node or the inclusion of topographic information on the network.

We have considered two applications of gossip, an information dissemination and a distributed aggregation. We first explained our methodology on a simple model of the GTP, that follows a timestamp exchange strategy similar to pull-based information dissemination. After that, the adapted version of the GTP has been used as a more sophisticated example for mean-field modelling. This model has very “natural” assumptions for large networks, including possible collision

during gossip interactions, and a gossiping frequency that adjusts according to the node state. Our modelling results are confirmed by large-scale simulations.

We also addressed the issue of finding the optimal constant gossip probability. We did this manually by computing results for different g and choosing the best value.

As for future work, we plan to investigate mean-field analysis for alternative stochastic models for the nodes, e.g., by moving to the continuous-time context or by introducing non-determinism using Markov decision processes [26].

7. REFERENCES

- [1] L. Afanassieva, S. Popov, and G. Fayolle. Models for transportation networks. *J. of Math. Sciences*, 84(3):1092–1103, 1997.
- [2] A. Allavena, A. Demers, and J. Hopcroft. Correctness of a gossip based membership protocol. In *Proc. 24th Annual ACM Symp. on Principles of Distributed Computing (PODC’05)*, pages 292–301. ACM Press, 2005.
- [3] F. Baccelli, A. Chaintreau, D. De Vleeschauwer, and D. McDonald. A mean-field analysis of short lived interacting TCP flows. *ACM SIGMETRICS Perform. Eval. Rev.*, 32(1):343–354, 2004.
- [4] F. Baccelli, A. Chaintreau, D. De Vleeschauwer, and D. McDonald. HTTP turbulence. *AMS Networks and Heterogeneous Media*, 1:1–40, 2006.
- [5] F. Baccelli, D. McDonald, and M. Lelarge. Metastable regimes for multiplexed TCP flows. In *42nd Allerton Conf. on Communication, Control, and Computing*, 2004.
- [6] F. Baccelli, D. McDonald, and J. Reynier. A mean-field model for multiple TCP connections through a buffer implementing RED. *Perform. Eval.*, 49(1-4):77–97, 2002.
- [7] R. Bakhshi, F. Bonnet, W. Fokkink, and B. Haverkort. Formal analysis techniques for gossiping protocols. *ACM SIGOPS Oper. Syst. Rev.*, 41(5):28–36, 2007.
- [8] A. Bobbio, M. Griboaud, and M. Telek. Analysis of large scale interacting systems by mean field method. In *Proc. 5th Conf. on the Quantitative Evaluation of Systems (QEST’08)*, pages 215–224. IEEE Computer Society, 2008.
- [9] F. Bonnet. Performance analysis of Cyclon, an inexpensive membership management for unstructured P2P overlays. Master thesis, ENS Cachan Bretagne, University of Rennes, IRISA, 2006.
- [10] J.-Y. Le Boudec, D. McDonald, and J. Munding. A generic mean field convergence result for systems of interacting objects. In *Proc. 4th Conf. on the Quantitative Evaluation of Systems (QEST’07)*, pages 3–18. IEEE Computer Society, 2007.
- [11] P. Costa, V. Gramoli, M. Jelasity, G. P. Jesi, E. Le Merrer, A. Montresor, and L. Querzoni. Exploring the interdisciplinary connections of gossip-based systems. *ACM SIGOPS Oper. Syst. Rev.*, 41(5):51–60, 2007.
- [12] D. Dawson, J. Tang, and Y. Zhao. Balancing queues by mean field interaction. *Queueing Syst. Theory & Appl.*, 49(3-4):335–361, 2005.
- [13] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database

- maintenance. In *Proc. 6th Annual ACM Symp. on Principles of Distributed Computing (PODC'87)*, pages 1–12. ACM Press, 1987.
- [14] D. Gavidia, S. Voulgaris, and M. van Steen. A Gossip-based Distributed News Service for Wireless Mesh Networks. In *Proc. 3rd IEEE Conf. on Wireless On demand Network Syst. and Services (WONS'06)*, pages 59–67. IEEE Computer Society, 2006.
- [15] K. Iwanicki. Gossip-based dissemination of time. Master's thesis, Warsaw University and Vrije Universiteit Amsterdam, May 2005.
- [16] K. Iwanicki, M. van Steen, and S. Voulgaris. Gossip-based clock synchronization for large decentralized systems. In *Proc. 2nd Workshop on Self-Managed Networks, Systems and Services (SelfMan 2006)*, volume 3996 of *LNCS*, pages 28–42. Springer, 2006.
- [17] M. Jelasity, R. Guerraoui, A.-M. Kermarrec, and M. van Steen. The peer sampling service: Experimental evaluation of unstructured gossip-based implementations. In *Proc. 5th ACM/IFIP/USENIX Int. Middleware Conf. (Middleware'04)*, volume 3231 of *LNCS*, pages 79–98. Springer, 2004.
- [18] M. Jelasity, W. Kowalczyk, and M. van Steen. Newscast computing. Technical Report IR-CS-006, Vrije Universiteit Amsterdam, Department of Computer Science, Amsterdam, The Netherlands., November 2003.
- [19] M. Jelasity, A. Montresor, and O. Babaoglu. Gossip-based aggregation in large dynamic networks. *ACM Trans. Comput. Syst.*, 23(3):219–252, 2005.
- [20] M. Jelasity, A. Montresor, G. P. Jesi, and S. Voulgaris. PeerSim: A peer-to-peer simulator. <http://peersim.sourceforge.net/>.
- [21] W. Kang, F. Kelly, N. Lee, and R. Williams. Fluid and Brownian approximations for an internet congestion control model. In *Proc. 43rd IEEE Conf. on Decision and Control (CDC'04)*, volume 4, pages 3938–3943, 2004.
- [22] F. Karpelevich, E. Pechersky, and Y. Suhov. Dobrushin's approach to queueing network theory. *J. of Applied Mathematics and Stochastic Analysis*, 9(4):373–397, 1996.
- [23] A.-M. Kermarrec and M. van Steen. Gossiping in distributed systems. *ACM SIGOPS Oper. Syst. Rev.*, 41(5):2–7, 2007.
- [24] S. Kumar and L. Massoulié. Integrating streaming and file-transfer internet traffic: fluid and diffusion approximations. *Queueing Syst. Theory Appl.*, 55(4):195–205, 2007.
- [25] A. Martinoli, K. Easton, and W. Agassounon. Modeling Swarm Robotic Systems: A Case Study in Collaborative Distributed Manipulation. *Int. Journal of Robotics Research*, 23(4):415–436, 2004. Special Issue on Experimental Robotics, B. Siciliano, editor.
- [26] M. Puterman. *Markov Decision Processes*. Wiley, 1994.
- [27] I. Stojanovic, M. Sharif, and D. Starobinski. Data dissemination in wireless broadcast channels: Network coding r cooperation. In *Proc. 41st Conf. on Information Sciences and Systems (CISS '07)*, pages 265–270. IEEE Press, 2007.
- [28] P. Tinnakornrisuphap and A. Makowski. Limit behavior of ECN/RED gateways under a large number of TCP flows. In *Proc. Conf. of the IEEE Computer and Communications Societies*, volume 2, pages 873–883. IEEE Computer Society, 2003.
- [29] S. Voulgaris, D. Gavidia, and M. van Steen. Cyclon: Inexpensive membership management for unstructured P2P overlays. *J. Network and Syst. Manage.*, 13(2):197–217, 2005.
- [30] N. Vvedenskaya and Y. Suhov. Dobrushin's mean-field approximation for a queue with dynamic routing. *Markov Proc. Rel. Fields*, 3(4):493–526, 1997.
- [31] Q. Zhang and D. Agrawal. Dynamic probabilistic broadcasting in MANETs. *J. of Parallel and Distributed Computing*, 65(2):220–233, 2005.