

Greedy Gossip with Eavesdropping

Deniz Üstebay, Mark Coates, and Michael Rabbat

Department of Electrical and Computer Engineering

McGill University, Montréal, Québec, Canada

Email: deniz.ustebay@mail.mcgill.ca, {mark.coates, michael.rabbat}@mcgill.ca

Abstract—This paper presents *greedy gossip with eavesdropping* (GGE), a new average consensus algorithm for wireless sensor network applications. Consensus algorithms have recently received much attention in the sensor network community because of their simplicity and completely decentralized nature which makes them robust to changes in the network topology and unreliable wireless networking environments. In the sensor network, each node has a measurement value and the aim of average consensus is computing the average of these node values in the absence of a central authority. We prove that GGE converges to the average consensus with probability one. We also illustrate the performance of the algorithm via simulations and conclude that GGE provides a significant performance improvement compared to existing average consensus algorithms such as randomized gossip and geographic gossip.

I. INTRODUCTION AND BACKGROUND

Efficiently reaching a consensus is a fundamental problem in decentralized systems such as wireless sensor-actuator networks. The prototypical example of a consensus problem is that of computing the *average consensus*: initially, each node has a scalar piece of information, y_i , and the goal is to compute the average, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, at every node in the network. Consensus can be viewed as a sort of synchronization or agreement, before the network makes a concerted action. The average consensus is also useful for distributed decision making under a Gaussian observation model. Our previous work examined the use of average consensus algorithms in distributed signal processing applications such as source localization [7] and data compression [6].

In this paper we propose a new average consensus algorithm, *greedy gossip with eavesdropping*, that takes advantage of the broadcast nature of wireless communications to accelerate convergence. Motivated by wireless sensor-actuator network applications, we assume the network is composed of battery-powered nodes, communicating via wireless radios. Since each wireless transmission consumes valuable energy resources, our goal is to reach the average consensus in as few transmissions as possible. We assume a broadcast model where all nodes within range of a transmitting node (the neighbors) successfully receive the message.

The two most widely studied algorithms for solving the average consensus problem are *distributed averaging* [8] and *randomized gossip* [3]. In distributed averaging, every node broadcasts information to its neighbors at every iteration. Let $x_i(k)$ denote the value at node i after the k th iteration. Each node i initializes its value to $x_i(0) = y_i$. At the k th iteration, after node i receives values $x_j(k-1)$ from each of its

neighbors, it replaces $x_i(k)$ with a weighted average of its own previous value and its neighbors' values. Under appropriate conditions on the weights used in the update step, one can show that the values $x_i(k)$ at every node converge to the average \bar{y} as $k \rightarrow \infty$ [8]. However, information diffuses slowly across the network in this scheme, and since the information at each node typically does not change much from iteration to iteration, this is not efficient use of the broadcast medium.

Randomized gossip operates at the opposite extreme, where only two neighboring nodes exchange information at each iteration. At the k th iteration, a node s is chosen uniformly at random; it chooses a neighbor, t , randomly; and this pair of nodes “gossips”: s and t exchange values and perform the update $x_s(k) = x_t(k) = (x_s(k-1) + x_t(k-1))/2$, and all other nodes remain unchanged. Again, one can show that under very mild conditions on the way a random neighbor, t , is drawn, the values $x_i(k)$ converge to \bar{y} at every node [8]. Although other neighbors overhear the messages exchanged between the active pair of nodes, they do not make use of this information in existing randomized gossip algorithms. The fact that nodes only exchange information with their immediate neighbors is attractive, from the point of view of simplicity and robustness to changing topologies and/or network conditions. However it also means that in typical wireless network topologies (grids or random geometric graphs), information diffuses slowly across the network. Boyd et al. prove that for random geometric graphs, randomized gossip requires $O(n^2)$ transmissions to approximate the average consensus well.

Slow convergence of randomized gossip motivated Dimakis et al. to develop geographic gossip. Assuming each node knows its geographic location and the locations of its neighbors, information can be exchanged with nodes beyond immediate neighbors. In [4], they show that these long-range transmissions improve the rate of convergence from $O(n^2)$ to roughly $O(n^{3/2})$ transmissions. Although geographic gossip is a significant improvement over randomized gossip in terms of number of transmissions, it comes at the cost of increased complexity, since the network must now provide reliable two-way transmission over many hops. Messages which are lost in transit potentially result in biasing the average consensus computation.

Unlike randomized gossip and geographic gossip, where the pair of nodes exchanging information at each iteration is drawn independently and identically from iteration to iteration, in GGE a node greedily chooses which neighbor to gossip with at each iteration, based on which update will result in the

largest decrease in local error. In order to perform the greedy update, each node must know the values at its neighbors. This is accomplished by exploiting the broadcast nature of wireless communication and having each node eavesdrop on its neighbors' transmissions. We show that making greedy updates in this fashion corresponds to decentralized asynchronous iterations of an incremental subgradient algorithm for minimizing a particular objective function. Using convexity properties of the objective function, we show that GGE converges to the average consensus almost surely. Moreover, simulation results indicate that GGE converges significantly faster than geographic gossip for networks of a few hundred nodes.

Recently proposed in [1], *broadcast gossip* also makes use of the broadcast nature of wireless networks. For this algorithm, at each iteration, a node is chosen uniformly at random to broadcast its value. The nodes in the broadcast range of this node calculate a weighted average of their own value and the broadcasted value, and they update their value with this weighted average. In GGE, the broadcast nature of the wireless medium is utilized in order to greedily choose the neighbor to gossip with. However, in broadcast gossip, the value of the broadcasting node is independently incorporated at each neighbor. Consequently, broadcast gossip does not preserve the network average at each iteration. The simulations show that GGE performs better than broadcast gossip.

The paper is organized as follows. In Sec. II we state the formal definition of the algorithm. In Sec. III we prove that greedy gossip with eavesdropping converges almost surely to the average consensus. We empirically investigate the rate of convergence in Sec. IV and find that greedy gossip converges at a rate superior to existing average consensus algorithms.

II. GREEDY GOSSIP WITH EAVESDROPPING

Network connectivity is represented as a graph, $G = (V, E)$, with vertices $V = \{1, \dots, n\}$, and edge set $E \subset V \times V$ such that $(i, j) \in E$ if and only if nodes i and j directly communicate. We assume that communication relationships are symmetric and that the graph is connected. Let $\mathcal{N}_i = \{j : (i, j) \in E\}$ denote the set of neighbors of node i (not including i itself). In addition to maintaining the local variable, $x_i(k)$, each node maintains a copy of the current values at its neighbors, $x_j(k)$, for $j \in \mathcal{N}_i$. To initialize the algorithm, each node sets its gossip value to $x_i(0) = y_i$, and broadcasts this value to all of its immediate neighbors.

At the k th iteration, a node s_k is chosen uniformly at random from $\{1, \dots, n\}$. This can be accomplished using the asynchronous time model described in Bertsekas and Tsitsiklis [2], where each node "ticks" according to a Poisson clock with rate 1. Unlike the randomized gossip algorithms described in [3], where s_k randomly chooses a neighbor to gossip with, in GGE, s_k gossips with a neighbor that is currently the most different from its own value. That is, s_k identifies a node t_k satisfying

$$t_k \in \arg \max_{t \in \mathcal{N}_{s_k}} \left\{ \frac{1}{2} (x_{s_k}(k) - x_t(k))^2 \right\}.$$

When s_k has multiple neighbors that are all equally (and maximally) different from s_k , it chooses one of these neighbors at random. Then s_k and t_k exchange values and perform the update

$$x_{s_k}(k) = x_{t_k}(k) = \frac{1}{2} (x_{s_k}(k-1) + x_{t_k}(k-1)), \quad (1)$$

while all other nodes $i \notin \{s_k, t_k\}$ maintain their values at $x_i(k) = x_i(k-1)$. Finally, the two nodes, s_k and t_k , whose values have changed broadcast these new values so that their neighbors have up-to-date information.

The main idea of GGE is to use each node's knowledge of its neighbors' values to greedily perform the update that will have the greatest impact at each iteration. The key factor that makes greedy updates possible is the broadcast nature of wireless communications: because each node overhears its neighbors transmissions, it only takes one transmission to inform the entire neighborhood of a change. The following sections illustrate that performing greedy updates does not affect the convergence behavior of gossip (the algorithm still converges to the average consensus), and, in fact, greedy updates significantly improve the rate of convergence.

III. CONVERGENCE ANALYSIS

The standard approach to analyzing convergence of distributed averaging and gossip algorithms essentially involves expressing each update (or expected update) in terms of a linear recursion, $\mathbf{x}(k+1) = \mathbf{W}\mathbf{x}(k)$, and then imposing properties on the matrix \mathbf{W} which guarantee convergence [3], [8]. In particular, the standard distributed averaging and gossip algorithms are such that \mathbf{W} is the same at every iteration and, more importantly, \mathbf{W} is independent of $\mathbf{x}(k)$. Other researchers have considered the case where $\mathbf{W}(k)$ may change from iteration to iteration, to model either noisy-transmissions or time-varying topologies. That work generally assumes that $\mathbf{W}(k)$ matrices are i.i.d. realizations of a random matrix, and convergence is guaranteed so long as $\mathbb{E}[\mathbf{W}(k)]$ satisfies certain properties. Again, a key assumption is that $\mathbf{W}(k)$ is independent of $\mathbf{x}(k)$.

GGE updates can also be expressed in the form

$$\mathbf{x}(k+1) = \mathbf{W}(k)\mathbf{x}(k)$$

where $\mathbf{W}(k)$ is a stochastic matrix with $W_{s_k, s_k}(k) = W_{s_k, t_k}(k) = W_{t_k, s_k}(k) = W_{t_k, t_k}(k) = \frac{1}{2}$, $W_{i, i}(k) = 1$ for all $i \notin \{s_k, t_k\}$, and 0 elsewhere. However, in our setting, $\mathbf{W}(k)$ explicitly depends on $\mathbf{x}(k)$ via the choice of t_k . Because of this dependence, standard analysis for gossip and consensus algorithms cannot be applied to guarantee convergence of the GGE algorithm. Instead, we prove convergence by viewing our algorithm as a particular instance of the randomized incremental subgradient algorithm [5].

A. Incremental Subgradient Methods

Before continuing with an analysis of GGE, we briefly review incremental subgradient¹ methods. Consider the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \quad & \sum_{i=1}^n f_i(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in X, \end{aligned}$$

where we assume that each $f_i(\mathbf{x})$ is a convex function, but not necessarily differentiable, and X is a non-empty convex subset of \mathbb{R}^n . An incremental subgradient algorithm for solving this optimization is an iterative algorithm of the form

$$x(k+1) = \mathcal{P}_X[x(k) - \alpha_k g(s_k, x(k))], \quad (2)$$

where $\alpha_k > 0$ is the step-size, $g(s_k, x(k))$ is a subgradient of f_{s_k} at $x(k)$, and $\mathcal{P}_X[\cdot]$ projects its argument onto the set X . The algorithm is randomized when the component updated at each iteration, s_k , is drawn uniformly at random from the set $\{1, \dots, n\}$, and is independent of $x(k)$. Intuitively, the algorithm resembles gradient descent, except that instead of taking a descent step in the direction of the gradient of the cost function, $f(\mathbf{x}) = \sum_{i=1}^n f_i(\mathbf{x})$, at each iteration we focus on a single component of $f(\mathbf{x})$. The projection, $\mathcal{P}_X[\cdot]$, ensures that each new iterate $x(k+1)$ is feasible. Under mild conditions on the sequence of step sizes, α_k , and on the regularity of each component function $f_i(\mathbf{x})$, Nedić and Bertsekas have shown that the randomized incremental subgradient method described above converges to a neighborhood of the global minimizer [5].

B. Convergence of Greedy Gossip with Eavesdropping

GGE is a randomized incremental subgradient algorithm for the problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^n \max_{j \in \mathcal{N}_i} \left\{ \frac{1}{2} (x_i - x_j)^2 \right\} \quad (3)$$

$$\text{subject to} \quad \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \quad (4)$$

where y_i is the initial value at node i . Clearly, the objective function in (3) has a minimum value of 0 which is attained when $x_i = x_j$ for all i, j . Thus, any minimizer is a consensus solution. Moreover, the constraint $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ ensures that the unique global minimizer is the average consensus.

To see the connection between the GGE update, (1), and the incremental subgradient update, (2), let us define $g(s_k, t_k, x(k))$ such that

$$\begin{aligned} g_{s_k}(s_k, t_k, x(k)) &= x_{s_k}(k) - x_{t_k}(k), \\ g_{t_k}(s_k, t_k, x(k)) &= x_{t_k}(k) - x_{s_k}(k), \end{aligned}$$

¹Subgradients generalize the notion of a gradient for non-smooth functions. The subgradient of a convex function f_i at x is any vector g that satisfies $f_i(y) \geq f_i(x) + g^T(y - x)$. The set of subgradients of f_i at x is referred to as the *subdifferential* and is denoted by $\partial f_i(x)$. If f_i is continuous at x , then $\partial f_i(x) = \{\nabla f_i(x)\}$; i.e., the only subgradient of f_i at x is the gradient. A sufficient and necessary condition for x^* to be a minimizer of the convex function f_i is that $0 \in \partial f_i(x^*)$. See [5] and references therein.

and $g_i(s_k, t_k, x(k)) = 0$ for all $i \notin \{s_k, t_k\}$. Note that subscripts denote components of the vector g . One can easily verify that $g(s_k, t_k, x(k))$ is a subgradient of $f_{s_k}(x(k)) = \max_{j \in \mathcal{N}_{s_k}} \left\{ \frac{1}{2} (x_{s_k}(k) - x_j(k))^2 \right\}$. Use constant step size $\alpha_k = \frac{1}{2}$, and substitute the expression for $g(s_k, t_k, x(k))$ into (2) to find that this update is equivalent to the GGE update described in Section II. (Note: The use of $\alpha_k = \frac{1}{2}$ in conjunction with the form of the subgradient $g(s_k, t_k, x(k))$ ensures that the constraint $\sum_{i=1}^n x_i(k) = \sum_{i=1}^n y_i$ is satisfied at each iteration, so we can drop the projection term in (2).)

Nedić and Bertsekas study the convergence behavior of randomized incremental subgradient algorithms in [5]. Unfortunately, for a constant step size, their analysis only guarantees the iterates $x(k)$ will reach a neighborhood of the optimal solution: with probability 1, $\min_k f(x(k)) \leq \alpha n C^2 / 2$, where $C \geq \|g(s_k, x(k))\|$ is an upper bound on the norm of the subgradient [5]. We wish to show that $x(k)$ converges to the average consensus, \bar{x} , the global minimizer of our optimization problem. By exploiting the specific form of our particular problem, we are able to prove the following stronger result.

Theorem 1: Let $x(k)$ denote the sequence of iterates produced by GGE. Then $x(k) \rightarrow \bar{x}$ almost surely.

Proof: To begin, we examine the improvement in squared error after one GGE iteration. Based on the discussion in the previous section, the recursive update for GGE has the form

$$x(k+1) = x(k) - \frac{1}{2} g(k), \quad (5)$$

where, given s_k and t_k ,

$$g_i(k) = \begin{cases} x_{s_k}(k) - x_{t_k}(k) & \text{for } i = s_k, \\ -(x_{s_k}(k) - x_{t_k}(k)) & \text{for } i = t_k, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Expanding $x(k+1)$ via the expression (5), we have

$$\begin{aligned} \|x(k+1) - \bar{x}\|^2 &= \|x(k) - \frac{1}{2} g(k) - \bar{x}\|^2 \\ &= \|x(k) - \bar{x}\|^2 - \langle x(k) - \bar{x}, g(k) \rangle + \frac{1}{4} \|g(k)\|^2. \end{aligned}$$

Based on the definition of $g(k)$ in (6), given s_k and t_k , we have

$$\|g(k)\|^2 = 2(x_{s_k}(k) - x_{t_k}(k))^2,$$

and (abusing notation by using \bar{x} to denote both the average value and the average consensus vector)

$$\begin{aligned} \langle x(k) - \bar{x}, g(k) \rangle &= \sum_{i=1}^n (x_i(k) - \bar{x}) g_i(k) \\ &= (x_{s_k}(k) - \bar{x})(x_{s_k}(k) - x_{t_k}(k)) \\ &\quad - (x_{t_k}(k) - \bar{x})(x_{s_k}(k) - x_{t_k}(k)) \\ &= (x_{s_k}(k) - x_{t_k}(k))^2. \end{aligned}$$

Therefore, we have $\|x(k+1) - \bar{x}\|^2 = \|x(k) - \bar{x}\|^2 - \frac{1}{4} \|g(k)\|^2$ with probability 1, since the expression holds independent of the value of s_k and t_k .

Recursively applying our update expression, we find that w.p. 1,

$$\begin{aligned} \|x(k+1) - \bar{x}\|^2 &= \|x(k) - \bar{x}\|^2 - \frac{1}{4} \|g(k)\|^2 \\ &= \|x(k-1) - \bar{x}\|^2 - \frac{1}{4} \sum_{j=k-1}^k \|g(j)\|^2 \\ &\vdots \\ &= \|x(0) - \bar{x}\|^2 - \frac{1}{4} \sum_{j=0}^k \|g(j)\|^2. \end{aligned}$$

Since, $\|x(k+1) - \bar{x}\|^2 \geq 0$, this implies that

$$\sum_{j=0}^k \|g(j)\|^2 \leq 4\|x(0) - \bar{x}\|^2$$

w.p. 1, and, consequently, the series $\sum_{j=0}^k \|g(j)\|^2$ converges a.s. as $k \rightarrow \infty$. Since each term $\|g(j)\|^2 \geq 0$, this also implies that $\|g(k)\|^2 \rightarrow 0$ a.s. as $k \rightarrow \infty$. However, by definition, $g(k)$ is the subgradient of a convex function, and $g(k) = 0$ is both a sufficient and necessary condition for $x(k)$ to be a global minimizer. Thus, $g(k) \rightarrow 0$ a.s. implies that $x(k) \rightarrow \bar{x}$ a.s., since \bar{x} is the unique minimizer of (3)-(4). ■

It is not surprising that GGE converges almost surely, given that standard gossip converges and, after a single iteration, GGE never does worse than standard gossip. More interesting is the finding that GGE generally converges significantly faster. The next section examines this point further via simulation.

IV. CONVERGENCE RATE (SIMULATIONS)

In this section we illustrate the convergence behavior of GGE through simulation. We compare the performance of GGE with randomized gossip [3], geographic gossip [4], and broadcast gossip [1]. The network consists of nodes that are distributed uniformly at random over the unit square. The transmission radius is set to $\sqrt{2 \log n/n}$ such that the

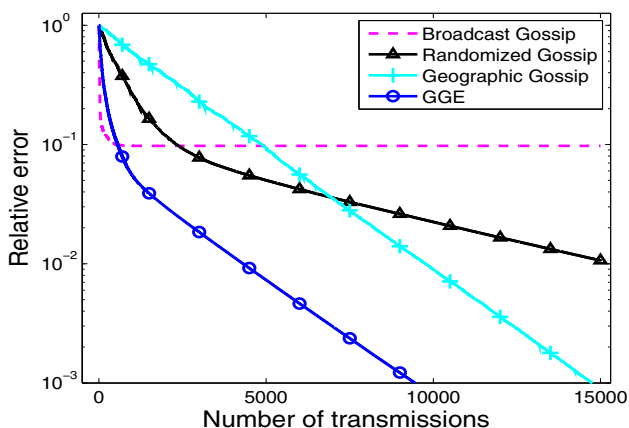


Fig. 1. Relative error versus number of transmissions for a network that has zero value everywhere except one random node with value 1 (averaged over 100 realizations of the graph).

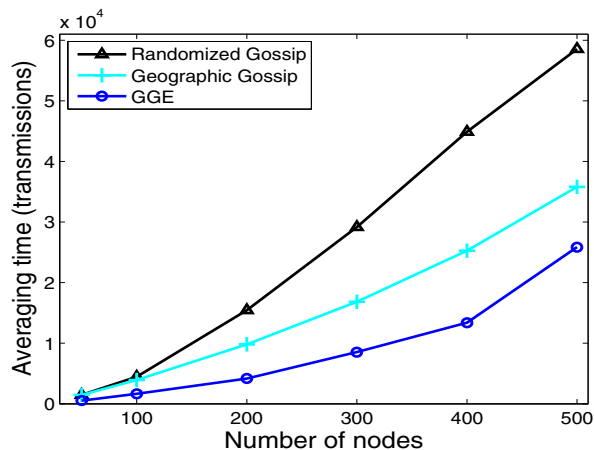


Fig. 2. The behavior of averaging time as the number of nodes in the network increases (for 100 realizations) for a network that has zero value everywhere except at one random node. Averaging time is the number of transmissions required to reach a relative error below 0.01.

random geometric graph is connected with high probability. All the figures show the average over 100 realizations of the random geometric graph. The first few transmissions during the operation of GGE are dedicated to broadcasting of initial node values such that each node collects the initial values of its neighbors.

Since consensus algorithms try to reach average consensus with as few radio transmissions as possible and different algorithms require different number of transmissions per iteration, we compare their performance in terms of number of radio transmissions instead of number of iterations. For GGE, the number of transmissions per iteration is three, one for initial transmission and two more for the broadcasting of the new values of the two nodes that have gossiped. The number of transmissions per iteration is two for randomized gossip and one for broadcast gossip. On the other hand, for geographic gossip this number depends on the distance between the nodes that exchange information.

Relative error is defined as $\frac{\|x(k) - \bar{x}\|}{\|x(0)\|}$ and its behavior for each algorithm is illustrated in Figure 1 for a network of 200 nodes with all nodes initially having value 0, except one random node with initial value of 1. For the same network we also demonstrate the rate of increase in convergence time for increasing network size. Figure 2 shows the *averaging time* versus the number of nodes, where the averaging time is the first time the relative error decreases below 0.01 (broadcast gossip is excluded since its relative error does not reach 0.01) in terms of number of transmissions. The simulations show that GGE converges significantly faster than randomized and geographic gossip algorithms since it achieves smaller relative error values with fewer transmissions. In the first few rounds the performance of broadcast gossip is similar to the performance of greedy gossip with eavesdropping, however as time passes the former converges to a false average whereas GGE converges towards the initial average.

We also consider a network with linearly varying initial values. Figure 3 shows the relative error behavior for a

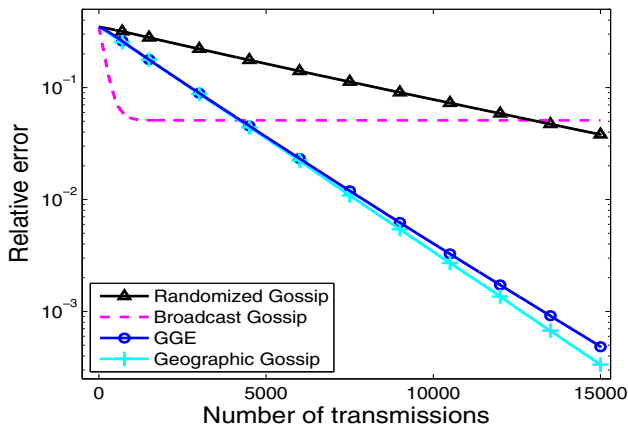


Fig. 3. Relative error versus number of transmissions for a linearly varying field (averaged over 100 realizations of the graph).

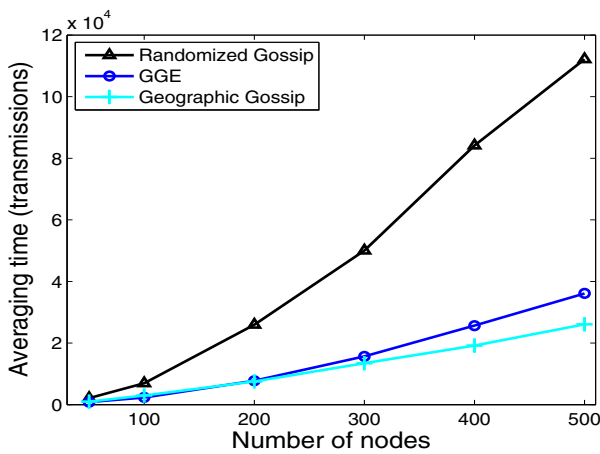


Fig. 4. The behavior of averaging time as the number of nodes in the network increases (for 100 realizations) for a linearly varying field. Averaging time is the number of transmissions required to reach a relative error below 0.01.

network of 200 nodes whereas Figure 4 illustrates the effect of increasing network size for linearly varying fields. For such an initialization, GGE performs better than randomized and broadcast gossip. However, its performance is slightly worse than that of geographic gossip's. This is because of the fact that a linearly varying field is nearly a worst case scenario for GGE, since all the subgradients yield the same improvement for this case.

V. DISCUSSION

In this paper we proposed a novel average consensus algorithm for wireless sensor networks. Greedy gossip with eavesdropping takes advantage of the broadcast nature of wireless communications and provides fast and reliable computation of average consensus. The idea of exploiting the broadcast nature of wireless communications in gossip algorithms also forms the core of the broadcast gossip procedure proposed by Aysal et al. in [1]. The primary difference is that a realization of broadcast gossip converges to a consensus value, but as illustrated in Figure 1 this value can indeed be quite biased

away from to the true average since broadcast gossip does not preserve the network average at each iteration.

Our current proposal of GGE involves uniform random selection of the node that performs gossip. We can change this selection procedure to focus on nodes that have a large differential with one of their neighbors, incorporating a "rejection" procedure, where a node chooses not to gossip if the maximum difference is smaller than a decaying threshold. Our current research is exploring the convergence behavior of such algorithms.

REFERENCES

- [1] T. Aysal, M. Yildiz, and A. Scaglione. Broadcast gossip algorithms. To appear in *Proc. 2008 Information Theory Workshop*, May 2008.
- [2] D. Bertsekas and J. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, MA, 1997.
- [3] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Trans. Info. Theory*, 52(6):2508–2530, June 2006.
- [4] A. Dimakis, A. Sarwate, and M. Wainwright. Geographic gossip: Efficient aggregation for sensor networks. In *Proc. Information Processing in Sensor Networks (IPSN)*, Nashville, TN, April 2006.
- [5] A. Nedić and D. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM J. on Opt.*, 12(1):109–138, 2001.
- [6] M. Rabbat, J. Haupt, A. Singh, and R. Nowak. Decentralized compression and predistribution via randomized gossiping. In *Proc. Information Processing in Sensor Networks*, Nashville, TN, Apr. 2006.
- [7] M. Rabbat, R. Nowak, and J. Bucklew. Robust decentralized source localization via averaging. In *Proc. IEEE ICASSP*, Phil., PA, Mar. 2005.
- [8] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53(1):65–78, Sep. 2004.