Efficient algorithms for topology control and information dissemination/ retrieval in large scale Wireless Sensor Networks

Leonidas M. Tzevelekas[1]

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
ltzev@di.uoa.gr

Abstract. Wireless Sensor Networks (WSNs) require radically new approaches for protocol/ algorithm design, with a focus towards energy efficiency at the node level. We propose two algorithms for energy-efficient, distributed clustering called Directed Budget Based (DBB) and Directed Budget Based with Random Delays (DBB-RD). Both algorithms improve clustering performance and overall network decomposition time when compared with state-of-the-art distributed clustering algorithms. Energy-efficient information dissemination in WSNs is proposed through modifying the movement of a simple Random Walk agent in a large scale Geometric Random Graph. The Random Walk with Jumps agent is compared against the Random Walk without backtracking agent. It is shown in simulations that the RW-J agent performs better than the RW agent in terms of Cover Time or Partial Cover Time. RW-J performs best when the underlying network topology has low connectivity, i.e. the graph has bad cuts. Information extraction from a large sensor field is a dual problem to information dissemination. We adopt the single mobile sink based information extraction methodology for collecting large amounts of data from a sensor field. The algorithms proposed are classified based on a. whether sensor nodes are allowed to transmit data over the wireless medium (single hop data forwarding) or not (almost zero hop data forwarding) and b. if there is detailed knowledge of sensor nodes locations in the field (deterministic algorithm) or not (randomized algorithm). All collection schemes are compared against required number of stops/ steps to completion and total physical distance covered by the mobile sink.

Keywords: Wireless Sensor Networks, distributed clustering, random walks, information dissemination, information harvesting from sensors, mobile sink based data harvesting

---

[1] Dissertation Advisor: Ioannis Stavrakakis, Professor

1 Dissertation Summary

Recent technological and scientific advances in the areas of solid state physics, integrated circuit design and telecommunications have led to enabling the functional design of innovative wireless networking models. These models comprise a large number of tiny sensor nodes working cooperatively towards building a wireless communication network. This doctoral dissertation focuses on the study/ proposal of energy-efficient algorithms for topology control and dissemination/ harvesting of information in large-scale Wireless Sensor Networks (WSNs). The doctoral dissertation comprises five chapters.

The first chapter of this dissertation is a detailed introduction in the subject of Wireless Sensor Networks. Specific details of recent technological advances towards making it technically feasible to produce large numbers of tiny sensor motes are explained. A large ensemble of sensor motes, which are embedded in the physical space, can produce a “smart environment”. The technical characteristics of “smart environments” are listed and the major models for Wireless Sensor Networks operation (as they have been proposed in literature) are explained. There is also a detailed report regarding the new technical challenges bound to be tackled by operation protocols designed for Wireless Sensor Networks. The most well known protocols in literature are explained at the end of the chapter (SPIN protocol, LEACH protocol, Directed Diffusion protocol, Publish/ Subscribe protocol).

Chapter 2 introduces two novel techniques for distributed clustering of sensor nodes in a large scale Wireless Sensor Network. The proposed Directed Budget Based (DBB) and Directed Budget Based with Random Delays (DBB-RD) algorithms have their basis on two previously published algorithms for distributed clustering of nodes in wireless networks, called Rapid and Persistent. The algorithms begin the distributed clustering process by distributing a set of coupons/ tokens offered to the initiator node evenly among the neighbors of that node; the process is then repeated in consecutive cycles of operation until the tokens are completely distributed or no more growth is possible in the network. The directed budget-based clustering algorithms called DBB and DBB-RD are proved through simulations to be energy-beneficial for the wireless sensor network due to both the reduced total number of exchanged messages in the network and the final cluster sizes achieved (close to the desired offered budget).

Another section of the dissertation describes an innovative technique for information diffusion in a large scale WSN. The described technique is based on random walks for information propagation inside the sensor network, which is modeled as a random geometric graph. The classic, well known, random walk involves the proliferation of the agent in the network by choosing uniformly at random among all next hop neighbors of the currently visited node. In contrast to this, the in-chapter-3-described technique involves the design of a “freezing” mechanism for the direction of movement of the random walk agent, such that the agent is allowed to be forwarded towards a specific direction in the network. The particular forwarding direction is retained by the random walk-with-jumps agent for as long as the agent will stay in the “freezing state”. It is shown through both simulations and analysis that the incorporation of such a freezing mechanism into the otherwise pure random walk movement of the random walk agent will be beneficial for the overall covering process of the sensor network.

In Chapter 4, the random walk based movement is tested for contributing in the data harvesting process when the mobile sink based data harvesting is assumed. The movement of the mobile sink inside the sensor field is designed based on a) the wireless transmission of sensed data at 1-hop distance away from the producing node and b) the wireless transmission of sensed data at almost 0-hop distance away from the producing node. Furthermore, the movement of the mobile sink involves a deterministic variant (deterministically scheduling the changes in location of the mobile sink) and a random walk based variant (choosing randomly among all possible next mobile sink locations).

The final chapter of the dissertation summarizes contributions and results of previous chapters and furthermore touches on subjects of further work, which can be directly extracted from topics/ results presented in this dissertation.

2 Results and Discussion

2.1 Introduction

One of the main challenges associated with large-scale, unstructured and dynamic networking environments is that of *efficiently* reaching out to all or a portion of the network nodes (i.e. *disseminating information*), in order to provide, e.g., software updates or announcements of new services or queries. The high dynamicity and the sheer size of such networking topologies ask for the adoption of decentralized approaches to information dissemination [1], [2], [3], [4]. One of the simplest approaches employed for disseminating information in such environments, is the traditional flooding approach. Under flooding ([5], [6], [7], [8]), each time a node receives a message for the first time from some node, it forwards it to all its neighbors except from that node. Despite its simplicity and speed (typically achieving the shortest cover time possible, upper bounded by the network diameter), the associated large message overhead is a major drawback.

As flooding is considered not to be an option for large scale, wireless networking environments due to strict energy limitations of individual sensor nodes, approaches based on random walks are viewed as reasonable choices [9], [10], [11], [12]. Random walks possess several good characteristics such as simplicity, robustness against dynamic failures or changes to the network topology, and lack of need for knowledge of the network physical and topological characteristics. The Random Walk agent (RW agent) employed within a network of wireless sensors moves from neighbor node to neighbor node in a random manner, frequently revisiting previously covered nodes in a circular manner, even without backtracking (returning to the node it just came from is not allowed); these revisits constitute overhead and impact negatively on the cover time [8].

The Jumping Random Walk (J-RW) mechanism is proposed as an efficient alternative against the RW agent for information dissemination/ retrieval in large scale environments, like wireless sensor networks. The proposed scheme exploits the benefits of the RW mechanism (simple, decentralized, robust to topology changes) while providing a `boost' in performance, i.e. accelerating the coverage process within the network. The latter is achieved by introducing a second state of operation to the RW agent in which the random movement paradigm is replaced by a non-random "directional" movement paradigm. It turns out that this improved significantly the cover time by "creating" virtual long links in topologies that lack them. It should be noted that the RW agent corresponds to a special case (parameter setting) of the proposed J-RW agent.

2.2 The RW agent in various topologies

A credible alternative to flooding for disseminating information in an unstructured environment, is the RW agent. In RW-based approaches, the initiator node employs an agent that will move randomly in the network, one hop/ node per time slot, informing (or querying) all the nodes in its path. Authors in [14] proposed a number of algorithms for RW-based searching in unstructured P2P networks, whereas probability-based information dissemination has been investigated for use in sensor networks [4], with data routing as the main consideration. Random walk in large-scale P2P nets has been shown to possess a number of good properties for searching and/or distributing of information within the network [15].

The overhead of RW-based solutions is considered to be much smaller than that of the flooding approaches, at the expense of a significant increase of cover time. *Cover time* is the expected time taken by a random walk to visit all nodes of a network. The generally relatively large (compared to flooding) cover time achieved under RW-based approaches depends on the network topology. For instance, it is O(N ln(N)) for the fully connected graph (best-case scenario) and $O(N^3)$ for clique topologies (worst case scenario) [13], [14]. Random walks on random geometric graphs $G(N; r_c)$ have been shown to have optimal cover time Θ(N ln(N)) and optimal partial cover time Θ(N) with high probability given that the connectivity radius of each node r_c fullfills a certain threshold property, i.e. given that [16]

$$r_c^2 \geq \frac{c8\ln(N)}{N}$$

For the rest of this work it is assumed that G(V;E) is a connected network. Let N be the number of network nodes (equivalently, the size of set V). In such a network, a RW agent moving according to the previously described mechanism, will eventually visit or cover all network nodes after some time (cover time). Let $C_r(t)$ be the fraction of network nodes covered (or visited) after t time units or movements of the RW agent (i.e., the RW agent start moving at t = 0), for a particular realization (sample path) of the walk and for a given initiator node. $C_r(t)$ will be referred to hereafter as the coverage at time t. Clearly, $C_r(t)$ depends on the network size, the network topology, the initiator node and other factors. If T_r denotes the cover time then $C_r(T_r) = 1$; clearly $C_r(0) = 0$. As time increases, the RW agent is expected either to move to a node that hasn't been covered previously (thus, $C_r(t)$ increases) or to move to an already covered node (thus, Cr(t) remains the same). Therefore, $C_r(t)$ is a non-decreasing function ($C_r(t_1) \cdot C_r(t_2)$, for $t_1 < t_2$).

The number of movements of RW-based solutions is much smaller than that under flooding approaches (where one movement of an agent corresponds to one message transmission), at the expense of a significant increase in cover time. For example, for the case of a fully connected network, a number of movements of the order of N ln(N), [17], is required under the RW mechanism, while under flooding approaches the corresponding number of movements (or messages) is of the order of N^2. On the other hand, cover time under the RW mechanism is of the order of N ln(N), while under flooding it is upper bounded by the network diameter plus 1 (e.g., for the case of a complete graph cover time under flooding is 2).

2.3 The J-RW agent

2.3.1 Motivation

Fig. 1 illustrates a RW agent movement path initiated from the initiator node depicted inside the dotted ellipsis. The random walker spends some time revisiting nodes in the depicted “upper-left" network part, while nodes in other network parts are left unvisited. Suppose now that after a few time units long enough to “cover" a certain network part the RW agent moves to a “new" (most likely uncovered) network part (“bottom right" network part in Fig. 2). It is more likely than before to cover nodes that have not been visited previously by the agent, and therefore, accelerate the overall network cover process.

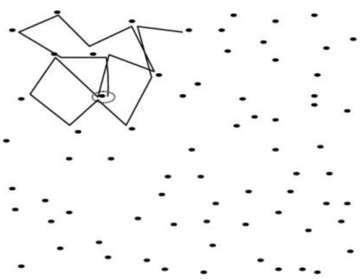

Fig. 1. RW Agent

One possible way for the agent to move away from a certain network region would be to carry out a number of consecutive directional movements, implementing a jump. This directional movement mechanism or jumping, initially proposed in [18], can be realized by switching occasionally away from the RW operation and engaging an operation implementing a directional movement. That is, such a RW agent (to be referred to as the Jumping Random Walk (J-RW)) operates under two states: State 0 under which it implements the typical RW mechanism without backtracking, and

State 1 during which the directional move is implemented; the time spent in state 1 (freezing state) will be referred to as the freezing (the direction) period.

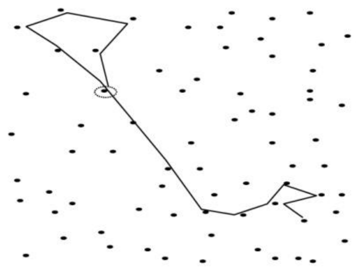

Fig. 2. J-RW Agent

The J-RW mechanism moves the agent - at the end of the freezing period – to networks that are expected (due to the directional freeze) to be geographically more distant than those reached by the RW agent after the same number of movements. That is, the introduction of the freezing state implements in essence jumps, defined as the physical distance between the nodes hosting the RW agent at the beginning and the end of the freezing period.

2.3.2 Description

The proposed J-RW mechanism is based on two underlying states. When in *State 0*, the J-RW agent operates as the already described RW agent. When in *State 1*, it implements a directional walk, by selecting as the next node to visit to be the neighbor of the current node that is the closest to the line connecting the current node and the node visited by the agent in the previous discrete time, in the direction away from the previously visited node. The directional walk may be easily implemented through a simple look up table involving the geographic locations of the neighbours of a node; this table determines the next node to forward the agent to under the directional walk, given that the agent came to this node from a given neighbour. The geographic information can be easily retrieved either at the time of deployment in the case of a static sensor field (with provisions for second, third, etc. choices when lower order choices are not available due to battery depletion), or after the deployment of the field with the help of a localization protocol run occasionally.

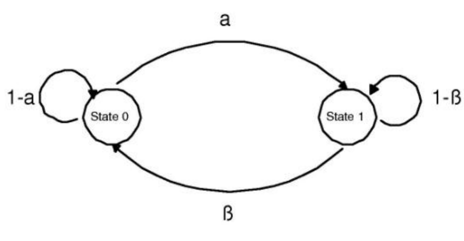

Fig. 3. Markov Chain mechanism for controlling J-RW agent movements

State transitions of the J-RW agent are assumed to occur at the discrete times according to a simple 2-state Markov chain, as shown in Fig. 3; let a (β) denote the transition probabilities from State 0 to State 1 (State 1 to State 0) and let T_0 = 1/a (T_1 = 1/β) denote the mean time (in discrete times of our reference time, or number of visits to nodes) that the agent spends in State 0 (State 1). Clearly, β (or, T_1) determines the length of time over which the directional walk is continuously in effect and, thus, the mean length of the induced jump. Similarly, a (or, T_0) determines the length of time over which the RW mechanism is continuously in effect. It should be noted that a and β should be carefully selected so that the mix of the two distinct operations is effectively balanced. It should be such that the implemented jump is sufficiently large to move the agent away from the current locality that is likely to be covered by the operation at State 0, and on the other hand, it should not be too large in which case it would leave uncovered large areas or require the random walk operation to operate long enough (at the increased cost of revisits) to cover the large areas between the start and the end of the jump. Similarly, a should be such that the time spent at State 0 be balanced so as to not over-cover or under-cover the current locality.

As previously for the RW mechanism, coverage and cover time under the J-RW mechanism may be defined in a similar manner, denoted by $C_j(t)$ and T_j respectively. $C_j(t)$ is a non-decreasing function of t taking values between 0 (for t = 0) and 1 (for t> T_j).

2.4 Coverage Analysis

2.4.1 Coverage under the RW and J-RW mechanisms

The main aim here is to derive an analytical expression for $C_r(t)$, which will serve as a tool for further understanding of random walk based information dissemination. Let’s assume that the network topology is fully connected (i.e., all nodes are connected to all other nodes). This is actually the case for large values of r_c in geometric random graphs. For example, for nodes scattered in the [0,1]x[0,1] 2-dimensional plane, any value of r_c>sqrt(2) ensures that there is a link among any pair of nodes.

In such a network, each time the RW agent decides to move to a new neighbor node at time t (thus, arriving at time t + 1), coverage $C_r(t)$: (a) may increase ($C_r(t + 1)$ = $C_r(t+1)$=N), provided that the new node has not been covered previously; or (b) remain the same ($C_r(t+1)$ = $C_r(t)$), provided that the new node has already been covered. Note that at time t, in a fully connected network the RW agent may select one out of N - 2 network nodes (i.e., all network nodes excluding the one the agent came from and the one that is currently located at). Since 1/N corresponds to the coverage contribution of the node the agent came from and 1/N to the coverage contribution of the node that is currently located at, then $C_r(t)$-2/N is the coverage corresponding to the remaining N-2 nodes and eventually, (N-2)x($C_r(t)$-2/N) corresponds to the number of nodes that have already been visited by the agent (excluding the one the agent came from and the one that is currently located at). It is easily derived now that (on average) the increment of the coverage after a RW agent moves at time t, is given by the probability 1 - $C_r(t)$ that it moves to a node not visited before multiplied by 1/N which is the contribution to coverage by each node that is visited for the first time. That is,

$$C_r(t+1) - C_r(t) = \frac{1}{N}\left(1 - C_r(t)\right) \tag{1}$$

Let t be continuous and let $\tilde{C}_r(t)$ denote the corresponding continuous and increasing function of $C_r(t)$. Based on previous equation we have

$$\frac{d\tilde{C}_r(t)}{dt} = \frac{1}{N}(1 - \tilde{C}_r(t)) \tag{2}$$

Equation (2) is a first class differential equation, and the solution is

$$\tilde{C}_r(t) = 1 - e^{-\frac{t}{N}} \tag{3}$$

Equation (3) was derived assuming a fully connected network. By reducing r_c in geometric random graphs, the number of neighbor nodes decreases and therefore a RW agent has fewer choices to move than before. Therefore, the fraction of nodes that have (not) been visited previously, is expected to deviate from $C_r(t)$. In order to account for the aforementioned decrease in the increase rate of $\tilde{C}_r(t)$ we introduce a positive constant k with 0<k<1, such that

$$\tilde{C}_r(t) = 1 - e^{-\frac{k}{N}t} \tag{4}$$

The case of k = 1 corresponds to the fully connected network topology (i.e., large values of r_c) as it is concluded from Equation (3).

Coverage under the J-RW mechanism is related to r_c, a and β. However, an analytical expression for the coverage considering r_c, a and β is difficult to be derived and its further investigation will be based on simulations presented in the following section.

2.5 Simulation results and evaluation

There are multiple simulation runs executed under specific sets of parameters for the network and the investigated schemes. During each simulation run there is a large-scale node set up, with node population varying from 100 to 3000 nodes depending on the case. The nodes are placed at random locations on a square plain [0,1] x [0, 1]. The random positions (x_u, y_u) of each node u in V are chosen within the set [0,1] using the uniform probability distribution. Each node u is aware about its own position: (x_u, y_u). Each node is connected to some other node if the euclidean distance among them is less or equal to r_c. Clearly, for r_c > sqrt(2) the resulting network is fully connected. Depending on N (the size of the network), the lower bound of r_c for which the topology remains connected varies (typically decreases as N increases). Four different values of r_c (0.05, 0.1 0.5 and 1.0) are considered in the sequel for those topologies of N = 1000. Note that all four values are less than sqrt(2).

2.5.1 The RW mechanism

Figure 5 presents simulation results for various topologies derived for r_c = 0.05, 0.1, 0.5 and 1.0. The first observation is that for the appropriate value of k (i.e. k = 0.3 for r_c = 0:05, k = 0.7 for r_c = 0.1, k = 0.9 for r_c = 0.5 and k = 1.0 for r_c = 1.0), the analytical expression for coverage, given by Equation (3) approximates well the simulation results.

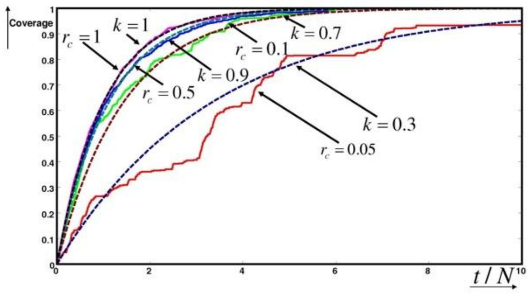

Fig. 4. Results for RW mechanism

2.5.2 The J-RW mechanism

Fig. 5-8 present simulation results under the J-RW mechanism for a network of 1000 nodes and various values of r_c and a. β has been kept constant and equal to 0:4, which means that as soon as State 1 is assumed (i.e., directional movement) the agent moves (on average) for 2-3 nodes towards a certain direction (more details are provided in the description of the J-RW mechanism in Section 3) before State 0 is assumed. In Fig. 5, coverage under the RW mechanism is clearly depicted and it is less than the coverage under J-RW for any value of a (e.g., 0.2, 0.4, 0.6 and 0.8). Note that the case depicted in Fig. 5 corresponds to a topology that is not highly connected (r_c = 0.05), thus even a relatively small value of β = 0.4 results in the J-RW doing significantly long jumps to get a performance improvement.

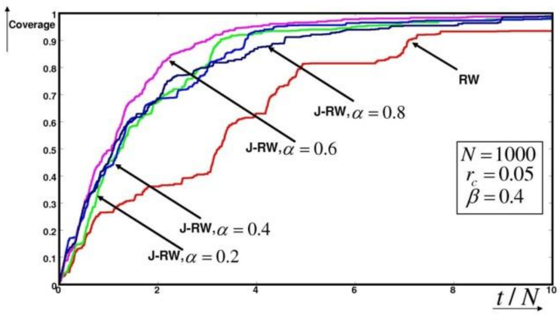

Fig. 5. Results for J-RW mechanism and low connectivity

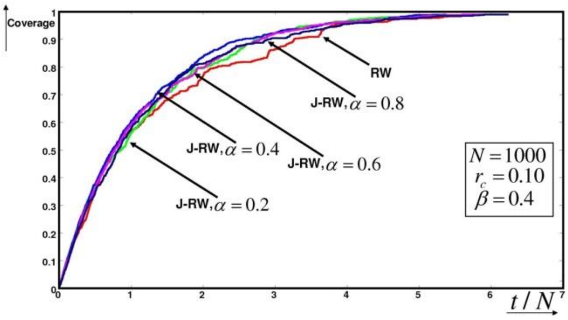

Fig. 6. Results for J-RW mechanism and medium connectivity

As the topology becomes more connected (r_c increases), the advantage of the J-RW mechanism is less obvious. For example, for the case depicted in Figure 6 ($r_c = 0.1$), coverage under the RW mechanism is still smaller than that under the J-RW mechanism (for any value of a), even though not that smaller as before, while for the case depicted in Figure 7 ($r_c = 0.5$), coverage under the RW mechanism is now larger than that under the J-RW mechanism (for any value of β). As r_c increases further, coverage under the RW mechanism is clearly higher than that under the J-RW mechanism for the specific combination of values of a and β. This is clearly depicted in Figure 8 for the case of $r_c = 1.0$ and can be attributed to the fact that the RW mechanism can now fully exploit the increased connectivity of the graph.

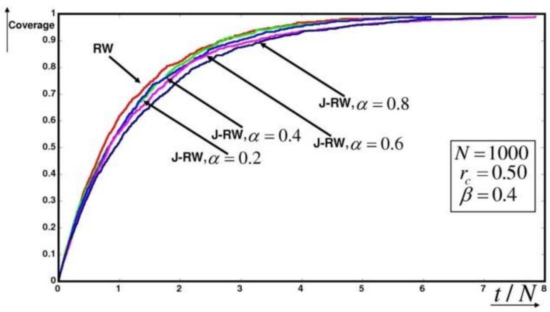

Fig. 7. Results for J-RW mechanism and higher connectivity

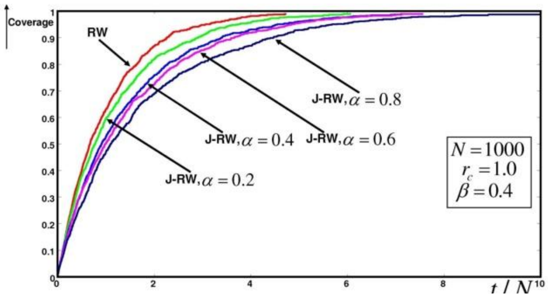

Fig. 8. Results for J-RW mechanism and highest connectivity

3 Conclusions

This dissertation presents results that are expected to foster technologies for next generation Wireless Sensor Networks. We develop energy efficient solutions for distributed clustering, for information dissemination and for information extraction in large scale Wireless Sensor Networks. We develop a methodology for distributed clustering in WSNs, called DBB, DBB-RD. We present a methodology for information dissemination that is energy efficient in terms of required number of steps of the RW agent to reach a given coverage level of nodes in the network. J-RW provides significantly higher coverage than RW because it is designed to avoid regions with bad cuts in the graph which trap the RW agent into revisiting already covered nodes. Finally we present a set of efficient algorithms for data collection in a Wireless Sensor Network when the single mobile-sink-based data harvesting methodology is adopted.

4 References

1. E. S. S. Dolev, J. Welch, Random walk for self-stabilizing group communication in ad hoc networks, IEEE Trans. on Mobile Computing 5 (7).
2. C. Avin, B. Krishnamachari, The power of choice in random walks: An empirical study, in: Proc. 9th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, (MSWiM), Malaga, Spain, 2006.
3. D. Braginsky, Rumor routing algorithm for sensor networks, in: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, 2002, pp. 22-31.
4. U. Feige, A spectrum of time-space trade-offs for undirected s-t connectivity, Journal of Computer and System Sciences 54 (2) (1997) 305-316.
5. A. Segal, Distributed network protocols, IEEE Trans. on Information Theory IT (29) (1983) 23{35.

6. B. Williams, T. Camp, Comparison of broadcasting techniques for mobile ad hoc networks, in: MOBIHOC 2002, 2002, pp. 194-205.
7. K. Oikonomou, I. Stavrakakis, Performance analysis of probabilistic flooding using random graphs, in: The First International IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC 2007), 2007.
8. D. Tsoumakos, N. Roussopoulos, Adaptive probabilistic search for peer-to-peer networks, in: 3rd IEEE International Conference on P2P Computing, 2003.
9. M. O.-K. M. Bani Yassein, S. Papanastasiou, On the performance of probabilistic flooding in mobile ad hoc networks, in: 11th International Conference on Parallel and Distributed Systems (ICPAD'05), 2005.
10. A. Abouzeid. Nabhendra Bisnik, Optimizing random walk search algorithms in p2p networks, Comnet 51 (6) (2006).
11. H. W. T. Lin, Search performance analysis in peer to peer networks, in: Third International Conference on Peer To Peer Computing (P2P'03), 2003, pp. 82-83
12. E. C. K. L. S. S. Q. Lv, P. Cao, Search and replication in unstructured peer-to-peer networks, in: ACM SIGMETRICS'02, 2002, pp. 258-259.
13. C. Avin, C. Brito, Efficient and robust query processing in dynamic environments using random walk techniques, in: Proc. IPSN 2004, Berkeley, California, 2004.
14. A. L.-O. R. Dorrigiv, P. Pralat, search algorithms for unstructured peer-to-peer networks, in: LCN 07: Proceedings of the 32nd IEEE Conference on Local Computer Networks, 2007, pp. 343-352.
15. M. Mihail. C. Gkantsidis, A. Saberi, Random walks in peer-to-peer networks, in: Proc. INFOCOM'04, Hong Kong, 2004.
16. C. Avin, G. Ercal, On the cover time of random geometric graphs, in: ICALP.(2005), pp. 677-689.
17. L. Lov'asz, Random walks on graphs: a survey, Combinatorics, Paul Erdos is Eighty, J. Bolyai Math. Soc., Vol. II 2 (1996) 353-397.
18. L. Tzevelekas, I. Stavrakakis, Improving Partial Cover of Random Walks in Large Scale Wireless Sensor Networks, 3rd IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC'09), 15 June 2009, Kos, Greece.