Improving Partial Cover of Random Walks in large-scale Wireless Sensor Networks [*]

Leonidas Tzevelekas
University of Athens
Dept. of Informatics & Telecommunications
Panepistimiopolis, Illisia 15784
Athens, Greece
Email:ltzev@di.uoa.gr

Ioannis Stavrakakis
University of Athens
Dept. of Informatics & Telecommunications
Panepistimiopolis, Illisia 15784
Athens, Greece
Email:ioannis@di.uoa.gr

Abstract

Random Walks (RWs) have been considered for information dissemination in large scale, dynamic and unstructured environments, as they are scalable, robust to topology changes and do not require topology information. Nevertheless, they are relatively slow in reaching out to the network nodes, particularly when applied to wireless networks (such as Wireless Sensor Networks (WSNs)), where the steps of the walk are of limited range, they reach only physically-close neighbours and are unable to move the walker far away from just covered area; as a consequence, time and resources are wasted in revisits of already covered nodes. In this paper we develop an innovative mechanism (referred to as the Jumping Random Walk (J-RW)) that alleviates drastically the aforementioned problems and is shown to improve substantially (compared to the RW) the cover time/overhead or coverage of WSNs modelled as a random geometric graph.

1. Introduction

One of the main challenges associated with large-scale, unstructured and dynamic networking environments is that of *effectively* reaching out to all or a portion of the network nodes (i.e., *disseminating information*) in order to provide software updates or announcements of new services, query them, etc; that is, disseminating information effectively. The efficiency of information dissemination schemes in such environments is measured in terms of the message overhead or the time (or number of sequential (transmission) steps) required to achieve a certain portion of network coverage, or derivative measures of those. The term *cover time, C* will be used to refer to that time, with the understanding that the network coverage is for a specific portion of the network which could be 100% or less; when needed, the specific portion will be mentioned.

One of the simplest approaches employed for disseminating information in an unstructured, dynamic and of large scale networking environment, is the traditional *flooding* approach. Under flooding ([10], [2]), 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)), the associated large message overhead is a major drawback. In order to reduce the (typically unacceptably large) overhead induced by flooding, *probabilistic flooding* ([7], [6], [8]) may be employed, under which the message forwarding to a node takes place with some probability less than 1. Although probabilistic flooding manages to reduce the message overhead, this reduction occurs at the expense of an increase in the cover time and a likely decrease in the coverage of the network. *Controlled flooding* may also be employed to reduce the large overhead of flooding; in this case flooding is limited to a predefined number of hops, namely K hops, away from the initiator node. Both the induced overhead and cover time increase with K. If the value of K is very large, then the K-hop controlled flooding scheme would approach the traditional flooding.

A credible alternative to flooding for disseminating information in an unstructured environment, is the *Random Walk (RW)* ([3], [4], [9]). 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. According to the typ-

---

[*] This work has been supported in part by the project ANA (Autonomic Network Architecture) (IST-27489), the PENED 2003 program of the General Secretariat for Research and Technology (GSRT) co-financed by the European Social Funds (75%) and by national sources (25%) and the NoE CONTENT (IST-384239)

978-1-4244-4439-7/09/$25.00©2009IEEE

ical RW-based information dissemination paradigm, each node receiving the *RW agent or packet*, chooses a forwarding neighbor arbitrarily based on the uniform distribution. For a network modelled as a graph $G(V, E)$ (where V and E denote its vertices and edges, respectively) a node $v \in V$ with connectivity degree $\delta(v)$ chooses each one of its next hop neighbors $u \in V$ with probability $p_u = 1/\delta(v)$ if $(v, u) \in E$ and $p_u = 0$ for all other $u \in V$. An important (for this paper) variant of RW-based propagation of information is the RW without backtracking. Under this scheme the RW visiting a node $v \in V$ will choose the next hop node $u \in V$ arbitrarily among the neighbors of node v with probability $p_u = 1/(\delta(v) - 1)$ if $(v, u) \in E$, where u is any neighbor of v except the one from last hop, and $p_u = 0$ for all other $u \in V$. 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 in the cover time.

In this paper we consider the problem of disseminating information (such as queries) across a large-scale, resource-limited, ad-hoc-structured wireless network, such as Wireless Sensor Networks (WSNs). As flooding is considered not to be an option for such environments due to the high resource requirements, RW-based approaches are viewed as reasonable choices and have been considered in the past in WSNs [11]. In addition to the low packet transmission overhead (compared to flooding), RWs 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 such as the neighbor's location or degree, the transmission radius, the symmetry of connections, etc.

The generally relatively large (compared to flooding) cover time C achieved under RW-based approaches depends on the network topology. For instance, it is $O(nlog(n))$ for the complete d-regular graphs (best-case scenario) and $O(n^3)$ for clique topologies (worst case scenario), where n is equal to the number of nodes [1]. Generally, it has been shown that C is lower for high connectivity network topologies, such as complete graphs, and it is higher in network topologies presenting bottlenecks. In the latter case, the number of revisits (which affect the induced overhead and cover) of already covered nodes (and resource waste) becomes particularly high.

The innovative RW-based information dissemination scheme proposed in this paper (to be referred to as the Jumping Random Walk (J-RW)) entails the key idea of freezing occasionally the course (or direction) of the RW, so as to help drive the RW agent away from the area of wandering just before the freeze. That is, the J-RW operates under two states: the freezing state during which the directional move is implemented, and the regular state under which it implements the typical RW without backtracking. This way, the proposed J-RW moves the RW agent at the end of the freezing period to geographic areas that are expected (due to the directional freeze) to be more distant than those reached by the regular RW after the same number of steps. That is, the introduction of the freezing state implements in essence jumps, defined as the physical distance between the nodes hosting the RW at the beginning and the end of the freezing period.

In WSNs the physical and the network topologies are typically correlated: a long path between two nodes in the network topology corresponds to a large physical distance between these nodes in the physical topology. As the J-RW moves the RW quickly into new regions, it is expected that the probability of revisiting previously visited nodes would decrease under the J-RW, compared to that under the RW without backtracking. The major concept to point out here is the fact that a simple random walker or even a random walker without backtracking has the tendency of performing circles (i.e. doing frequent revisits of already covered nodes in a circular manner) within the physical/ network topology of the sensor network. Therefore, the scheme proposed under J-RW focuses on alleviating this misbehavior of random walkers and reducing covered nodes revisits. As a result, it is expected that a larger portion of *distinct* network nodes would be visited over a given time (number of steps) under the J-RW, thus improving the cover time C.

The improvement in the cover time may be viewed as a consequence of ”sampling” the network more uniformly, by moving the sampling agent (ie., the RW) into remote and likely new (not yet sampled) areas, as opposed to keeping the RW wandering around a certain locality according to the RW paradigm and (over)sampling predominately a certain locality. When a network graph has long links (that can take an agent into a remote network region in a topological sense), it has been shown in [3] that a RW produces a uniform sampling of the network nodes. In essence, the proposed J-RW applied over a network with no long (physical) links (as a WSN) creates virtual long links in this network and results in an environment that is equivalent to that of applying the RW over a network with some long links. Thus, the proposed J-RW is expected to result in a more uniform sampling of the network nodes, which - as argued earlier - leads to a better cover time.

Besides the improved cover time, the increased uniformity of the (node) sampling under the proposed J-RW may be on its own another important property of the proposed information dissemination scheme when considered in conjunction with certain specific and fairly common applications in WSNs such as those related to sensing the environment. In such applications and due to the typically high spatial correlation of nearby nodes, a dissemination of a query on the state of the field may target *only* a portion of the network nodes to conserve energy [1]. Since the J-RW possesses the uniform sampling capability as argued earlier, it is expected that the query dissemination and collected responses would better represent the state of the WSN field and contain less redundant information. For such environments it is reasonable to base the evaluation of information dissemination schemes on the partial cover time as opposed to the 100% cover time.

Network cover is typically measured in terms of the portion of the nodes *actually visited or reached*. This definition may be too restrictive and not sufficient for establishing the effectiveness of an information dissemination scheme in certain cases. For instance when the information dissemination scheme carries advertising information about a new service, an effective advertising (in this case) scheme may be defined as the one that brings the announcement within a distance L (hops) from each network node, as opposed to each network node [5], as a low intensity (L hop) search for this information launched by a node can easily discover the information. In this case, mechanisms that yield high $L - cover$ of the network (as defined above), for $L > 0$, (as opposed to 0-cover, corresponding to the standard cover C) are considered to be effective.

The notion of L-cover may well be applicable to a WSN in which the disseminated information is not service advertisement but a query for the state of the WSN, represented collectively by its nodes. As there is redundancy in the information carried by closely located nodes, it may be worth querying and retrieving the state of a uniformly spaced subset of the network nodes only, to conserve energy. In this case, the query disseminating scheme should ensure that it reaches at least one (or only one) network node over a given locality. Or, equivalently, an efficient such scheme would be the one that induces high L-cover of the network, as this value would capture the portion of the network that has been effectively sensed (in the L-hop sense discussed before). In addition to having major applications run over WSNs whose efficiency is more effectively measured in terms of the L-cover for $L > 0$, it is important to mention that there is another argument stressing the relevance and usefulness of the L-cover for WSNs. It is known that the basic operation of a WSN requires the exchange of $hello$ messages between neighbors. By piggybacking state information in these messages, each node could possess the state of neighboring nodes almost for free and could include it in a response to a query. In such cases, the 0-cover of the network is equivalent to the 1-cover (which is achieved with lower overhead and with shorter cover time), as all the neighbors of a node receiving the query are also considered to be receiving it. By allowing L-hop neighbour information to be carried by the $hello$ messages, one can extend the previous arguments and establish an equivalence of the 0-cover with the L-cover.

2. The proposed Jumping Random Walk (J-RW) scheme

In this section the proposed scheme is introduced and described. In view of the discussion in the previous section and without loss of generality, the remaining of the paper will consider a WSN as the typical representative of the environment to which the proposed information dissemination scheme would be applied.

The WSN considered here will be modelled as a geometric random graph. Such a network can be constructed by laying the n network nodes on a 2-dimensional plane using some 2-dimensional distribution, such as the Uniform or Poisson distributions. The set of nodes V is then complemented with a set of edges E such that for any two nodes $u, v \in V$ there will exist a link $(u, v) \in E$ iff $||u - v|| \leq r_c$, where r_c is the connectivity radius of each node and $|| \cdot ||$ denotes the euclidian distance. Let $G(n, r_c)$ denote such a random geometric graph.

The proposed J-RW is a RW-based scheme that determines the movement of a single agent through the network. The reference time that governs the operation of the RW-based information dissemination process is discrete, with the discrete times defined to coincide with the time of arrival of the agent to a node.

The proposed J-RW is a 2-state RW operating in two distinct modes. When in state 0, the J-RW operates as the already described random walk without backtracking. When in state 1, it implements a directional walk, by selecting as the next node to visit to be the neighbour 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 previous visited node. The directional walk may be easily implemented through a simple look up table involving 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 location of each node and those of all it's neighbors should be registered in order for the look-up tables to be set up. Such 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.

State transitions of the J-RW are assumed to occur at the discrete times according to a simple 2-state Markov chain, as shown in Fig. 1; let α (β)denote the transition probabilities from state 0 to 1 and let $T_0 = 1/\alpha$ ($T_1 = 1/\beta$) denote the mean time (in discrete times of our reference time, or number of visits to nodes) that the Markov chain spends in state 0 (1). Clearly, β (or, T_1) determines the length of

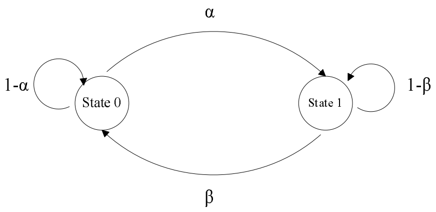

Figure 1. Markov chain for controlling random long jumps

time over which the directional walk is continuously in effect and, thus, the mean length of the induced jump. Similarly, α (or, T_0) determines the length of time over which the random walk without backtracking is continuously in effect. It should be noted that α and β should be carefully selected so that the mix of the two distinct operations is effectively balanced. β 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 random walk operation, 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, α should be such that the time spent under the random walk without backtracking operation be balanced so as to not over-cover or under-cover the current locality. A realization of the J-RW is depicted in Fig. 2, where the times of operation of the directional walk (jumps) and the random walk without backtracking (wandering around the nodes initiating jumps) are clearly shown.

The performance of the proposed J-RW can be evaluated in term of the cover time C. In view of the discussion in the previous section it is reasonable to focus on *partial* L-cover time, for $L >= 0$, to be denoted by $C(L)$.

3. Performance Evaluation and Results

3.1. Set Up of the Simulations

The performance of the proposed J-RW and the RW without backtracking are evaluated through simulations in this section. In the simulations it is assumed that each of the J-RW and the RW without backtracking is run for a given number of steps or discrete times H (to be referred to as the time/overhead available *budget*), and the resulting *partial* L-cover - to be denoted by $C(L)$ - is measured for $L = 0, 1, 2, 3$. The dissemination scheme that achieves a higher cover for the same time would clearly be the best performing.

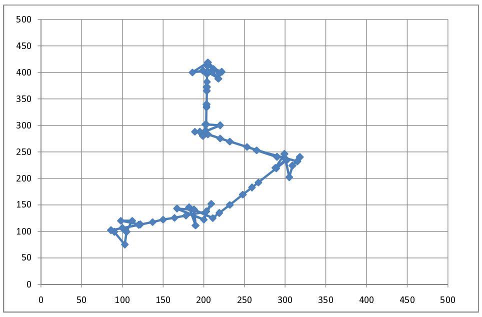

Figure 2. Random walk packet moving with jumps

Table 1. Simulated Modes for Random Walk with jumps

Model	α	β	E(stay in 0)	E(stay in 1)
jrw-1	0.6	0.3	1.67	3.34
jrw-2	0.5	0.5	2	2
jrw-3	0.3	0.6	3.34	1.67

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 WSN set up, with a total node population of N=6000. The nodes are placed at random locations on a square plain with total length X and total width Y. The random positions (x_u, y_u) of each node $u \in V$ are chosen within the sets $[0, X)$ for x_u and $[0, Y)$ for y_u using the uniform probability distribution, i.e. $P(x_u = x) = 1/X$ and $P(y_u = y) = 1/Y$. The locations of the nodes are guaranteed to be different for each simulation run because the seed values are chosen explicitely different prior to each run. Furthermore, the transmission range of each node r_c (identical for all nodes in the network) is selected such that the resulting topology is connected with a mean connectivity degree typical for WSN implementations ($\delta(u)_{mean} = 11.78$). The local neighborhood of each node u is comprised of all other nodes $v \in V$ located within the transmission range of node u (a Euclidian distance not greater than the transmission range). All presented results are calculated as mean values of the results obtained during each run. The standard deviations of the mean values are given to take account for the statistical accuracy.

As indicated in the previous section the proposed J-RW will be parameterised in terms of α and β; the pairs of these parameters are considered in the simulations, referred to as cases as shown in table 1. Case Jrw-1 yields the largest average jumps and shortest time of continuous operation of the random walk; the reverse holds true for case Jrw-3, while the lengths of the times spent continuously operating in the directional walk and the random walk are equal under case

Table 2. Simulation Results for $C(0)$

Budget H	2000	8000	10000	16000	18000
$C(0)_{rwnb}$	0.1037± 0.0263	0.3566± 0.0502	0.4385± 0.0385	0.5883± 0.0602	0.6106± 0.0666
$C(0)_{jrw-1}$	0.1843± 0.0166	0.4254± 0.0414	0.4885± 0.0526	0.5229± 0.0825	0.5768± 0.0571
$C(0)_{jrw-2}$	0.1547± 0.0247	0.4357± 0.0351	0.5119± 0.0615	0.6467± 0.0861	0.6301± 0.0953
$C(0)_{jrw-3}$	0.1601± 0.0276	0.4265± 0.0464	0.4649± 0.0834	0.5832± 0.0945	0.6548± 0.1082

Jrw-2. The RW without backtracking case results are referred to using the subscript $rwnb$

3.2. Results on Coverage $C(0)$

Coverage $C(0)$ is measured by counting the total number of distinct nodes that the RW without backtracking or J-RW visit before the available budget H is exhausted. The comparative results for the three cases of the J-RW and the RW without backtracking are shown in table 2 for various values of the available cover time or budget H. Clearly, the J-RW outperforms the RW without backtracking as it yields a higher $0-$cover for the majority of values of H shown. For instance, for a cover time or budget of $H = 2000$ the resulting 0-cover is increased by 78% for Jrw-1, by 49% for Jrw-2 and by 54% for Jrw-3, compared to that under the RW without backtracking for the same H. It is also worth noting that for a relatively large value of H (e.g., 10000 in table 2), the performance improvement is relatively smaller (11% for Jrw-1, 16% for Jrw-2 and 6% for Jrw-3), suggesting that the proposed J-RW exhibits its greatest advantages for low to medium cover times H. This reduction in the gain achieved by the J-RW may be attributed to the fact that the jumps implemented by the J-RW are increasingly likely to bring the walk in an area already covered to some extent (more so for a larger H). For this reason, only results for low values of H will be presented in the remaining of this section. Such results are shown in Fig. 3 for values of H up to 3500. From this figure one can argue that there is a linear relationship between $C(0)$ on H for such low values of H, or when the network cover is low. The J-RW variants outperform the RW without backtracking throughout most values of H, with Jrw-1 performing best regularly as compared to Jrw-2, Jrw-3.

3.3. Results for $C(1)$, $C(2)$, $C(3)$

The results for $C(1)$ are shown in table 3. It is clear that the conclusions drawn for $C(0)$ are still valid and further emphasized. For instance, for $H = 2000$, $C(1)$ is increased by 122% for Jrw-1, by 57% for Jrw-2 and by 70% for Jrw-3 compared to that achieved under the RW without backtracking.

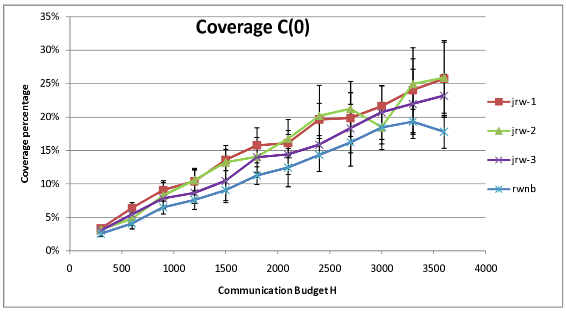

Figure 3. Coverage $C(0)$ for lowbudget H

Table 3. Simulation Results for $C(1)$

Budget H	2000	8000	10000	16000	18000
$C(1)_{rwnb}$	0.1711± 0.0621	0.4697± 0.1087	0.5411± 0.0796	0.7481± 0.0686	0.7924± 0.0977
$C(1)_{jrw-1}$	0.3807± 0.0522	0.6613± 0.0783	0.7315± 0.1046	0.7119± 0.1302	0.7683± 0.0897
$C(1)_{jrw-2}$	0.2689± 0.0874	0.6721± 0.0778	0.6871± 0.1158	0.8450± 0.1175	0.8013± 0.124
$C(1)_{jrw-3}$	0.2908± 0.0861	0.6047± 0.1035	0.6302± 0.1405	0.7215± 0.1369	0.8141± 0.1383

For low values of H (as seen in Fig. 4) one can clearly observe a linear dependence of $C(1)$ on H, as for the case of $C(0)$. As expected, one can clearly observe that $C(1) > C(0)$. There is an almost 100% increase in the partial cover which - as argued in the introduction - could be gained at almost no cost, by adding a few more bits to the *hello* packets' field.

Results for $C(2)$ and $C(3)$ are presented for low budgets H, as seen in Fig. 5 and Fig. 6, only for low values of H to conserve space. As expected, the cover achieved is higher than for lower values of L. Still the proposed J-RW greatly outperforms the RW without backtracking.

Finally, the coverage results for $L > 0$ presented earlier also indicate to some extent the uniformity of the coverage achieved by the J-RW and RW schemes. For instance, a scheme that results in a larger C(3) coverage implies that the

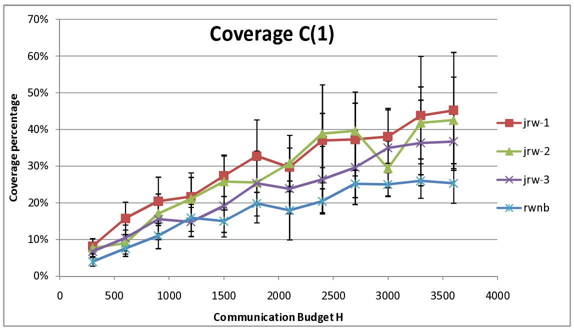

Figure 4. Coverage $C(1)$ for lowbudget H

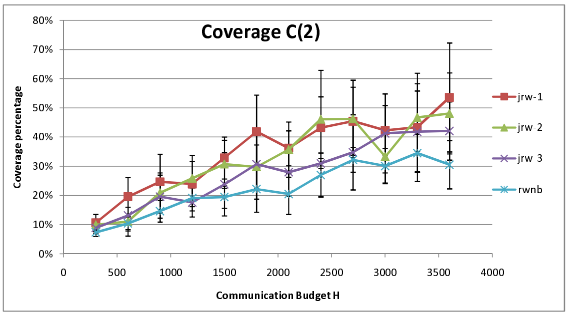

Figure 5. Coverage $C(2)$ for lowbudget H

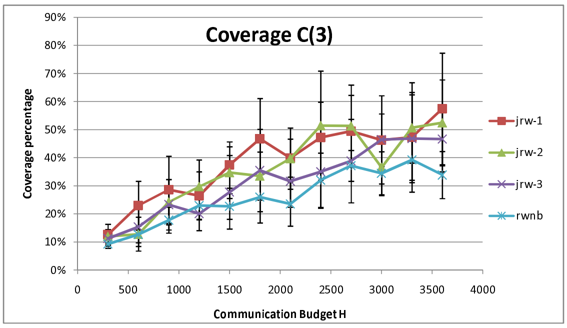

Figure 6. Coverage $C(3)$ for lowbudget H

informed nodes are more uniformly spread since the overlap of the circles of radius 3 around each informed node is smaller.

4. Conclusion

In this paper, an innovative RW-based information dissemination scheme is proposed and studied for large-scale, resource-limited, ad-hoc-structured wireless networks, such as Wireless Sensor Networks (WSNs), aiming at improving on the (partial) network cover or cover time achieved by the common Random Walk (RW).

The key idea behind the proposed Jumping Random Walk (J-RW) is that of freezing occasionally the course (or direction) of the RW, so as to help drive the RW agent away from the area of wandering just before the freeze and likely reduce the revisits to previously visited nodes. That is, the J-RW operates under two states: the freezing state during which the directional move is implemented, and the regular state under which it implements the typical RW. The introduction of the freezing state implements in essence jumps, defined as the distance travelled by the agent during the freezing period.

Looking at it from a different angle, the proposed scheme creates (some) virtual long links in an environment with only very short physical links, which help in covering the network more effectively; the latter is also expected since long links are also known to yield a more uniform sampling of the network under a RW. In essence, the proposed J-RW applied over a network with no long (physical) links (as a WSN) creates virtual long links in this network and results in an environment that is equivalent to that of applying the RW over a network with some long links. Thus, the proposed J-RW is expected to result in a more uniform sampling of the network nodes, which - as argued earlier - leads to a better cover time or network cover.

Finally, by extending the definition of coverage to include all nodes that are neighbours to already covered nodes (which neighbours are almost instantaneously and for free reachable through the $hello$ messaging mechanism), or to include nodes that are $L > 1$ hops away from already covered nodes, it is shown that the enhanced performance of the proposed J-RW mechanism is further exacerbated.

Numerical results are derived and show that the proposed J-RW can improve substantially the (partial) cover time of WSNs modelled as a random geometric graph.

References

[1] C. Avin and C. Brito. Efficient and robust query processing in dynamic environments using random walk techniques. In *Proc. IPSN 2004*, Berkeley, California, April 2004.

[2] B.Williams and T.Camp. Comparison of broadcasting techniques for mobile ad hoc networks. In *MOBIHOC 2002*, pages 194–205, 2002.

[3] M. M. C. Gkantsidis and A. Saberi. Random walks in peer-to-peer networks. In *Proc. INFOCOM'04*, Hong Kong, 2004.

[4] M. M. C. Gkantsidis and A. Saberi. Hybrid search scemes for unstructured peer to peer networks. In *Proc. INFOCOM'05*, Miami, USA, 2005.

[5] K. O. D. Kogias and I. Stavrakakis. Study of randomly replicated random walks for information dissemination over various network topologies. In *Proc. WONS 2009)*, 2009.

[6] D.Tsoumakos and N.Roussopoulos. Adaptive probabilistic search for peer-to-peer networks. In *3rd IEEE International Conference on P2P Computing*, 2003.

[7] K.Oikonomou and I.Stavrakakis. Performance analysis of probabilistic flooding using random graphs. In *Proc. AOC2007*, June 2007.

[8] M. O.-K. M.Bani Yassein and 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.

[9] A. L.-O. R.Dorrigiv and P.Pralat. search algorithms for unstructured peer-to-peer networks. In *LCN 07: Proceedings of the 32nd IEEE Conference on Local Computer Networks*, pages 343–352, 2007.

[10] A. Segal. Distributed network protocols. *IEEE Trans. on Information Theory*, IT(29):23–35, 1983.

[11] S. D. Servetto. Constrained random walks on random graphs: routing algorithms for large scale wireless sensor networks. In *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, pages 12–21, 2002.