Study of Randomly Replicated Random Walks for Information Dissemination Over Various Network Topologies

Dimitris Kogias
National And Kapodistrian
University of Athens
Department of Informatics
& Telecommunications
Panepistimiopolis, Ilissia 15 784
Athens, Greece
Tel: +30 210 7275315, Fax: +30 210 7275333
Email: dimkog@di.uoa.gr

Konstantinos Oikonomou
Ionian Univeristy
Department of Informatics
Corfu, Greece
Email: okon@ionio.gr

Ioannis Stavrakakis
National And Kapodistrian
University of Athens
Department of Informatics
& Telecommunications
Panepistimiopolis, Ilissia 15 784
Athens, Greece
Tel: +30 210 7275315, Fax: +30 210 7275333
Email: ioannis@di.uoa.gr

Abstract—Flooding is well known to be the "fastest" way to propagate information throughout a network (achieve 100% coverage), at the expense of typically unacceptably large message-forwarding overhead; when the overhead is controlled or limited, then the achieved coverage is reduced. A Single Random Walker (SRW) is another popular mechanism for information dissemination that is very "slow" compared to flooding but utilizes less overhead and can potentially achieve better coverage than flooding. The latter may be attributed to the better "stretching" properties (ability to visit further away network regions) of the SRW and is particularly observed if coverage is defined as the set of network nodes which are at most L hops away from a network node which received the information (notion of L-coverage, $L \geq 0$).

Randomly Replicated Random Walkers (RRRWs) are considered in this paper as a broad class of information dissemination mechanisms whose members are defined by a given value of the (first) replication probability; clearly a SWR corresponds to the RRRWs with zero (first) replication probability, while for values of the (first) replication probability close to one a large number of information disseminating agents (RWs) are generated resembling the operation of flooding. By considering various performance metrics (such as the achieved coverage for a given message-forwarding budget and the time required to spend the available message-forwarding budget), and by considering various network topologies (such as Random Geometric, Erdös- Rényi and Clustered environments), we show that the class of RRRWs can "fill" the performance space, between the performance induced by flooding and the SRW, thus offering a richer set of information dissemination strategies that can better match desirable performance constraints or objectives.

I. INTRODUCTION AND MOTIVATION

Information dissemination is important to the robustness or functionality of modern, highly dynamic and of large scale networking environments. Depending on its nature the disseminated information may be destined to the entire network or a subset of its nodes. In this work, and without loss of generality, it will be assumed that the disseminated information provides information about the existence of a service and its location within the network and will also be referred to as service advertisement. This service advertisement may target all the network nodes or a subset of them (to keep the associated overhead low). In the former case, each network node desiring the particular service will know right away about its existence and location. In the latter case, each network node desiring a service for which it has not received any advertisement will have to look for it by launching a service search process. This search process will try to find an informed node in order to obtain the desired information. Clearly, the closer the location of an informed node, the smaller the expected overhead (in terms of message overhead, delay etc) of the searching process. In view of the above, an effective information dissemination process should either reach out to the entire network with the least possible overhead (if coverage of the entire network is desired), or should try to get the information as close as possible to as many nodes as possible, thus exhibiting good information stretching properties, as explained more later.

The node that initiates the advertising process is referred to as the *service node* (or the *initiator node*). The set of informed nodes is referred to as the *advertising network*, whose size would depend on the intensity of the advertising process, expressed here in terms of the number of messages available to advertising (the *advertising budget (H)*).

In this paper, the effectiveness of two very popular information dissemination schemes the traditional *flooding* and the *single random walker (SRW)*, in disseminating information over various unstructured network topologies will be investigated, with respect to metrics such as the size of the generated advertising network, the time needed for the completion of the dissemination process and the formation of the advertising network, and the stretchiness of the resulting advertising network (explained later) that reflects coverage of a broader area of the network.

In addition to the two aforementioned information dissem-

ination schemes, the *Randomly Replicated Random Walkers (RRRWs)* information dissemination scheme will also be introduced and studied. The main idea behind the introduced class of RRRWs, is to allow for each random walker to *replicate* itself. Each time a random walker arrives to a node, a replica of the original walker is created (referred to as the "child" walker) with some (not necessarily fixed) probability, and the available advertising budget of the "mother" walker is split equally between the "mother" and the "child"; for efficiency reasons only one child will be generated each time. Clearly, the higher the value of the replication probability, the higher the number of replicas generated during the advertising process. A high value of the replication probability, close to one, would result in a scheme that generates a large number of walkers, resembling the operation of flooding. A value of the replication probability that equals zero would result in a scheme with no replications, thus resulting in the single random walker scheme. The replication probability considered here is not fixed but decreases exponentially after each successful replication event, in order to stimulate the generation of the replicas in the earlier stages of the advertising process, when the available advertising budget of the "mother" walker is relatively high and, thus, both the "mother" and the "child" walkers will have enough steam (advertising budget) to carry on and, thus, have better chances to cover diverse regions (compared to exhausting their budgets after a few hops). In view of the above, it appears that the class of RRRWs could be able to "fill" the performance space between the performance induced by flooding and the SRW.

The rest of this paper is organized as follows: in Section II some widely-used techniques for information dissemination are presented. In Section III the general networking environment is described along with a wider notion of network coverage and other important definitions. The introduced RRRWs scheme is described in Section IV. In Section V the various network topologies considered in this paper are described and the performance evaluation results are presented. Finally, some conclusions are drawn in Section VI.

II. RELATED WORK

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 ([1], [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 (achieving a very small completion time), the associated large message overhead is a major drawback. In order to reduce the (unacceptably large) overhead induced by flooding, *probabilistic flooding* ([3], [4], [5], [6], [7]) 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 a small increase in the completion 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 needed to cover the network and the completion time increase with K. If the value of K is very large, then the K-hop controlled flooding scheme would approach the traditional flooding approach.

A popular alternative to flooding, for service advertising in an unstructured environment, is the *Single Random Walker (SRW)* ([8], [9], [10]). In SRW, the initiator node will employ a single agent that will move randomly in the network, one hop/ node per time slot, informing all the nodes in its path. The message overhead of the SRW is considered to be much smaller than that of flooding approaches, at the expense of a significant increase in the completion time, which is now closely related to the number of messages (i.e., available advertising budget).

Flooding and SRW are considered as two rather "extreme" approaches. On one hand, the use of several parallel agents (which work independently) under flooding results in flooding's smaller completion times. On the other hand, the single agent (moving in a sequential manner) under the SRW, results in a rather slow dissemination of the information and relative large completion times. It should also be noticed that the lack of "coordination" between the numerous "parallel agents" under flooding tends to increase (unnecessarily) the induced message overhead for achieving a certain network coverage, as these agents may try to cover overlapped areas. At the same time, although the SRW's single agent, is still probable of revisiting areas already covered, it is less likely to do it so in the "early stages" of the process when the already covered region will still be small.

In view of the above discussion, the use of "multiple" random walkers, that are born probabilistically one at a time and *during* the advertising process, is considered in this paper in order to trade effectively message overhead (large under flooding) with completion time (large under SRW). It is expected that the existence of more than one agents, born at different times and in, potentially, different networking areas can help disseminate the advertising information "deeper" in the network, as the various walkers are anticipated to cover probabilistically different networking areas. Alternatively, the various random walkers could all be created at the beginning of the information dissemination process yielding the *k-simultaneous random walkers* scheme (where k is the number of initial entities) ([11], [12]). The simultaneous creation of all random walkers (RWs) is expected to lead to a behavior (depending strongly on the value of k) similar to flooding, as the RWs will spend a fraction of their available budget (overhead) and time over largely overlapping areas, before moving to more diverse ones; during this period revisits to already visited (by some RWs) nodes are expected to be high. Under the RRRWs scheme, outlined earlier, the previous problem is avoided by keeping the multiplicity of the RWs generated each time to the minimum (only two, the "mother" and the "child"), and the times of replications spread in time to increase the likelihood that the next replication will happen

in an area not likely to be mostly already covered by the previously replicated RWs. With the RRRWs, we basically attempt to engage more agents in the advertising process (as under flooding), trying to avoid at the same time the intrinsic problems of the flooding process. The latter arguments can be even more appropriate if one views them in the context of the L-coverage of the network introduced next.

III. L- COVERAGE

Let an undirected connected graph $G(V, E)$ represent a networking topology consisting of a set of V nodes and a set of bidirectional edges E, among the nodes. Let $d(u, v)$ denote the distance, in number of nodes/ hops over the shortest path between the nodes u and v; of course $d(u, u) = 0$.

Consider an advertising process that is initiated from the service node (s). The advertising information is disseminated through the forwarding of the (information) message over the network edges, informing the nodes attached to them. The edges involved in the spreading of the advertising information define the set E_a ($E_a \subseteq E$) and the informed nodes define the set V_a ($V_a \subseteq V$). The graph $G_a(V_a, E_a)$ is the part of the $G(V, E)$ graph that was used for the advertising purposes and, therefore, it will be referred to as the *advertising network*. As it can be easily seen, the advertising network is a connected network by construction.

Depending on the employed advertising scheme, different advertising networks may be created. The efficiency of an advertising scheme can be measured by the number of informed nodes (i.e., $|V_a|$, the size of the advertising network); the higher the value of $|V_a|$, the more efficient the advertising scheme. Apart from this metric, the efficiency of an advertising scheme will be measured, in this work, in terms of its capability of *stretching* the advertising information in broader areas of the network and, therefore, bringing it as close as possible to as many nodes as possible. In order to define and evaluate the stretchiness, the notion of *L-coverage* is introduced here.

Let the L-property of a network node u be defined as the existence of at least one informed node v (i.e., $v \in V_a$) at a distance of *at most* L hops away from u. Let the *L-coverage* (or *coverage*) of an advertising scheme, denoted as $C(L)$ be defined as the proportion (%) of the network nodes for which the L-property is satisfied. A high value of $C(L)\%$ for an advertising network would suggest good stretching capabilities of the advertising scheme that created it. For example, consider two different advertising networks (e.g., $G_a(V_a, E_a)$ and $G'_a(V'_a, E'_a)$). If the L- coverage of G_a is better than that of G'_a (i.e., $C_a(L) > C'_a(L)$), then the advertising network $G_a(V_a, E_a)$ is said to be more stretched than the advertising network $G'_a(V'_a, E'_a)$, since the advertising network $G_a(V_a, E_a)$ is such that its L-hop neighborhood is larger, implying that its nodes are closer to more network nodes under the L-property notion.

Notice that if the L_1-property is satisfied for a node then, the L_2-property (for any $L_2 > L_1$) is also satisfied for this node. If the 0-property is satisfied for a node, then this node is an informed one and part of the advertising network (i.e.,

$u \in V_a$). If the L-property is satisfied for all network nodes, then $C(L) = 100\%$. Obviously, if $C(0) = 100\%$, then $V = V_a$, that is all the nodes are part of the advertising network. Finally, it should be noted that $C(L)$ is an increasing function of L; that is, $C(L_1) \geq C(L_2)$ for $L_1 > L_2 \geq 0$.

Referring also to the discussion in Section I, it should be noted that the notion of L-coverage is not of academic value only but has practical significance. A higher L-coverage implies that more of the uninformed nodes can obtain the information through an L-hop flooding, while a lower L-coverage implies that more of the uninformed nodes will have to launch a larger than L-hop, thus more expensive, flooding in order to obtain the information. Therefore, an advertising scheme that induces higher L-coverage is considered to be more effective.

IV. RANDOMLY REPLICATED RANDOM WALKERS (RRRWs)

The RRRWs scheme introduced in this paper is presented in this section. At the beginning of the advertising process a single random walker/ entity (e_0) is created at the service node (s). This entity will start moving in the network, disseminating the service location information to all the visited nodes, creating an advertising network. The random walker will move to one node in each time slot and this movement will correspond to one message (one unit of message overhead or of its available advertising budget).

Assume that entity e has reached node u. The choice of the next hop node to be visited will be made randomly over the set of the neighbor nodes of node u (S_u), excluding the node that was visited by entity e at the very previous time slot. In other words, a movement restriction rule is applied, according to which the entity keeps a one hop memory of the nodes in its path in order to prevent itself from revisiting the very previously visited node, increasing potentially the probability of reaching out to and informing nodes in new network areas (stretching). Clearly, this movement restriction rule does not prevent a revisit to the previously visited node in the future or by another entity. An exception to the aforementioned movement restriction rule is applied when the node where the entity resides has only one neighbor (i.e., $|S_u| = 1$). Only then the entity is allowed not to follow the movement restriction rule to avoid a certain deadlock and a premature termination of the random walk (see Section V).

The initial entity e_0 is assigned an advertising budget of H hops to spend (corresponding to H messages and H time slots for this single entity). This budget will be reduced by one hop for every move of an entity. When, according to the replication policy defined below, an entity (e_x) must split and create only *one* replica of itself (e_y), then the remaining advertising budget of the entity at the time of replication will be equally shared (or almost equally since h_e is an integer) among the "mother" (h_{e_x}) and the "child" entity (h_{e_y}).

The number of entities that will be born during the advertising process depends on the applied replication policy. When the number of entities is large then, because of the

sharing of the advertising budget at each replication, the available budget for each replica would be small. As a result, the random walkers might get "stuck" in small localities – depending on the networking topology– and lead to a small network coverage. On the other hand, a smaller number of entities can lead to better stretching capabilities because the entities could use more efficiently their –potentially– larger available advertising budget. In addition, the completion time would be decreased with an increase in the number of entities, due to the aforementioned sharing of the advertising budget at each split. Clearly, the replication process should be designed carefully, in order to be possible to shape the performance characteristics according to the desired objectives.

A reasonable approach to follow, is to apply a replication policy that will allow for frequent replications at the beginning of the advertising mechanism and to significantly reduce the number of replications in the latter stages, where the entities will not have enough "steam" to explore new networking areas. The replication policy presented in the sequel employs the aforementioned approach.

A. Replication Policy

The replication policy presented here, aims at generating more frequent replications at the beginning of the advertising process and less frequent afterwards. More specifically, the replication policy will allow an agent e to replicate itself according to an exponentially decreasing *replication probability*,

$$p(k_e) = p^{2^{k_e}} \tag{1}$$

where p is a constant probability that is equal to the probability that the first replication takes place and k_e is the number of replications that have taken place in the past for entity e and all its parent entities. For the initial entity e_0, $k_{e_0} = 0$. Each time a replication takes place only *one* new entity (i.e., "child") will be created.

It can be easily seen, that according to equation (1) an increase in the number of replications would result in a sharp decrease in the value of the replication probability $p(k_e)$. This particular policy is easily applied, since the only requirement is to maintain information regarding k_e along with entity e and to make it available to its child entities (if any).

When $p = 0$ (RRRW($p = 0$)), then no splitting will take place and the approach will resemble to an enhanced SRW with one hop memory, since only the initial entity will carry out the advertising (completion time of H time slots). On the other hand, when p is close to 1, then the number of replicas that are born is large, resulting in a significant decrease in the completion time, closely resembling the operation of flooding. The difference between RRRWs with $p = 1$ and flooding resides in the fact that the number of entities born when a replication takes place is only one and that the number of replications is exponentially decreased over time.

A main interest in this work is in exploring whether the class of RRRWs (whose members are defined for different first replication probabilities) can "fill" the performance space (defined in terms of specific performance metrics to be introduced in the following section) that lies between the performance induced by flooding and the SRW (i.e., RRRW($p = 0$) here). For the rest of this paper the terms RRRWs and RRRW(p) (for $0 \le p \le 1$) will be used interchangeably.

V. PERFORMANCE EVALUATION

In this section, the experimental performance comparison of the introduced RRRWs scheme (which includes the SRW) with flooding is presented over various networking environments. In Subsection V-A the methodology of the experiments along with the performance metrics are presented, by a presentation of the examined topologies followed in Subsection V-B. Finally, in Subsection V-C the results of our simulations are presented.

A. Methodology

A simulation program written in C programming language is used for simulating various random graph topologies that are presented later. Each topology consists of 10.000 nodes, some of which are randomly selected to play the role of the service node. For each run of an advertising scheme only one service node is present and this node initiates the advertising process. The final results, to be presented, are the averages values of the results taken by running each advertising scheme twice for each service node.

In the following experiments, the main interest is in exploring the coverage achieved by each examined advertising scheme under the different networking topologies, for a certain advertising budget H, along with the associated completion time for the same value of H. Since the interest, here, is not only in the size of the generated advertising networks, but also in the stretching capabilities of the various advertising schemes, the L- coverage of the network will be examined for different values of L (e.g., $L = 0, 1, 2$). The following quantities are considered in the experimental evaluation:

L- Coverage: The L- coverage of the network is the performance metric that captures the size of the generated advertising networks ($L = 0$), along with the stretching capabilities of the employed advertising approach. It measures the proportion (%) of the network nodes for which the L-property is met. The performance for coverage for all the information dissemination approaches will be examined for different L values. It is expected that the performance of the RRRWs scheme ($0 < p \le 1$) with respect to the L-coverage will be better than that of flooding and the SRW (RRRW($p = 0$)).

Completion Time: The completion time of an advertising process is equal to the number of time slots that are needed for a process to exhaust all the available advertising budget H and terminate. In order to ensure that all the available advertising budget H is exhausted before the advertising process terminates, the exception of the movement restriction rule when an entity visits a node with only one neighbor was introduced. Furthermore, to measure the completion time, it is considered that the nodes process and forward all the messages

that have been received in the very previous timeslot in one time slot.

The completion time of the RRRW($p = 0$) is equal to the available advertising budget H and is expected to be significantly larger than the completion time of flooding. At the same time, the completion time of the RRRWs scheme (for $0 < p \leq 1$) is expected to be somewhere between that of the two aforementioned approaches.

Advertising Budget H: The advertising budget H is the available number of messages for the advertising process. This budget is shared among the entities when a replication takes place. The coverage and the completion time of the advertising schemes are examined for various values of H. An increase in their performance is expected with an increase in H.

First Replication Probability p: The value of the first replication probability p will shape the information dissemination mechanism, since it controls the number of entities that will be born during the advertising process. The performance of the RRRWs approach, regarding coverage and the completion time, is examined for different values of the probability p and, it is compared to that of flooding and RRRW($p = 0$).

B. Topologies

Several random graph topologies are considered in this work: Random Geometric, Clustered topologies and Erdös-Rényi (binomial) Graphs. In addition to being viewed as realistic network models, the Random Geometric Graphs capture better the sense of the geographical coverage of the network, facilitating the understanding and examination of the notion of stretching. The Clustered environment is examined due to its capability of modeling adequately the structure of a modern communication network, such as many modern Peer-to-Peer (P2P) and social networks. Finally, the Erdös- Rényi (binomial) Graphs will be examined for comparison reasons.

Random Geometric Graphs: Random Geometric Graphs are constructed by dropping n nodes randomly uniformly over a square area (i.e., $1m^2 * 1m^2$ in our case) and adding edges to connect any two nodes whose distance is at most r. The higher the value of r, the larger the number of links and the more connected the resulting network will be, also exhibiting a larger average degree.

In our example, the value of r was 0.023 leading to a networking environment of 81.313 links and an average degree of 16.

Clustered Topologies: These topologies are used to model real networks consisting of a small number of well-connected, relatively isolated clusters that are connected with each other through some representatives from each one. The number of links between the clusters is limited, while the clusters inside exhibit very good connectivity.

In particular, in our experiments, the network will be composed of a small number of clusters, namely 5 clusters, and each one would be connected by means of a d-regular graph (i.e., d-regular graphs are used to model the demanding good connectivity inside the clusters).There will be clusters of size of 2000 nodes each and 20 representatives from each

cluster will be chosen to connect the clusters between them by means of a d'-regular graph (i.e., $d > d'$). Two different topologies are considered: in the first topology $d = 6$ and $d' = 2$ (forming 30.100 links), while in the second $d = 10$ and $d' = 2$ (forming 50.100 links).

Binomial Graphs: The binomial graphs were initially introduced by Erdös and Rényi. They are based on a simple construction model. Let's consider n nodes. Then a link between each one of them is created with a probability p_e ($0 \leq p_e \leq 1$) and is not with probability $(1 - p_e)$. The total number of links present in the graph is equal to $p_e \cdot n \cdot (n-1)/2$. As the value of p_e increases, so does the number of links forming small isolated islands, at the beginning, until some of these islands merge to form a *giant component*. When all the isolated islands become a part of this component, a connected network is created.

In the binomial graphs, the geographic sense of "neighborhood" is not present since the links are formed probabilistically and therefore a node could be arbitrary close to any other node no matter what its position is in the network; the latter makes the examination of such topologies, at least with respect to stretching, interesting.

For our evaluation purposes, a binomial graph of average degree 12 and 60.253 links is created.

C. Simulation Results

Here, the simulation results concerning the performance of flooding and RRRWs, for the topologies described above are presented. Tables I- XII show the performing results regarding the L-coverage, for various values of L and H and for the different topologies. Tables XIII- XVI show the results for the completion time for each topology for the same advertising budgets. In each Table the first column depicts the applied advertising scheme.

1) Random Geometric Graphs: In Random Geometric Graphs, the network coverage is generally slightly decreased with the first replication probability p (especially for large values of the advertising budget), while the completion time is decreased with p, as expected. RRRWs outperform flooding with respect to network coverage for all the examined values of p. This is the case even for $p = 1$ –in which case the completion time of the RRRWs approach is the smallest– for small advertising budgets. Furthermore, RRRWs outperform RRRW($p = 0$) with respect to network coverage for certain advertising budgets and small values of the first replication probability ($p = 0.05, 0.1$), decreasing the completion time as well (e.g., $H = 2000, 9000$).

Tables I-III and Table XIII show the results for coverage and completion time respectively, for the Random Geometric Graphs. From these results it can be seen that RRRWs outperform flooding with respect to L-coverage for all the examined values of L; the performance difference increases with H (except when $p = 1$, where RRRWs outperforms flooding only for $H < 2000$). The price paid for the latter increase is an increase in the induced completion time.

2) Clustered Topologies: In Tables IV-VI and Tables XIV, the results regarding network coverage and completion time, respectively, are presented, for the case where the structure inside each cluster is a 10-regular graph. Once again, the network coverage is slightly decreased with p (especially for $p \geq 0.1$), while the completion time is decreased with p, as expected. RRRWs outperform flooding for $p \leq 0.8$, but the advertising budget for which this better coverage performance is achieved increases with p. For example, RRRW($p =$ 0) outperforms flooding for all advertising budgets, while RRRWs($p = 0.8$) outperforms flooding for $H \geq 3000$. This is noticed because as p increases, so does the number of entities present, thus demanding more budget to widely explore the network. On top of this, RRRWs($p \leq 0.8$) significantly outperform flooding regarding stretching for the aforementioned advertising budgets (e.g., around 60% for $L = 1$ and 40% for $L = 2$, for $H = 9000$) . Furthermore, RRRWs($p \leq 0.1$), also manage to outperform RRRW($p = 0$), for certain advertising budgets, and combine this performance with a smaller completion time as well. The same results hold for the other clustered environment, with the observation that flooding is outperformed for smaller advertising budgets (i.e., the clusters have smaller connectivity, making it easier for a walker to move to not overlapping network areas). They are presented in Tables VII-IX and Table XV.

From the results described above, it can be stated that RRRWs outperforms RRRW($p = 0$) for small values of p, both in network coverage and completion time, for certain advertising budgets. At the same time, RRRWs manages to outperform flooding regarding coverage, but this improved performance comes with a larger than in the Geometric Graphs increase in completion time.

3) Binomial Graphs: In the binomial graphs the decrease in the performance of coverage is not so noticeable, as in the other topologies, for an increase in p, while the decrease in the completion time is still noticeable. RRRWs's network coverage performance is similar to that under flooding and RRRW($p = 0$), for most of the examined values of p (i.e., $p \leq 0.8$), providing much smaller completion times compared to the latter or reducing the difference compared to the former.

As a final comment to the results that can be found in Tables X-XII and Table XVI, it can be said that flooding's network coverage performance is increased, compared to the other examined topologies, but RRRWs manage to perform similarly well even for large values of p (i.e., $p \approx 0.8$), thus reducing the difference from flooding in the completion time and increasing it compared to RRRW($p = 0$). Despite this fact, flooding's smaller completion time make it a better approach for information dissemination in binomial graphs.

VI. CONCLUSIONS

In this paper, a thorough investigation of the performance of various information dissemination schemes is presented, with respect to various performance metrics and over various popular network topologies. In addition to flooding and the Single Random Walker (SRW), a newly introduced class

TABLE I
RANDOM GEOMETRIC GRAPH, r=0,023: $L = 0$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	0.5	0.86	2.25	3.31	5.04	8.56
RRRW(p=0)	1.38	3.11	9.62	14.87	22.33	32.33
RRRW(p=0.05)	1.39	3.07	10.25	14.17	21.88	33.85
RRRW(p=0.1)	1.28	2.86	9.73	14.04	20.23	33.23
RRRW(p=0.3)	1.29	2.89	9	13	18.52	30.99
RRRW(p=0.5)	1.25	2.58	8.32	11.54	17.58	28.01
RRRW(p=0.8)	1.03	2.24	6.6	9.06	13.07	21.47
RRRW(p=1)	0.85	1.24	2.09	2.47	3.07	3.36

TABLE II
RANDOM GEOMETRIC GRAPH, r=0,023: $L = 1$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	1.09	1.55	3.22	4.46	6.36	10.24
RRRW(p=0)	3.88	7.4	17.66	28	39	49.97
RRRW(p=0.05)	3.62	7.33	20.29	26.28	38.12	52.83
RRRW(p=0.1)	3.46	6.61	18.56	25.58	34.71	53.3
RRRW(p=0.3)	3.54	6.67	16.62	23.3	30.71	48.17
RRRW(p=0.5)	3.3	5.43	15.29	19.74	28.28	42.67
RRRW(p=0.8)	2.59	4.66	11.88	15.17	21.23	31.66
RRRW(p=1)	1.91	2.44	3.54	3.92	4.71	5.1

TABLE III
RANDOM GEOMETRIC GRAPH, r=0,023: $L = 2$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	1.77	2.35	4.42	5.73	7.78	12.03
RRRW(p=0)	5.55	10.13	21.52	33.91	46.15	56.04
RRRW(p=0.05)	5.17	9.84	25.31	31.77	44.88	59.5
RRRW(p=0.1)	4.89	8.93	22.89	31.18	40.97	61.08
RRRW(p=0.3)	5.17	8.89	20.45	27.89	35.97	54.73
RRRW(p=0.5)	4.71	7.25	19.06	23.48	33.03	48.36
RRRW(p=0.8)	3.78	6.29	14.72	18.6	25.37	36.17
RRRW(p=1)	2.93	3.51	4.83	5.2	6.09	6.57

TABLE IV
CLUSTER GRAPH 10-REG INSIDE EACH CLUSTER: $L = 0$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	1.92	4.52	13.66	17.16	20.52	22.47
RRRW(p=0)	1.92	4.54	15.27	21.99	34.31	53.74
RRRW(p=0.05)	1.92	4.5	15.78	21.58	33.52	52.19
RRRW(p=0.1)	1.92	4.58	15.52	21.29	33.58	52.02
RRRW(p=0.3)	1.9	4.44	14.64	19.11	30.45	49.65
RRRW(p=0.5)	1.9	4.46	14.18	20.03	29.25	42.25
RRRW(p=0.8)	1.87	4.41	13.5	18.12	23.8	34.8
RRRW(p=1)	1.73	4.01	11.72	15.09	18.61	20.24

TABLE V
CLUSTER GRAPH 10-REG INSIDE EACH CLUSTER: $L = 1$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	12.35	19.36	24.26	25.59	26.83	30.17
RRRW(p=0)	12.57	21.29	43.57	57.5	79.52	92.31
RRRW(p=0.05)	12.37	20.78	49.66	55.84	76.26	93.45
RRRW(p=0.1)	12.41	22.6	46.33	55	77.47	92.38
RRRW(p=0.3)	12.47	19.09	41.13	46.32	72.15	91.01
RRRW(p=0.5)	12.1	20.23	36	51.81	69.93	85.45
RRRW(p=0.8)	11.9	18.46	29.95	40.77	52.63	76.44
RRRW(p=1)	11.19	17.73	21.19	22.72	25.73	25.87

TABLE VI
Cluster Graph 10-Reg Inside Each Cluster: $L = 2$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	23.37	26.31	33.07	34.12	42.17	55.62
RRRW(p=0)	25.81	30.93	55.13	67.26	89.18	96.92
RRRW(p=0.05)	24.67	31.16	65.78	69.82	85.53	99.68
RRRW(p=0.1)	24.24	34.23	61.72	71.44	88.15	97.18
RRRW(p=0.3)	23.86	26.64	56.04	64.71	88.2	96.96
RRRW(p=0.5)	22.68	26.59	54.88	64.19	85.29	95.63
RRRW(p=0.8)	21.25	25	51.84	67.58	77.82	93.4
RRRW(p=1)	20.66	24.02	28.95	34.9	43.42	48.99

TABLE VII
Cluster Graph 6-Reg Inside Each Cluster: $L = 0$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	1.94	4.57	14	17.44	21.77	28.61
RRRW(p=0)	1.93	4.54	16.39	23.28	35.85	53.68
RRRW(p=0.05)	1.91	4.57	16.28	23.07	36.22	53.94
RRRW(p=0.1)	1.91	4.53	16.03	22.91	35.25	53.15
RRRW(p=0.3)	1.9	4.52	15.74	22	32.51	52.48
RRRW(p=0.5)	1.87	4.47	14.78	20.33	31.96	50.86
RRRW(p=0.8)	1.85	4.36	14.33	18.02	25.3	39.47
RRRW(p=1)	1.57	3.5	10.64	14.12	17.34	20.75

TABLE VIII
Cluster Graph 6-Reg Inside Each Cluster: $L = 1$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	8.07	14.98	23.12	24.11	28.92	51.16
RRRW(p=0)	8.3	15.97	46.41	58.74	77.31	89.35
RRRW(p=0.05)	8.12	16.5	45.1	57.4	79.38	92.16
RRRW(p=0.1)	8.24	16.26	43.47	56.73	75.85	92.59
RRRW(p=0.3)	8.06	15.87	41.23	52.02	69.03	90.61
RRRW(p=0.5)	7.97	15.3	35.5	45.75	66.97	87.75
RRRW(p=0.8)	7.85	14.71	32.11	34.57	49.78	74.29
RRRW(p=1)	6.79	12.62	20.61	22.06	23.85	25.95

TABLE IX
Cluster Graph 6-Reg Inside Each Cluster: $L = 2$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	19.41	23.47	31.39	34.96	46.66	86.44
RRRW(p=0)	21.79	28.39	69.88	79.51	92.72	96.97
RRRW(p=0.05)	21.05	31.07	66.37	77.81	96.15	99.07
RRRW(p=0.1)	21.48	30.9	67.66	78.58	93.09	99.95
RRRW(p=0.3)	20.15	28.32	64.49	73.87	90.54	99.06
RRRW(p=0.5)	19.53	25.88	56.6	71.78	89.08	97.32
RRRW(p=0.8)	21.15	23.92	53.99	59.52	81.42	95.45
RRRW(p=1)	18.59	22.05	25.28	29.16	30.15	40.29

TABLE X
$G_{n,p}$ Graph with n=10.000,p= 12/n: $L = 0$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	1.99	4.9	18.16	26.11	39.58	59.72
RRRW(p=0)	2	4.89	17.99	25.61	38.72	57.97
RRRW(p=0.05)	2	4.88	17.96	25.7	38.72	57.98
RRRW(p=0.1)	2	4.89	17.97	25.66	38.83	57.94
RRRW(p=0.3)	2	4.86	18	25.71	38.73	58
RRRW(p=0.5)	1.98	4.89	17.96	25.54	38.69	57.95
RRRW(p=0.8)	1.97	4.84	17.93	25.6	38.65	58.01
RRRW(p=1)	1.83	4.39	16.39	24.01	36.16	54.42

TABLE XI
$G_{n,p}$ Graph with n=10.000,p= 12/n: $L = 1$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	21.36	44.78	88.75	95.6	99.14	99.93
RRRW(p=0)	21.24	44.59	88.63	95.43	99.06	99.92
RRRW(p=0.05)	21.37	44.41	88.48	95.52	99.08	99.91
RRRW(p=0.1)	21.33	44.48	88.56	95.48	99.06	99.91
RRRW(p=0.3)	21.21	44.31	88.37	95.47	99.06	99.9
RRRW(p=0.5)	21.37	44.67	88.66	95.4	99.06	99.91
RRRW(p=0.8)	21.18	44.22	88.41	95.45	99.04	99.92
RRRW(p=1)	19.93	41.29	86.18	94.43	98.73	99.87

TABLE XII
$G_{n,p}$ Graph with n=10.000,p= 12/n: $L = 2$

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	92.32	99.55	100	100	100	100
RRRW(p=0)	92.17	99.54	100	100	100	100
RRRW(p=0.05)	92.28	99.51	100	100	100	100
RRRW(p=0.1)	92.22	99.53	100	100	100	100
RRRW(p=0.3)	92.11	96.69	100	100	100	100
RRRW(p=0.5)	92.35	99.55	100	100	100	100
RRRW(p=0.8)	92.11	99.52	100	100	100	100
RRRW(p=1)	90.0	99.31	100	100	100	100

TABLE XIII
Completion Time: Random Geometric Graph, r=0,023

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	2.3	3.2	5.2	6.3	7.9	10.6
RRRW(p=0)	200	500	2000	3000	5000	9000
RRRW(p=0.05)	106.7	255.6	855.6	1041.6	1549.9	2615.5
RRRW(p=0.1)	98.4	198.1	593.5	855.7	1342	2334.3
RRRW(p=0.3)	63.5	132.8	390.9	513.1	751.2	1246.1
RRRW(p=0.5)	42.8	89.3	274.2	389.7	588	872.7
RRRW(p=0.8)	25.3	44.9	122.8	163.6	221.8	359.3
RRRW(p=1)	9	10	12	13	14	15

TABLE XIV
Completion Time: Cluster Graph 10-Reg Inside Each Cluster

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	3	3	4	4	4	5
RRRW(p=0)	200	500	2000	3000	5000	9000
RRRW(p=0.05)	112.6	257.1	811	1065.9	1522.7	2538
RRRW(p=0.1)	96.4	208.3	585.1	835.8	1315.3	2333.5
RRRW(p=0.3)	61.9	130.2	400.5	530.2	738.1	1294.5
RRRW(p=0.5)	48.9	84.1	269.4	391.4	613.9	836.2
RRRW(p=0.8)	25.9	42.6	126.9	172.8	222.5	339.4
RRRW(p=1)	9	10	12	13	14	15

TABLE XV
Completion Time: Cluster Graph 6-Reg Inside Each Cluster

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	3.9	4	5	5	5.9	7
RRRW(p=0)	200	500	2000	3000	5000	9000
RRRW(p=0.05)	109.2	257.9	815.1	1055.6	1536.5	2533.8
RRRW(p=0.1)	94.2	185	586.1	853.4	1339	2321.3
RRRW(p=0.3)	60.1	129.9	402	510.2	790.6	1286
RRRW(p=0.5)	43.6	83.5	273.5	378.5	598	913.5
RRRW(p=0.8)	26.3	43.9	125.8	165.6	231	351.4
RRRW(p=1)	9	10	12	13	14	15

TABLE XVI
COMPLETION TIME: $G_{n,p}$ GRAPH WITH n=10.000,p= 12/n

H	$C(L)$ (%)					
	200	**500**	**2000**	**3000**	**5000**	**9000**
Flooding	2.8	3	3.7	4	4	4
RRRW(p=0)	200	500	2000	3000	5000	9000
RRRW(p=0.05)	107.9	247.6	810.7	1043.7	1516.3	2544.8
RRRW(p=0.1)	101.8	203.3	562.7	830.2	1337.3	2344
RRRW(p=0.3)	61.6	131.8	376.5	516.9	760	1252.5
RRRW(p=0.5)	46.1	89.1	269	390.2	611.1	904.7
RRRW(p=0.8)	25.7	44.3	124.2	166.2	229.7	352.6
RRRW(p=1)	9	10	12	13	14	15

of Randomly Replicated Random Walkers (RRRWs) is also considered in the study. This class includes the SRW approach, as well as it can be parameterized so that it can exhibit a performance approximately close to flooding's. Consequently, RRRWs can "fill" the performance space between the performance of flooding and SRW, thus offering a richer set of information dissemination strategies that can better match desirable performance constraints or objectives.

In RRRWs the random walker's replications take place according to an exponentially decreasing probability, so that the number of random walkers generated is higher at earlier stages than in later, which presents advantages as explained in the paper; this probability can also be selected so that the number of generated random walkers is controlled. When the (first) replication probability is close to one, the number of random walkers that are born is large and the resulting RRRWs scheme can resemble the operation of flooding. On the other hand, when the (first) replication probability is zero then no additional random walker is born and, thus, the behavior of the resulting RRRWs scheme is the same as that of SRW.

The various information dissemination approaches are examined regarding the number of the network nodes that are informed at the end of the advertising process (i.e., the size of the advertising network) and regarding their capability for stretching the information dissemination, bringing it close to as many nodes as possible; the stretching capabilities are captured through the introduced notion of $L - coverage$. The performance metrics used in our simulations are the L-coverage achieved for a given message-forwarding budget and the completion time required to spend the available message-forwarding budget.

The simulation results reveal that RRRWs outperform flooding, in Random Geometric and Clustered environments, for most of the examined values of the parameter p of the exponentially decreasing first replication probability (i.e., $p \leq 0.8$ for Clustered and $p \leq 1$ for Random Geometric) both regarding the size of the generated advertising networks and the stretching capabilities shown, at the expense of an increase in the completion time. In the above environments, RRRWs manage also to outperform RRRW($p = 0$) both in coverage and in the completion time, for small values of p (e.g., $p = 0.05, 0.1$) and for certain advertising budgets. Furthermore, the value of p can be selected to yield a performance that better matches the desirable one, or respect certain constraints. In

Erdös- Rényi (binomial) graphs the performance of RRRWs is similar to the corresponding performance of flooding and RRRW($p = 0$), even for large values of p. But, even though the difference in the completion time, compared to RRRW($p = 0$), is large, flooding's speed is still better. As a result, flooding seems to be preferable in the binomial graphs.

REFERENCES

[1] A.Segall, "Distributed network protocols", *IEEE Transactions on Information Theory*, vol.IT-29, pp.23-35, 1983.
[2] B.Williams, T.Camp, "Comparison of broadcasting techniques for mobile ad hoc networks", MOBIHOC 2002, pp.194-205, 2002.
[3] V. Dimakopoulos and E.Pitoura, "On the performance of flooding-based resource discovery", *IEEE Transactions on Parallel and Distributed Systems*, 17(11):290-297, November 2006.
[4] K.Oikonomou and I.Stavrakakis, "Performance analysis of probabilistic flooding using random graphs", in *The First International IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC 2007)*, Helsinki, Finland, 18 June, 2007.
[5] Y.Sasson, D.Cavin, and A.Schiper, "Probabilistic broadcast for flooding in wireless mobile ad hoc networks", in *Swiss Federal Institute of Technology (EPFL), Technical Report IC/2002/54*, 2002.
[6] D.Tsoumakos and N.Roussopoulos, "Adaptive probabilistic search for peer-to-peer networks", in *3rd IEEE International Conference on P2P Computing*, 2003.
[7] M.Bani Yassein, M.Ould-Khaoua 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.
[8] C.Gkantsidis, M.Mihail and A.Saberi, "Random Walks in Peer-to-Peer Networks", *in Proceedings of IEEE INFOCOM*, 2004.
[9] C.Gkantsidis, M.Mihail and A.Saberi, "Hybrid Search Scemes for Unstructured Peer-to-Peer networks", *in Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05), vol. 3, pp.1526-1537, Miami, Fla, USA*, March 2005.
[10] N.Alon, C.Avin, M.Koucky, G.Kozma, Z. Lotker, M.Tuttle, "Many Random Walks Are Faster Than One", *ArXive-prints. Vol. 705*, May 2007.
[11] R.Dorrigiv, A. Lopez-Ortiz and P.Pralat, "Search Algorithms for Unstructured Peer-to-Peer Networks", *in LCN '07: Proceedings of the 32nd IEEE Conference on Local Computer Networks*, pp.343-352, 2007.
[12] S. Dhillon and P. Van Mieghem, "Searching With Multiple Random Walk Queries", *in The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)*, pp. 1-5, Athens, September 2007.