Efficiency of search methods in dynamic wireless networks

Gerhard Haßlinger

T-Systems
D-64295 Darmstadt, Germany
gerhard.hasslinger@telekom.de

Thomas Kunz

Dept. of Systems and Computer Engineering
Carleton University, Ottawa, Canada
tkunz@sce.carleton.ca

Abstract – **Search methods in dynamic networks usually cannot rely on stable topology from which shortest or otherwise optimized paths through the network are derived. When no reliable search indices or routing tables are provided, other methods like flooding or random walks have to be considered to explore the network. These approaches can exploit partially available information on network paths, but the search effort naturally increases with the lack of precise paths due to network dynamics. This problem is especially relevant for wireless technology with strict limitation on power consumption.**

We compare the efficiency of random walks and flooding for exploring networks of small to medium size. Several scenarios are considered including partial path information support for search. Transient analysis and a bound is applied in order to evaluate the messaging overhead.

I. INTRODUCTION

Exploration and search methods are generally required to enable services and content retrieval in communication networks. Even in fixed network areas of the Internet where the topology is stable enough to establish standard routing protocols and search engines to locate nodes and information on them, they have to cope with a steadily ongoing changes. But more dynamic network structures are often encountered on large scale, e.g. peer-to-peer (P2P) overlays [6][14] as well as in wireless sensor or mobile ad hoc networks (MANETs) on smaller scale.

A search may refer to users, network nodes, information, content or services of any kind residing on network resources based on identifiers like IP addresses or hash values used in P2P networks. Although a single node is often addressed by a search, this can be extended to a set of nodes denoted as the target node set, each of which is able to respond successfully. Other cases, where several nodes have to be involved to get a result in a production chain or a distributed scheme, are not considered here.

Regarding search methods in unstructured networks, random walks approved to be a promising alternative in general and especially in large scale networks [1][4][4][5][7][11][14][16], where the standard method by flooding may be prohibitive, even with a limited hop count as in the first version of the Gnutella P2P network [13].

In this study we focus on dynamic networks with a planar graph structure as for sensor and mobile networks, where basic random walks are often less efficient than flooding. Several alternatives for routing and search methods are discussed in recent work in this area [3][10][15]. Poor performance of routing protocols in MANETs in general and in highly dynamic environments in particular is shown by [8][12]. Collecting state information to improve routing protocol performance in particular for QoS routing, also leads to a highly inaccurate view of the network state [9]. However, it is demonstrated [2] that random walks essentially benefit even from imprecise and only partially valid information in support of a search or when many nodes in the network are able to respond.

Within the limited space of an extended abstract we focus on the basic case of a planar grid. The analysis method is summarized in section II and evaluation in sections III confirm the promising performance of biased random walk shown in examples in [2] from a more general view. In section IV, a lower bound on the performance is obtained by focusing on the distance to the target leading to a linear chain approach. The result gives a clear insight in the performance tradeoff between biased random walk and flooding.

II. ANALYSIS OF THE EFFICIENCY OF RANDOM WALKS AND FLOODING

The structure of considered communication networks is denoted by a graph $G = (V, E)$ with sets V of nodes and E of edges. The degree $d(a)$ of a node $a \in V$ is the number of edges attached to the node $d(a) = |\{k \mid k \in V, (a, k) \in E\}|$. The minimum node degree is $d_{\min} = \min\{d(a) \mid a \in V\}$.

With regard to 2-dimensional grid networks each node $(x, y) \in V$ is connected to its 4 neighbors $(x–1, y)$, $(x+1, y)$, $(x, y–1)$ and $(x, y+1)$ except for missing neighbors beyond the boundaries for $x = 0$, $x = N$, $y = 0$, and $y = M$.

We follow a random walk through the network as a stepwise process, which proceeds from a node to a neighbor at the next hop. A random walk R of length L is denoted by the series $R = (r_0, r_1, r_2, \ldots, r_L)$ of visited nodes, where an edge $(r_{k–1}, r_k) \in E$ is chosen for the k-th hop ($1 \leq k \leq L$). Usually a random walk chooses its next hop with the same probability among $d(a)$ options

$$\forall\, a, k;\ (a, k) \in E\!:\ p_{ak} = \Pr(r_{n+1} = k \mid r_n = a) = 1 / d(a). \quad (1)$$

The corresponding transition matrix $P = (p_{ak})$ determines a random walk in the network as a Markov process, where network nodes correspond to the states of the underlying Markov chain and edges to transitions from a state to another.

II.A *Transient analysis of random walks*

When the exploration of a network by random walks of predefined length is evaluated using simulation [5][11][15], the results are subject to confidence levels with long simulation runs being required to achieve tight confidence intervals. General analytical bounds for the convergence are valuable to ensure the principle behaviour, but often are not tight [4]. We

follow an alternative classical approach using transient analysis [2][7], which stepwise determines the probabilities $p_m^{(R)}(a)$ of a random walk to enter a network node a at its m-th hop. When the random walk starts at a specific node s, then we have

$$p_0^{(R)}(s) = 1 \quad \text{and} \quad \forall a \neq s: p_0^{(R)}(a) = 0$$

as the initial distribution. In general, any arbitrary initial distribution can be considered. The transient analysis iteratively computes the distribution of the next hop location. From knowledge of $p_m^{(R)}(a)$ the distribution $p_{m+1}^{(R)}(a)$ for the next step is computed by

$$p_{m+1}^{(R)}(a) = \Sigma_{k: (k, a) \in E} \ p_m^{(R)}(k) \ p_{ka} \qquad (2)$$

For the purpose of network search performance, the probability $q_m^{(R)}(t)$ that a node t has been reached during a random walk of length m is the most important measure:

$$q_m^{(R)}(t) = \Pr(\exists j; 0 \leq j \leq m: r_j = t).$$

The probabilities $q_m^{(R)}(t)$ are computed by the same transient step-by-step approach. The only modification required is to introduce an absorbing state at node t, which accumulates the probabilities for visiting that node during the walk. Therefore the transition equations (1) need to be changed only for departures from node t:

$$\forall a, k; (a, k) \in E: q_{ak} = 1/d(a) \ \text{ if } a \neq t;$$

$$\forall k; (t, k) \in E: q_{tk} = 0 \text{ if } k \neq t; \qquad q_{tt} = 1. \qquad (3)$$

$q_m^{(R)}(t)$ is determined by applying equation (2) to the modified transition matrix:

$$q_{m+1}^{(R)}(a) = \Sigma_{k: (k, a) \in E} \ q_m^{(R)}(k) \ q_{ka}. \qquad (4)$$

The random walk defined by equations (3) and (4) has unchanged behaviour at all nodes except for t. Once it has reached t it stays there for ever. Consequently, $q_m^{(R)}(t)$ is monotonously increasing with m. Provided that the graph is irreducible and finite, node t will be reached sooner or later with probability 1 even in periodical cases:

$$q_m^{(R)}(t) \leq q_{m+1}^{(R)}(t); \quad \lim_{m \to \infty} q_m^{(R)}(t) = 1.$$

The computational complexity for the transient analysis due to equations (2-4) is proportional to the number $|E|$ of edges in the network for each hop. Therefore the effort to compute $p_m^{(R)}(a)$ is of the order $O(m|E|)$, which makes transient analysis applicable to large scale networks with millions of nodes.

II.B *Multiple random walks in parallel*
Random walks often can reduce the communication overhead, but they traverse the hops sequentially and thus usually spend much more time than flooding. Multiple random walks in parallel may be applied as a compromise between demands for low delay and overhead. If a random walk is assured to have a success rate of $\sigma < 1$ within k steps, then m random walks of the same type in parallel each with k steps reduce the failure rate from $1 - \sigma$ to $(1 - \sigma)^k$. Thus a success rate of $\sigma = 90\%$ is improving up to 99.999% by using $m = 5$ walks in parallel. A single walk most often achieves this success rate with less than mk steps but with a longer run time.

II.C *Flooding*
In the simplest case, flooding will spread a request from a node to all its neighbors which repeat to contribute into an exhaustive flood covering the entire network. When the same request is received several times at a node from different neighbors, then only the first receipt is forwarded and later ones are discarded. With broadcasting in wireless networks, the number of flooding messages sent is equal to the number of traversed nodes, whereas a node has to send multiple messages for flooding over point-to-point connections and in peer-to-peer networks. The fact that at most one message is sufficient per node, makes flooding more competitive in broadcast environments.

In order to reduce the messaging overhead, flooding is usually done only up to a predefined hop limit h [13]. An appropriate value of h may be derived from knowledge of the network structure and a demand for coverage. Without that knowledge, h may be initialized with a small value and stepwise increased if the previous search radius turned out to be insufficient. In a 2-dimensional grid there are a maximum of $2h^2 + 2h + 1$ nodes within a distance of $\leq h$ hops, if the surrounding of the node is not further restricted by the grid boundary.

When some network nodes have partial information about which of their neighbors are in downstream direction to the target, then the forwarding can be restricted to those nodes, which reduces the messaging overhead accordingly.

III. EVALUATIONS
In order to study the tradeoffs in the performance of flooding and random walks we consider large 2-dimensional grids. The transient analysis approach can easily cope with a size of several hundred in both dimensions. The results are valid for infinite planar grids in the sense that we extend the size until we are sure that the grid boundary has no impact. The start and target node are located in the middle. Note that this is expected to be a worst case scenario in which any finite random walk without bias would miss a single target node with considerable probability.

The search for a single target node is carried out by a biased random walk [2] utilizing partial routing information available in the network. This model simply assumes that a node is able to forward an incoming search request with probability $0 < \rho < 1$ into the direction downstream to the target. With probability $1 - \rho$ the node fails or has no valid information in support of the search due to dynamics, e.g. mobility or churn of nodes in the network.

Assuming that the random walk is currently at a node (a, b) with distance $\Delta = |a - x| + |b - y|$ to the target (x, y), we have to distinguish whether (a, b) is in line with the target, such that $a = x$ or $b = y$, or not. If they are in line, then a unique of the 4 neighbors of (a, b) is closer to the target while three other neighbors have distance $\Delta+1$. The biased random walk chooses each of those three with probability $(1-\rho)/4$ and the closer one with probability $(1 + 3\rho)/4$. Otherwise, if they are not in line then there are two neighbors at distance $\Delta-1$ and two at $\Delta+1$. Both closer neighbors are then chosen with probability $(1+\rho)/4$ each and both farer ones with $(1-\rho)/4$, respectively.

The model surely does not cover all realistic scenarios, since it is left open how a node decides if it has enough and up to date information to direct to the next step to the target. In addition, a homogeneous information distribution is implied, whereas networks structures are often inhomogeneous e.g. hierarchical and more precise information can be assumed near the target.
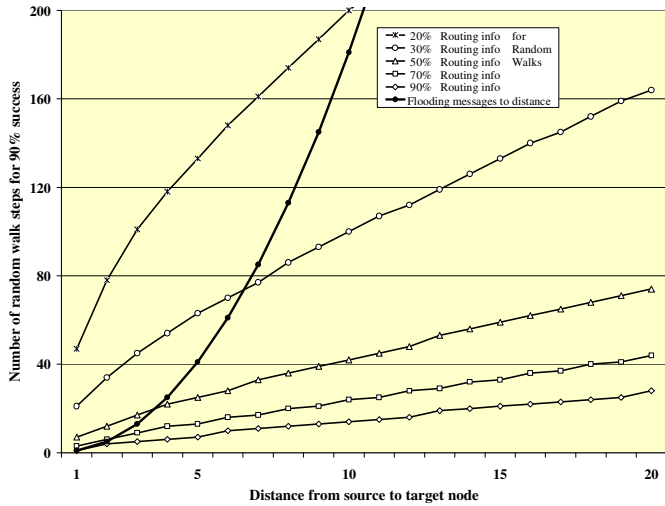

Figure 1: Overhead of biased random walk search for 90% success

Figure 1 shows how many steps are required for a biased random walk to reach the target at hop distances of 1, 2, … 20 with 90% probability. Start and target node are assumed to be located on a common line in the grid. The 3 lower curves for ρ = 50%, 70% and 90% routing information availability show that random walks are efficient as compared to flooding, which is included by the quadratic curve $2\Delta^2 - 2\Delta + 1$ for the number of nodes within distance $\Delta - 1$. This does not reflect the fact that flooding could also benefit from partial routing information, but on the other hand has to try with different hop count limits since the distance Δ is not known a priory.
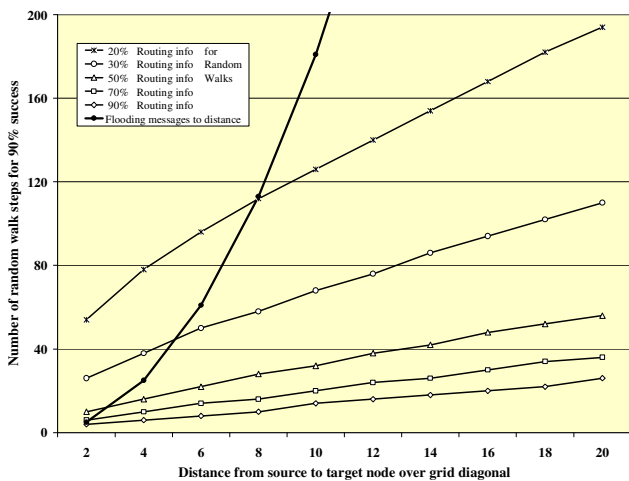

Figure 2: Effort of random walk with target reached over grid diagonal

While the analysis in figure 1 is for targets that can be reached over a line of the grid, we next consider the opposite cases of

targets to be reached over a diagonal in the grid. Figure 2 shows the corresponding results for hop distances up to 20 and the same parameters, revealing an even better performance of the random walks. The improvement is plausible from the fact, that if the current node and the target are on a line, then the distance to the target is reducing for one of four next hops and increasing for the three others. Otherwise, if the target is not in line with the current node, then two of its four neighbors are closer to the target which enables a faster convergence in such steps.
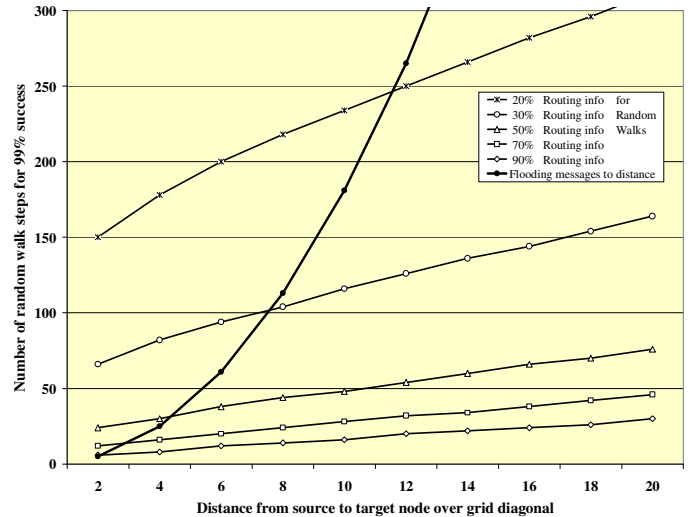

Figure 3: Overhead of biased random walk search for 99% success

Figure 3 gives the corresponding results for a success probability of 99% instead of 90% in figure 2. Note that two independent random walks in parallel can reach this success level at twice the number of steps in total. For small distances, a single random walk often needs more than twice the effort, which makes multiple walks preferable. For larger distances, a single walk reaches 99% success essentially below twice the number of hops required for 90%.
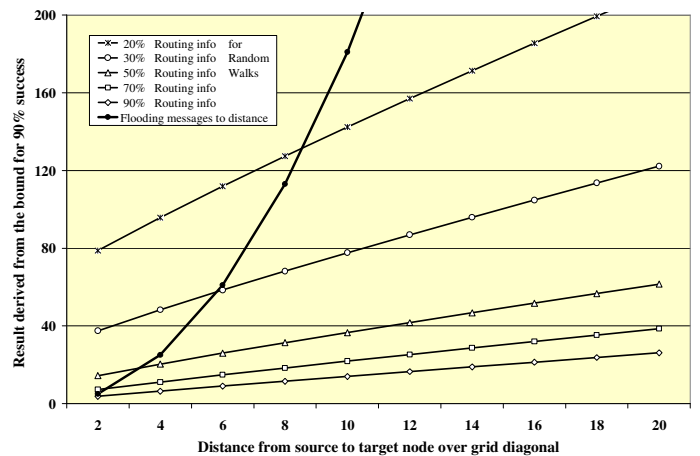

Figure 4: Applying the bound of section 4 to the examples of figure 2

IV. A BOUND ON THE BIASED RANDOM WALK PERFORMANCE

Assuming that a biased random walk decrements the distance Δ to the target in each step with probability q or increments the distance with probability $p < q$ or stays at the same distance with probability $1 - p - q$, the process of approaching the target becomes a simple birth-death process as is well known e.g. for M/M/1 queueing systems. The reduction to a 1-dimensional view is appropriate, when no boundaries have to be regarded in a grid or other network structure.
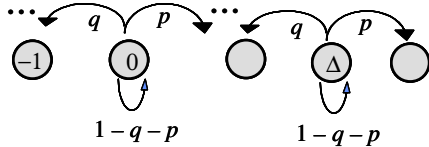

Figure 5: Linear Markov chain approach for the distance to the target

For the transient behaviour of this birth-and-death process we obtain: The number of hops until the next change in the distance Δ is geometrically distributed with parameter $1-p-q$; for 2-dimensional grids the distance changes in each step, i.e. $p + q = 1$. Starting at a distance Δ from the target, the distance Δ_m after m in- or decrements has a binomial distribution:

$$\Pr\{\Delta_m = \Delta + m - 2k\} = \binom{m}{k} \omega^k (1 - \omega)^{m-k} \text{ for } k = 0, \ldots, m, \text{ where}$$

$\omega = q/(p+q)$. This includes states of negative distances, which are only reachable by previously traversing the distance 0. The complete probability mass on states below 0 indicates that the search was already successful, i.e. state 0 should be absorbing.

But we continue the analysis of the non-truncated birth-death chain. The mean $E(\Delta_m)$ of Δ_m is given by $E(\Delta_m) = \Delta - (2\omega -1) m$. For $E(\Delta_m) \leq 0$, i.e. for $m = \lceil \Delta / (2\omega - 1) \rceil$ we conclude that the target has been reached at least with probability 50%, since a binomial distribution is symmetrical and has most of its probability mass in the negative part when the mean is negative. This gives a clear and simple hint on the number m of steps required for a successful biased random walk.

For success probabilities $1-10^{-k}$ we include the variance $\sigma^2(\Delta_m)$ of the distribution in the analysis via Chebychev's bound

$$\Pr\{\Delta_m > E(\Delta_m) + 10^k \sigma(\Delta_m)\} \leq 10^{-2k}.$$

The mean $E(\Delta_m) = \Delta - (2\omega -1)m$ and variance $\sigma^2(\Delta_m) = \omega(1-\omega)m$ of the binomial distribution, leads to a bound on $\Pr\{\Delta_m > 0\}$, from which we derive:

$$m \geq \frac{\Delta}{2\omega - 1} + \frac{10^k \omega(1-\omega)}{2(2\omega-1)^2} \left(1 + \sqrt{1 + \frac{4(2\omega-1)\Delta}{10^k \omega(1-\omega)}}\right)$$

We apply the bound to random walks in a grid for the same cases as depicted in figure 2. In the grid, we obtain

$$q = \Pr\{\Delta_{m+1} = \Delta_m - 1\} = (2+2\rho)/4 \text{ and}$$

$$p = 1 - q = \Pr\{\Delta_{m+1} = \Delta_m + 1\} = (2-2\rho)/4$$

if the current node is not in a line with the target or otherwise $q = (1+3\rho)/4$; $p = (3 - 3\rho)/4$. We ignore the latter case for nodes in line with the target and adopt $q = (2 + 2\rho)/4$, $p = (2 - 2\rho)/4$.

In this way, we get an approximation rather than a bound, which is valid for most of the traversed nodes except for those on both grid lines crossing at the target, Figure 4 compares the evaluations with the transient analyses of figure 2 to the derivation from the bound confirming a good overall match.

CONCLUSIONS AND FURTHER WORK

Using a basic planar grid model, we have shown that random walks can be efficient for search in dynamic networks, especially when partial routing information is still available. Transient analysis and a bound are utilized to analyze such biased random walks.

Among further cases that had to be omitted in the extended abstract are networks with a larger node degree corresponding to a larger radius for reachability by a broadcast message as well as different network structures taken from application scenarios. Target sets consisting of multiple nodes are of interest to model different information levels about a target in its surrounding.

REFERENCES

[1] C. Avin and B. Krishnamachari, The power of choice in random walks: An empirical study, Computer Networks 52 (2008) 44-60

[2] R. Beraldi: Service discovery in MANET via biased random walks, Proc. Autonomics, Rome, Italy (2007)

[3] C. Carter, S. Yi and P. Ratanchandani: Manycast: Exploring the space between anycast and multicast in ad hoc networks, Proc. MobiCom'03, San Diego, CA, USA (2003)

[4] D. Coppersmith, U. Feige and J. Shearer, Random walks on regular and irregular graphs, SIAM Journal on Discrete Mathematics 9/2 (1996) 301-8

[5] C. Gkantsidis, M. Mihail and A. Saberi, Hybrid search schemes for unstructured P2P networks, Proc. IEEE Infocom (2005)

[6] G. Hasslinger, ISP Platforms under a heavy peer-to-peer workload, Proc. Peer-to-Peer Systems and Applications, Eds.: R. Steinmetz and K. Wehrle, Springer LNCS 3485 (2005) 369-382

[7] G. Hasslinger and S. Kempken, Applying random walks in structured and self-organizing networks: Evaluation by transient analysis, Proc. Value tools workshop, Nantes (Oct. 2007) extended version to appear in PIK journal, special issue on self-organizing networks

[8] T. Kunz, On the inadequacy of MANET routing to efficiently use the wireless capacity, Proc. IEEE Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob 2005), Montreal, Canada (2005) 109-116

[9] T. Kunz and R. Alhalimi, Load-balanced routing in wireless networks: State information accuracy using OLSR, Proc. 3rd IEEE Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob 2007), New York, USA (Oct. 2007)

[10] L. Lima and J. Barros. Random Walks on Sensor Networks Proceedings of the 5th International Syposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt 2007), Limassol, Cyprus (April 2007)

[11] Q. Lv, P. Cao, E. Cohen, K. Li and S. Shenker: Search and replication in unstructured peer-to-peer networks, Proc. Internet. Conf. on Supercomputing, ACM (2002)

[12] L. Qian and T. Kunz, Mobility metrics for adaptive routing, Proc. 3rd IEEE ComSoc Conf. on Sensor and Ad Hoc Communications and Networks (SECON 2006), Vol. 3 (2006) 803-808

[13] M. Ripeanu and A. Iamnitchi: Mapping the Gnutella network, IEEE Internet Computing (2002) 50-57

[14] W. Terpstra, J. Kangasharju, C. Leng and A. Buchmann, BubbleStorm: Resilient, probabilistic and exhaustive P2P search, Proc. ACM SIGCOMM, Kyoto, Japan (2007)

[15] L. Tzevelekas and I Stavrakakis: Random walks with jumps in wireless sensor networks, Proc. of Med-Hoc-Net 2007, Corfu, Greece, June 2007.

[16] M. Zhong and K. Shen, Popularity-biased random walks for peer-to-peer search under the square root principle, Proc. 5th Internat. Workshop on P2P systems, IPTPS (2006)