When and How Much Should Random Walkers Proliferate for a Fast and Efficient Search?

Subrata Nandi, Ajit Pal
Department of Computer Science and Engineering
Indian Institute of Technology
Kharagpur, India
snandi,apal@cse.iitkgp.ernet.in

Lutz Brusch
Center for Information Services and High Performance Computing
TU Dresden
Dresden, Germany
lutz.brusch@tu-dresden.de

Niloy Ganguly
Department of Computer Science and Engineering
Indian Institute of Technology
Kharagpur, India
niloy@cse.iitkgp.ernet.in

ABSTRACT

The time-dependent coverage of a network by a varying number of random walkers is considered. Broadening the existing, coverage-focused perspective, we here explicitely address search efficiency, i.e. bandwidth consumption, which is essential for distributed systems. However, defining a single objective that combines speed and efficiency for estimation of the optimal number of walkers is a non-trivial issue, as speed increases but efficiency decreases with more numerous walkers. First, we define a combined objective to maximize coverage with least sacrifice in efficiency. Secondly, to fulfill the proposed objective, we design a novel search strategy inspired by the statistical mechanics of multiple, overlapping random walks. Our new algorithm is demonstrated on a four-dimensional Euclidean grid and it can be extended to other topologies with minor modifications. Finally, we formulate a suitable objective function and compare the optimal proliferation rates from theory and simulation.

1. INTRODUCTION

The increasing popularity of peer-to-peer systems threatens the quality of service of established search strategies. Besides scalability issues, existing algorithms are also vulnerable to ad-hoc changes of neighborhood relations. As a testbed for advanced search strategies we here consider multiple random walkers on a regular network and pose two challenges: (1) Can we design a local control protocol to obtain a fast and efficient random walk, i.e. simultaneous maximization of coverage rate and efficiency? (2) If that is possible, then when and how many walkers should self-replicate so that the above objective is fulfilled? Answers to the above questions are so far unknown even for the simple regular topology of the Euclidean grid. The strategy developed here will help designing fast, but still efficient enough algorithms for information dissemination and search in large scale distributed systems (e.g. unstructured peer-to-peer systems).

In a *simple N-walker* unrestricted random walk algorithm, the originating node sends out N independent walkers, each of them to a randomly chosen neighbor. In every intermediate step a node forwards each of the received random walkers to any one of its randomly chosen neighbor. The walker count does not increase with time and a walker terminates when its (time-to-live) TTL expires.

Random walk dynamics has been well studied [1]. Most attention comes from the physical sciences with problems such as diffusion-limited reaction, excitation trapping and multiple scavenger problems, where speed of the walk is the primary concern and thermodynamics drives the random process at no cost. However, in recent years, random walk has been widely used in the domain of distributed systems e.g. for unstructured search, information dissemination etc. In such applications, each hop of a random walker consumes some bandwidth which is a valuable resource and therefore, needs to be used judiciously. Hence the random walk phenomena need to be probed from quite a different perspective. The probing should help in designing novel algorithms to attain the following objective.

Objective: *Maximize the coverage subject to the condition that the efficiency of the walk is at its peak.*

Here *efficiency* at time t (E_t), is the ratio of the number of distinct nodes visited at time t to the number of visits at time t. Therefore, peak efficiency implies least number of redundant visits or overlaps which is the phenomena of multiple walkers visiting the same node. *Coverage* or spread [1] at time t (C_t), measures the mean number of distinct sites visited by the random walkers up to t time steps.

In a simple N-random walk, especially in the initial stage, a large value of N results into high inefficiency (see for example [4] and [5]). Therefore, in order to avoid this inefficiency, the strategy may be changed, by starting a walk with few initial walkers and introduce new walkers gradually with time as the trajectories of the previous walkers separate. Hence, walkers need to proliferate (self-replicate) with time in a regulated manner so that the walk may spread fast without much sacrifice of efficiency.

In the subsequent sections, it will be shown that a proliferating random walk strategy essentially fulfils the objective. However, we will see that estimating the optimal proliferation rate is a challenging problem because it is solely controlled by the walker dynamics on the underlying graph. Although some existing works [7] design proliferation regulation schemes based on application driven conditions, to the best of our knowledge, strategies based on walker dynamics have not been investigated. The above observations provide motivation for this work.

In this paper, we formally design the suitable proliferation strategy and formulate a maximizing objective function which is inspired by the statistical mechanics of multiple walkers. We estimate the optimal proliferation rate based on the objective function. In this work, we restrict our analysis to the infinite-size Euclidean grid. The proposed idea can be extended to other topologies with some minor modifications.

The rest of the paper is organized as follows: Section II designs the proposed proliferation strategy from the understanding of walker dynamics in existing literature and formulates the objective function. In Section III, we estimate the optimal proliferation rate

[1]We use the term coverage throughout the paper.

from simulation and also verify the effectiveness of the proposed strategy. Finally, we conclude in section IV.

2. PROLIFERATION STRATEGY

The following three subsections will (1) review the statistical mechanics of multiple random walkers, (2) design the proliferation strategy and (3) formulate the objective function required to evaluate the optimal proliferation rate.

2.1 Multiple Random Walkers

The average coverage of an infinite graph by N independent random walkers ($N \gg 1$) all starting from the same source node has been solved analytically [4]. The results are summarized in Table I. Three distinct time regimes are observed. Crossover time from the regime I to II and the regime II to III are denoted as t^{I-II} and t^{II-III}, respectively. The generalized expression for dimension $d > 3$ in the regimes II and III are conjectured from the results in $d = 1, 2$ and 3.

Analyzing the results shown in Table I for $N \gg 1$, the following are observed:

- **Observation 1:** The system remains in regime I for a very short duration $t^{I-II} = \ln N$. Let the *coverage rate* at time t be denoted as C'_t, which is $\frac{dC}{dt}|_t$ i.e., the number of distinct nodes visited at time t. In regime I, $C'_t = d \times t^{d-1}$. Hence, regime I is similar to flooding.

  Explanation: It is due to the fact that very few nodes close to the source contain walkers (Fig. 1(I)). As $N \gg 1$, each node has too many walkers compared to the number of unvisited neighbors. As a result, all the neighbors of the already visited nodes are reached at the next step. Spread rate observed is similar to flooding.

- **Observation 2:** As time passes, the trajectories of the walkers separate and the dynamics enters regime II. Then coverage $C_t \sim [t \times \ln(N \times t^{1-\frac{d}{2}})]^{\frac{d}{2}}$ for $d \geq 3$. In this regime, $\ln N < t < N^{\frac{2}{d-2}}$, therefore considering the number of walkers $N \sim 10^2, 10^3$, coverage can be further approximated to $C_t \approx t^{\frac{d}{2}}$. It implies coverage rate $C'_t \approx \frac{d}{2} \times t^{\frac{d-2}{2}}$, which shows that it increases but the rate of increase is less than that of regime I. The duration of regime II decreases with increase in grid dimension d.

  Explanation: The rate increase of $C't$ in regime II is less compared to regime I because, the walkers gradually move away from each other in regime II (Fig 1(II)), which results in a decrease in the number of walkers at each node.

- **Observation 3:** Finally, from time $t = t^{II-III}$ onwards the system remains in regime III. Analyzing results from Table I we obtain $C'_t = N$ in regime III, which means the walkers visit new nodes at a constant rate.

  Explanation: In this regime the walkers move *sufficiently far apart* from each other (Fig 1(III)), such that on average each node contains a single walker with almost all its neighbors unvisited. As a result from time $t = t^{II-III}$ onwards, each walker walks in a non-overlapping exploration space with respect to other walkers.

The following can be inferred from the above observations:

Inferences:

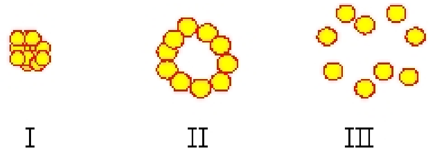

Figure 1: The increasing spatial separation and decreasing mutual overlap of random walkers is depicted as the dynamics passes three regimes while time progresses.

1. The coverage rate ($C't$) is a function of the number of walkers N i.e. $C'_t = f_1(N)$. The coverage rate can be increased by using a larger value of walker count N but the impact of walkers is significant only in regime III.

2. For $N \gg 1$, the efficiency (i.e. $E_t = \frac{C'_t}{N}$) increases at a faster rate with time in regime I than in regime II and reaches the peak value 1 at $t = t^{II-III}$ i.e. the boundary of regime II and regime III and remains steady throughout regime III.

3. The crossover time from regime II to III is a function of N i.e. $t^{II-III} = f_2(N)$. It is noted that peak efficiency is reached early if the number of walkers N used in the walk is small.

4. From the existing theory the functional form of coverage rate in regime III and crossover time from regime II to III can be expressed as $C'_t = k_1 \times N$, and $t^{II-III} = k_2 \times N^{\frac{2}{d-2}}$ respectively, where both k_1, k_2 are constants and $k_1 \leq 1$.

2.2 Design of the Proposed Strategy

The objective is to maximize speed without sacrificing efficiency at every time instant. Being inspired from our understanding of walker dynamics explained above, the proposed strategy is stated formally in this subsection, followed by the intuitive reasoning and physical implication.

The proliferation strategy: *Start with a small initial set of walkers N_{init} and proliferate in every step at a rate P, so that the system always remains at the boundary of regimes II and III.*

Intuitive basis: Let's see why proliferation is required to fulfil the stated objective. Consider that at time $t = 1$, N walkers start from a single source node and let $t^{II-III}|_N$ be denoted as the crossover time from regime II to III for N walkers. Then walker count must not be increased during regime II and III i.e. at time $t < t^{II-III}|_N$ as it will have little impact on coverage rate (refer to inference 1). It can also be noted, that with an objective to maximize coverage without sacrifice of efficiency, the walker count must be increased *at the earliest* i.e. at time $t = t^{II-III}|_N$ (refer to inference 1 and 2). Therefore, given initial N walkers, proliferation becomes necessary at $t = t^{II-III}|_N$.

Let us now consider ΔN new walkers are added through proliferation at time $t = t^{II-III}|_N$ and the system now contains N' walkers where $N' = N + \Delta N$. To fulfil the objective, ΔN should be such that the condition $(t+1) = t^{II-III}|_{N'}$ is satisfied. The condition essentially means that, the resulting system of N' walkers after proliferation should also remain at the boundary of regime II and III. Thus, we observe that proliferation is essential and it should be regulated in such a way that the system always remains at the boundary of regime II and III.

To reduce the initial inefficiency during regimes I and II, the walk needs to start initially with a small number of walkers (refer to inference 3). Hence the novel strategy proposed above fulfils the objective at each time step.

Table 1: Summary of the expressions for the average number of distinct sites visited (C_t) for different density regimes as well as the crossover times separating the regimes (Courtesy [4]). Averaging has to be performed over repeated search runs, especially in cases of small numbers of walkers N.

Dimension	Coverage in Regime I	t^{I-II}	Coverage in Regime II	t^{II-III}	Coverage in Regime III
1	$C_t \sim t$	$\ln N$	$C_t \sim t \times \ln N$	∞	
2	$C_t \sim t^2$	$\ln N$	$C_t \sim t \times \ln(\frac{N}{\ln t})$	e^N	$C_t \sim \frac{N \times t}{\ln t}$
3	$C_t \sim t^3$	$\ln N$	$C_t \sim [t \times \ln(\frac{N}{\sqrt{t}})]^{\frac{3}{2}}$	N^2	$C_t \sim N \times t$
d	$C_t \sim t^d$	$\ln N$	$C_t \sim [t \times \ln(N \times t^{1-\frac{d}{2}})]^{\frac{d}{2}}$	$N^{\frac{2}{d-2}}$	$C_t \sim N \times t$

Physical implication: At the boundary of regime II and III the walkers just start to be *sufficiently far apart*, implying that each walker has some non-overlapping exploration space with respect to other walkers such that there is almost no redundant visit. Therefore, the proposed strategy ensures that, a new walker once introduced gets its own non-overlapping exploration space, thereby increasing coverage rate in the most efficient way.

The key to the successful implementation of the strategy lies in the precise estimation of the (1)proliferation rate P and (2)defining an objective function to measure the performance of the walk.

2.3 Objective Function

This subsection formulates a maximizing objective function which computes the total overall gain, both in terms of speed and efficiency of a random walk, from the understanding of the walker dynamics in section 2. It is used to quantify the goodness of any random walk strategy in terms of the framed objective. Let's evaluate the gain in efficiency and speed achieved by n walkers at time t.

Efficiency achieved at time t, by n random walkers, can be expressed as $\frac{C'_t|_n}{n}$, i.e. the ratio of the coverage rate at time t to the number of visits at time t. The steady maximum achievable efficiency by n walkers is $\frac{C'steady|_n}{n}$, where $C'steady|_n$ is the maximum achievable coverage rate. Therefore, gain in efficiency ($\mathcal{GE}_t|_n$) is:

$$\mathcal{GE}_t|_n = \frac{\frac{C'_t|_n}{n}}{\frac{C'steady|_n}{n}} = \frac{C'_t|_n}{C'steady|_n} \quad (1)$$

Coverage rate of the walk is known to be proportional to the walker count during regime III (refer to inference 1), therefore faster coverage could have been achieved if n walkers are introduced early in the system. Let the time, when $C'steady|_n$ is achieved, be denoted by $Tsteady|_n$. Given the objective, there is no advantage of introducing n ahead of time $Tsteady|_n$ as this does not result in enhancement of speed (refer to inference 3). Therefore, gain in coverage speed ($\mathcal{GS}_t|_n$) is:

$$\mathcal{GS}_t|_n = \begin{cases} \frac{Tsteady|_n}{t} & : \ if \quad t > Tsteady|_n \\ 1 & : \ otherwise \end{cases} \quad (2)$$

The overall gain ($\mathcal{OG}_t|_n$) of the walk at time t can be expressed as:

$$\mathcal{OG}_t|_n = \mathcal{GE}_t|_n \times \mathcal{GS}_t|_n \quad (3)$$

Values of $\mathcal{GE}_t|_n$, $\mathcal{GS}_t|_n$ and $\mathcal{OG}_t|_n$ lie in the range [0,1]. At time t, the n walkers used, may be obtained either from a non-proliferating n-walker strategy or from a proliferating walk strategy with rate P. For a non proliferating walk, $\mathcal{OG}_t|_n$ gradually increases with time, reaches the peak value and then falls gradually. Therefore, there exists an optimal proliferation rate $P^{opt}(t)$ which maximizes $\mathcal{OG}_t|_n$ at each time step throughout the walk.

The total overall gain of the walk until time T, is defined as the sum of overall gain estimated at each time step t for $t = 1$ to T. Let $\mathcal{TOG}_T|_{N,0}$ be the normalized total overall gain until time T for N non proliferating walkers and $\mathcal{TOG}_T|_{1,P}$ be for a proliferating walk with rate P, starting with a single walker. In general, $\mathcal{TOG}_T$ is defined as follows:

$$\mathcal{TOG}_T = \frac{\sum_{t=0}^{T} \mathcal{OG}_t|_n}{T} \quad (4)$$

Therefore, $P^{opt}(t)$ maximizes $\mathcal{TOG}_T|_{1,P}$ for sufficiently large T with
$\mathcal{TOG}_T|_{1,P} =$

$$\sum_{t=0}^{T} \frac{C'_t|_n}{C'steady|_n} \times \frac{Tsteady|_n}{t} \quad . \quad (5)$$

Now, $Tsteady|_n$ and $C'steady|_n$ can be obtained from Table I as $t^{II-III}|n$ and $C'|n$ at regime III, respectively but $C'_t|_n$ i.e. coverage rate at any time t due to proliferating walkers is not known from existing literature. In the next section we employ numerical simulations to determine the optimal dependency $P^{opt}(t)$ after implementing and testing the predicted proliferation strategy which works at the boundary of regimes II and III.

3. NUMERICAL SIMULATIONS

This section is organized into two parts. The first section estimates the values of $Tsteady$ and $C'steady$ from simulation. In the second section, we estimate the optimal proliferation rate from simulation and compare it to the above theory.

3.1 Verifying maximum coverage rate and crossover times through simulation

We simulate a simple random walk with N walkers with the following objectives:

1. To check the presence of three distinct regimes and to check the functional form predicted by theory.

2. To obtain values of k_1 and k_2 to predict $C'steady|_N$ and $Tsteady|_N$ precisely and not just their qualitative time dependence.

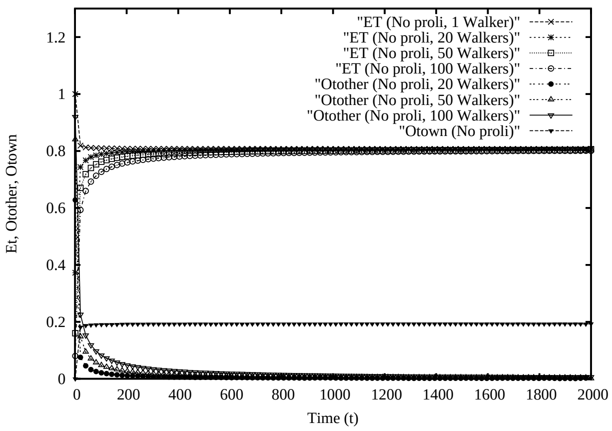

Figure 2: Plot of efficiency (E_t), overlap probability with other's trail $(Oother_t)$ and overlap probability with its own trail $(Oown_t)$ vs time (t) for simple random walk using $N = 1, 20, 50, 100$ walkers, $d = 4$.

3. To check the probability to overlap. It can be used as an important metric to verify that for $N \gg 1$ efficiency E_t increases indeed due to decrease in overlap probability with time.

To observe the regimes distinctly, we plot efficiency $E_t = \frac{C'_t}{N}$ which lies in the range [0,1]. In addition to the metrics efficiency (E_t) and coverage rate (C'_t), we also measure the overlap probability metrics defined as follows:

- $Oother_t$ - Mutual overlap probability i.e. probability that a walker overlaps with other walker's trail, i.e. it visits a node at time step t, which has already been visited by some other walkers (excluding itself) in some previous time steps, averaged over all walkers.

- $Oown_t$ - Own overlap probability, i.e. probability that a walker overlaps with its own trail at time step t.

The set of N walkers, all starting from the same source node gradually dilutes. Therefore, $Oother_t$ decreases with time. From the above definitions it is expected that $E_t + Oother_t + Oown_t = 1$.

Simulation parameters: 4 - dimensional Euclidean torous grid topology is used with Von Neuman neighborhood, i.e. each node has two neighbors along each dimension. Size of the grid $236421376 = 124^4$. Simulation is performed upto 2000 time units. Number of runs considered for each experiment is 3×10^5. A sufficiently large finite size graph is used along with a relatively small simulation time, so that during a simulation the finite size effect does not take place, i.e. walkers never bounce back from dimension boundaries. Torous grid is considered so that each node including the boundary nodes have the same degree. Each metric is measured by taking the average value of 3×10^5 runs. Here, we restrict our study only to a grid of dimension 4, therefore $Tsteady|_N$ is expected to be $k_2 * N$ (refer to Table I).

Results: To observe the regimes we plot E_t, $Oother_t$ and $Oown_t$ versus time t in Fig. 2. E_t and $Oother_t$ have been plotted for different values of walkers count, N =1, 20, 50 and 100. $Oown_t$ is plotted only for $N = 1$ as it remains the same for any N.

It is found that $E_t|_{N=1} \approx 0.8$ remains almost steady at $t \gg 1$. It is because $Oother_t|_{N=1} = 0$ and $Oown_t|_{N=1} \approx 0.2$, as for $d = 4$, each node has 8 neighbors, and probability to jump to the last visited node is $\frac{1}{8}$, therefore the most trivial lower bound of $Oown_t$ is $\frac{1}{8} \approx 0.125$.

$E_t|_{N=20,50,100}$ is found to increase very fast for a short while, which can be interpreted as regime I. It is also observed that mutual overlap $Oother_t|_{N=20,50,100}$ decreases drastically during this regime. As time progresses, $E_t|_{N=20,50,100}$ still increases, but at a slower rate, which refers to regime II. Correspondingly in this regime, $Oother_t|_{N=20,50,100}$ decreases, but slowly compared to that in regime I.

Finally, $E_t|_{N=20,50,100}$ reaches a plateau and becomes steady at a value $\approx N \times 0.8$ which signifies regime III. It is further observed that, $Oother_t|_{N=20,50,100}$ also takes a steady value ≈ 0 during this regime. Therefore, it is noted that in asymptotic time the maximum efficiency achieved by N walkers is N times the maximum efficiency of a single walker (regime III). This confirms the mutually non-overlapping subspaces predicted by theory. It have been verified that for all N, at any time t, $E_t|_N + Oother_t|_N + Oown_t|_N$ = 1 holds true.

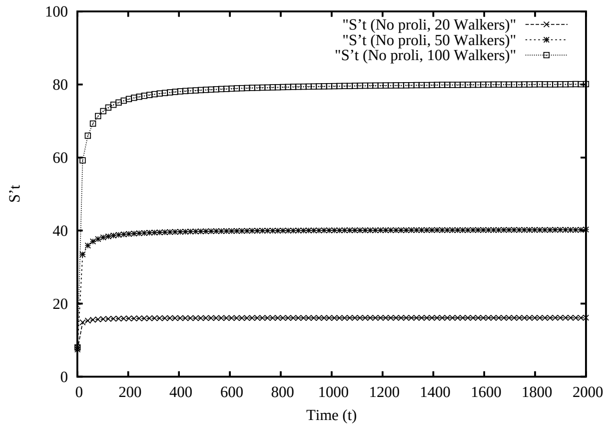

Figure 3: Plot of coverage rate (C'_t) vs time (t) for simple random walk using $N = 20, 50, 100$ walkers, $d = 4$.

To get the estimate of $C'steady|_N$ and $Tsteady|_N$, we plot C'_t versus time t for N =20, 50 and 100 in Fig. 3. A closer look into the plot shows that at $t = N$ coverage rate $C'_t \approx N \times 0.73$, close to $N \times 0.8$ as expected from theory. It is also observed that for a given N at $t = N$, $(N \times 4)$, and $(N \times 15)$, coverage rate is $C'_t \approx (N \times 0.73)$, $C'_t \approx (N \times 0.78)$ and N, respectively.

To summarize, the simulation results show that for a 4-d grid, the coverage rate becomes steady at a time $Tsteady|_N \approx N \times 15$ with $C'steady|_N = N \times 0.8$. It is also noted that at time $t = N \times 15$ and $N \times 4$, efficiency is 8.7 and 2.5 percentage less only than the peak value, respectively. The next subsection estimates the optimal proliferation rate through simulation using the above mentioned results.

3.2 Estimation of the optimal proliferation rate

Let $\hat{P}_S$ be denoted as the optimal proliferation rate obtained from simulation. We perform two experiments with the following objectives:

1. To obtain $\hat{P}_S$.

2. To see that regulated proliferation using optimal proliferation rate $\hat{P}_S$ is an essential strategy to fulfil the objective to maximize coverage with least sacrifice in efficiency, for a given

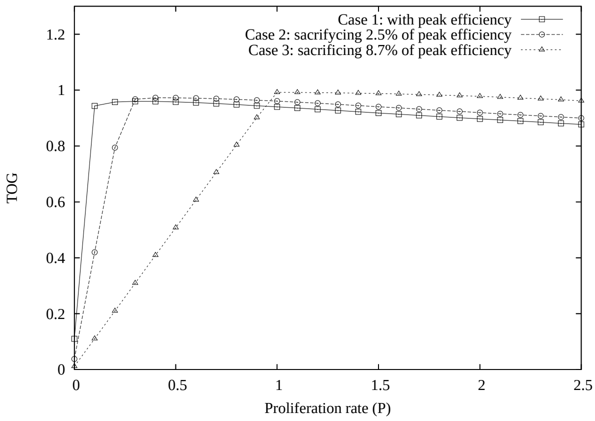

Figure 4: Plot of Normalized total overall gain (TOG) vs proliferation rate (P) for three different cases of efficiency.

time which cannot be achieved otherwise using a simple random walk using equivalent number of hops.

Experiment 1:- Estimation of $\hat{P}_S$

In the experiment, the total overall gain is computed for different values of proliferation rate P. The simulation considers $N_{init} = 1$ and varies P from 0 to 2.5 with step size 0.1. The total simulation time is assumed to be 500 units. $\hat{P}_S$ is estimated as the value of P that maximizes total overall gain $\mathcal{TOG}_{500}|_{1,P}$. The rest of the simulation parameters are same as used in subsection 3.1. $\hat{P}_S$ is of interest with respect to peak efficiency constraint but also for stepwise relaxed constraint. Through simulation we estimate $\hat{P}_S$ considering three different cases:

- *Case 1* - Estimate $\hat{P}$ without any sacrifice of the peak efficiency, i.e. $Tsteady$ and $C'steady$ is substituted by $N \times 15$ and $N \times 0.8$, respectively to calculate overall gain $\mathcal{OG}_t|_{1,P}$ at each time step t (refer equation 6). It is in accordance to the rule framed by the optimal proliferation strategy.

- *Case 2* - Estimate $\hat{P}$ by sacrificing 2.5 percent of the peak efficiency i.e. $Tsteady$ and $C'steady$ is substituted by $N \times 4$ and $N \times 0.78$ respectively.

- *Case 3* - Estimate $\hat{P}$ by sacrificing 8.7 percent of the peak efficiency, i.e. $Tsteady$ and $C'steady$ is substituted by N and $N \times 0.73$, respectively.

Results - The plot of total overall gain $\mathcal{TOG}_{500}|_{1,P}$ versus proliferation rate P is shown in Fig. 4 corresponding to the above three cases. It is found that for case 1, $\hat{P}_S = 0.3$ with $\mathcal{TOG}_{500}|_{1,0.3} = 479.8903$. For case 2, $\hat{P}_S = 0.4$, i.e. the system is allowed to proliferate in a marginally higher rate at the cost of 2.5 percent efficiency. For case 3, $\hat{P}_S = 1.0$, i.e. by using the same value of $Tsteady|_N$ as analytically predicted and at the cost of 7.3 percent efficiency, the obtained proliferation rate approaches 1. Finally, it is observed that the coverage obtained up to simulation time 500 time steps using $\hat{P}_S$ = 0.3, 0.4 and 1, i.e. optimal rates by sacrificing 0, 2.5 and 8.7 percent efficiency, respectively are 29987, 39555 and 95078 nodes. There is 217 and 8.7 percent increase obtained by sacrificing 2.5 and 8.7 percent efficiency, respectively.

To summarize, the precise value of $Tsteady|_N$ and $C'steady|_N$ is essential to obtain the correct estimate of $\hat{P}$. If a very small percentage of efficiency is sacrificed, then the system is allowed to proliferate at a marginally higher optimal rate, which results in a huge increase in coverage. Though ideally ensuring peak efficiency is a desired condition, but it is not an application specific constraint. Therefore, a small amount of overlap may be allowed in practice to get the benefit of increase in coverage.

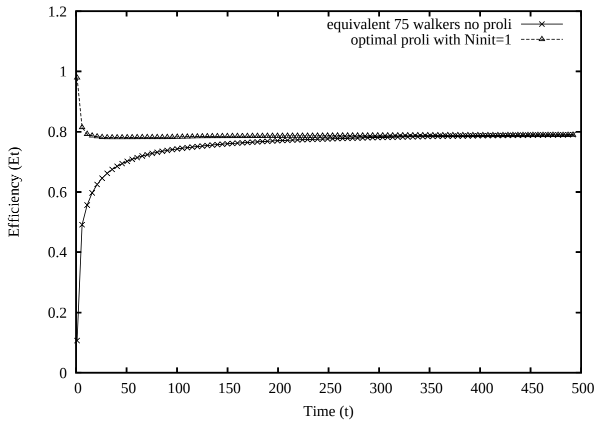

Figure 5: Plot of Efficiency (E_t) vs time (t) for optimal proliferation with rate 0.3 and equivalent simple random walk.

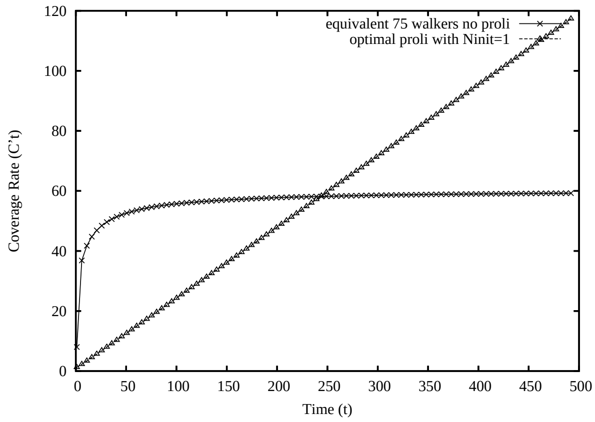

Figure 6: Plot of (C'_t) vs time (t) for optimal proliferation with rate 0.3 and equivalent simple random walk.

Experiment 2:- Testing the effectiveness of the regulated proliferation strategy

This experiment compares the performance of the proposed optimal proliferation strategy using $\hat{P}_S = 0.3$ (say $\mathcal{PR}$**)** with a simple random walk utilizing equal amount of bandwidth resource (i.e. total number hops up to specified simulation time T). The simple random walk uses N walkers. In this experiment, we choose $N = 75$ and simulation time $T = 492$ judiciously so that the total hops used matches. Let the proliferating random walk and simple random walk be denoted as $\mathcal{PR}$ and $\mathcal{SR}$, respectively. Performance is compared in terms of the E_t, C'_t and $\mathcal{TOG}|_{500}$ for both the strategies.

The plot of E_t for both the strategies is given in Fig. 5. It shows that the efficiency of $\mathcal{PR}$ remains close to the efficiency of a single walker (refer Figure 2) throughout the proliferating walk, in spite of the fact that the walker count increases with time, because proliferation is regulated properly. The efficiency of $\mathcal{SR}$ is less than

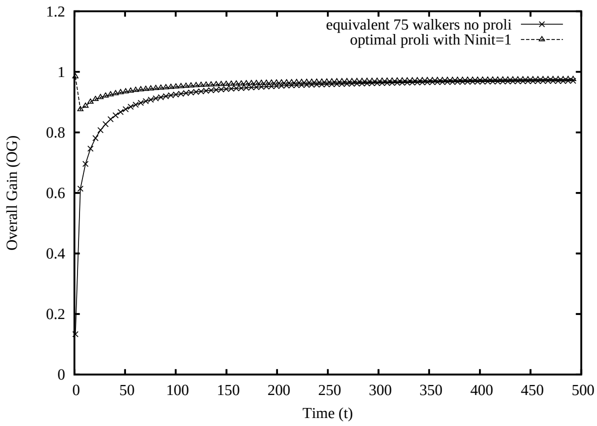

Figure 7: Plot of $(\mathcal{OG})$ vs time (t) for optimal proliferation with rate 0.3 and equivalent simple random walk.

$\mathcal{PR}$ in the initial stage of the walk, as it passes through regime II and III. Plot of C_t' versus t given in Figure 6 shows that the coverage rate of $\mathcal{PR}$ is higher than $\mathcal{SR}$ almost from the half way mark onwards, because the number of walkers used by $\mathcal{PR}$ exceeds $\mathcal{SR}$. Fig. 7 shows the plot of $\mathcal{O}G$ versus t for both the strategies, which is found to be greater for $\mathcal{PR}$ in the initial stage as simple walk is penalized due to initial inefficiency captured by $\mathcal{G}E_t$. The plot of $\mathcal{G}E_t$ closely matches with $\mathcal{O}G_t$ as in this experiment $\mathcal{G}S_t$ is always 1 for both the strategies.

It may be noted that $\mathcal{TOG}|_{\mathcal{PR}} = 472.16$, $\mathcal{TOG}|_{\mathcal{SR}} = 459.321$ shows $\mathcal{PR}$ performs better than $\mathcal{SR}$ considering efficiency and coverage combined. Further, the total coverage $C_{\mathcal{PR}} = 29035$ and $C_{\mathcal{SR}} = 27855$, showing that given a time 500 units, the regulated proliferation strategy covers 4.2 percent more nodes compared to simple random walk. The results show that given the time and fixed amount of resource (bandwidth), properly regulated proliferating random walk strategy always covers more nodes than a simple random walk.

4. CONCLUSIONS

The optimal proliferation rate estimated in this paper answer the fundamental questions raised in the introduction, which are highly relevant in the domain of distributed systems. The proposed regulated proliferation strategy is solely topology dependent and developed from insights on walker dynamics, hence provides an application independent generic approach towards designing fast and efficient random walk strategies. The optimal rate has been estimated both based on theory and simulation and the reasons for deviation have also been explained. Results show that regulated proliferation is an essential strategy to ensure optimal use of resources.

In this work, we considered infinite 4-d Euclidean graph. Though it is a simple topology, the proposed approach can be well applied to other dimensions of grid as well as other network topologies (e.g. random graph, small world, power law), on which work is in progress.

5. REFERENCES

[1] L. Lovasz, “Random Walks on Graphs: A Survey," *Combinatorics*, vol. 2, pp. 1-46, Keszthely Hungary, 1993.

[2] L. K. Gallos, “Random walk and trapping processes on scale-free networks," *Physical Review E*, 70, 046116, 2004.

[3] E. Almaas, R.V. Kulkarni, D. Stroud, “Scaling properties of random walks on small-world networks," *Physical Review E*, 68, 056105, 2003.

[4] H. Larralde, P. Trunfio, “Number of distinct sites visited by N random walkers," *Physical Review A*, 45, 7128, 1992.

[5] S.B. Yuste, L. Acedo, “Number of distinct sites visited by N random walkers on an Euclidean lattice," *Physical Review E*, 61, 2340, 2000.

[6] S. A. Pandit, R.E. Amritkar, “Random spread on the family of small-world networks," *Physical Review E*, 63, 411041, 2001.

[7] N. Ganguly, A. Deutsch, “Developing Efficient Search Algorithms for P2P Networks Using Proliferation and Mutation," *Proceedings : International Conference on Artificial Immune Systems*, Catania, Italy, September 2004.

[8] N. Alon et al., “Many Random Walks Are Faster Than One," *arXiv:0705.0467v1 [math.PR]*, 3 May 2007.