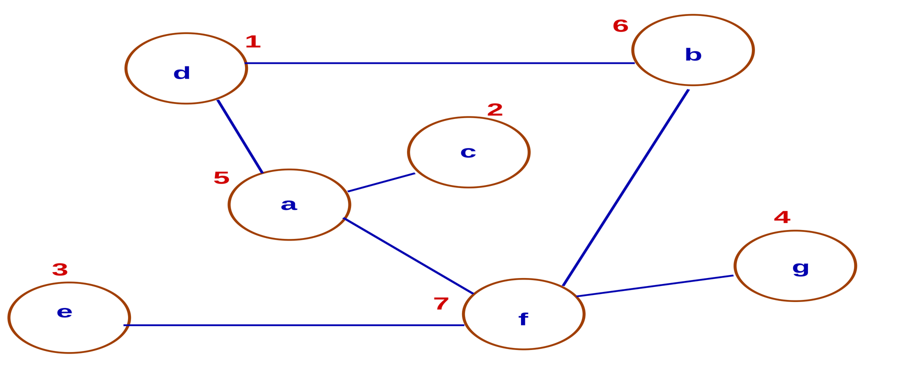
A Biology-Inspired Multi-Agent System For Efficient Search Of Unstructured Networks

Niloy Ganguly, Andreas Deutsch
Center for High Performance Computing, Dresden University of Technology

Peer-to-peer Networks

Salient Features:

- Totally decentralized networks unlike traditional client server architecture.
- Direct connection between peers.
- Unlike traditional Internet where peers are mere recipients of information, in these systems, peers are also providers of information.
- Popular systems - Napster, Kazaa, Freenet, Gnutella etc.



Picture of an unstructured network. The network comprises of nodes $< a, b \cdots >$, connections between nodes and data stored in the node $<1, 2 \cdots >$. Note: there is no relation between the nodes and data stored in the nodes.

Agent Based Search in Peer-to-peer Networks

Methodology

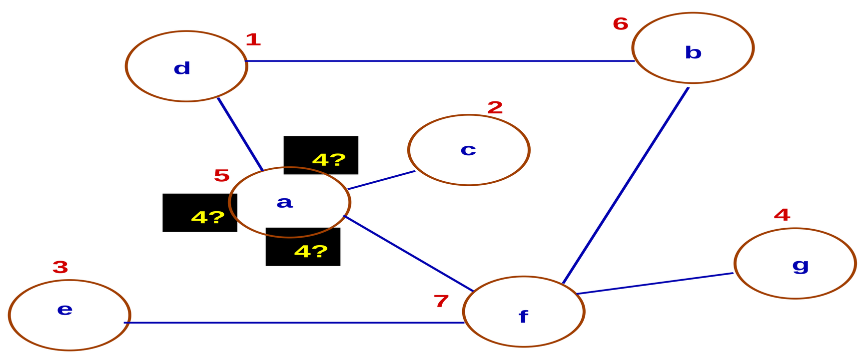
- Nodes initiate a search by sending out message packets (agents) in the network. The agents carry with them the search query.
- The agents travel through the network following a random path until they find the correct node.

Problem

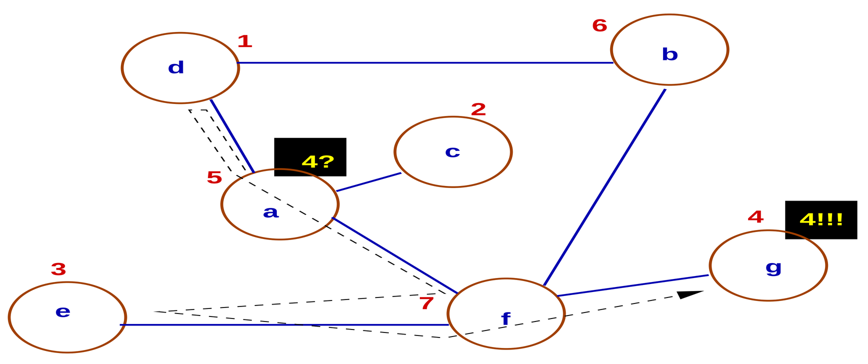
- The agents, undergoing random walk, take a long time to reach the correct node. Consequently, searching for content in peer-to-peer networks is very slow.

Solution

- Regulate agent behavior according to an immune-inspired mechanism of *opportunistic proliferation.*



When search is initiated by a node (here a), it generates many message agents with information of what is to be searched (here 4).



The movement of a single agent. The agent's movement from node a to the final destination g is plotted with a dotted line.

Agent Behavior

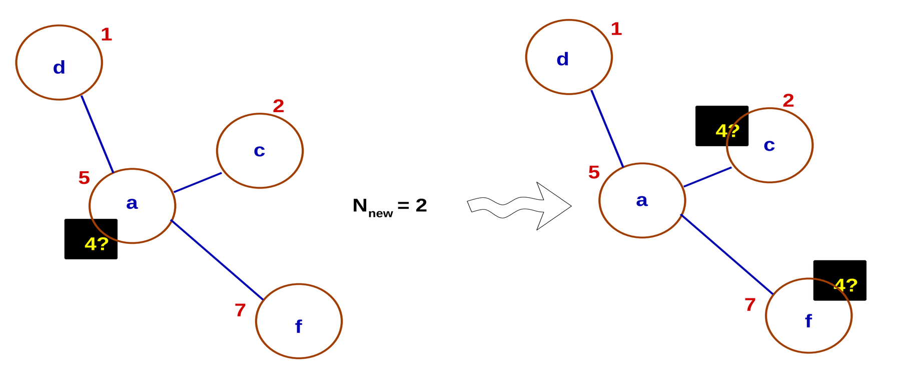
Two types of agent behavior are proposed in this work.

- Opportunistic proliferation.
- Restricted opportunistic proliferation.

The agent on visiting a node (say a) expresses the above mentioned behaviors.

Opportunistic proliferation

Algorithm 1 Opportunistic proliferation OP(a)
Input : Message agent (M) visiting the node a
Produce N_{new} [a] agents (M)
Spread the N_{new} packets to N_{new} randomly selected neighbors of a

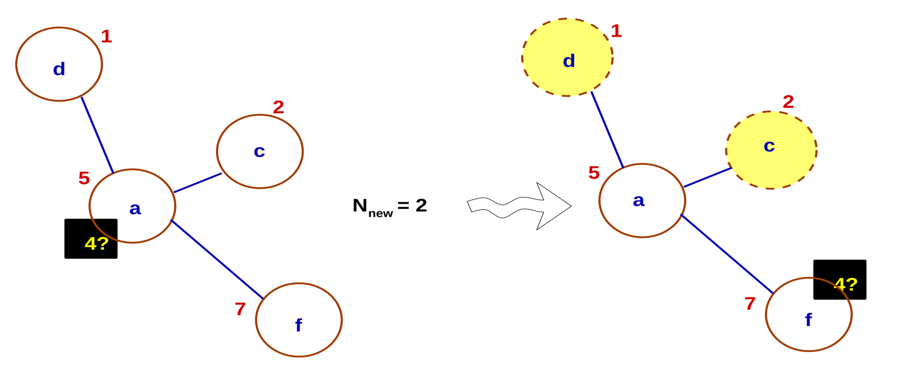


The picture shows the opportunistic proliferation behavior of the message agent visiting the node a. The agent multiplies into 2 ($= N_{new}$) packets and spread to the neighboring nodes c and f.

Restricted opportunistic proliferation

Algorithm 2 Restricted opportunistic proliferation ROP(a)
Input : Message agent (M) visiting the node a
Produce N_{new} agents (M)
Spread the N_{new} packets to N_{new} randomly selected neighbors of a
if N_{new} neighboring nodes are 'free' [b]
else
send to only the 'free' nodes. Destroy the others.
If no 'free' node
Forward one to a randomly selected neighbor of a

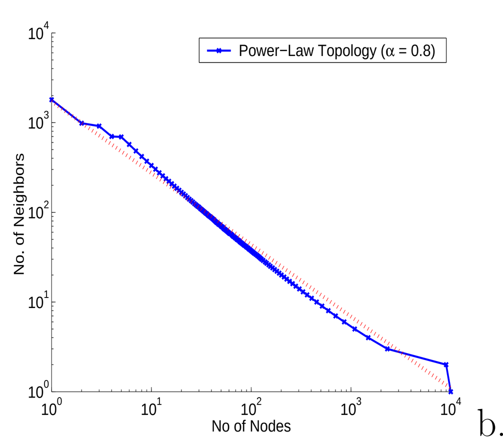
[a] N_{new} is generated through a *proliferation controlling function* which is discussed later
[b] 'free' nodes imply those nodes which have not been previously visited by any message agents



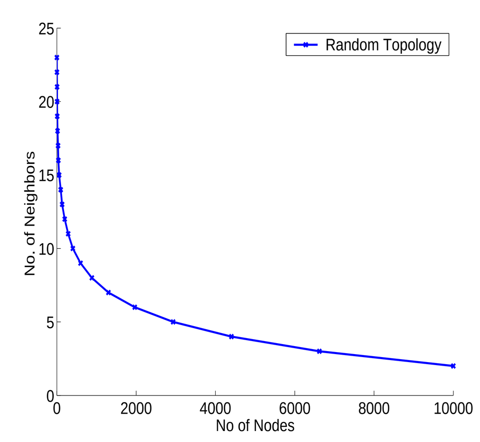
The picture shows the restricted opportunistic proliferation behavior of the message agent visiting the node a. The agent wants to multiply into 2 ($= N_{new}$) packets and spread to the neighboring node of a. However, 2 out of the 3 neighbors of a have already been visited by some other agents. So the agent stops multiplying and moves to the yet not visited node f.

Modeling Abstraction

- Peers: 10.000, each peer carries (a) information profile (P_I) (b) search profile (P_S) [(c) agents (M)]
- Overlay network: random network and power-law network
- Each node hosts a peer
- Profile & agent are represented by a 10-bit binary string
- Distribution of profiles according to Zipf's law
- Profile/message affinity: denoted by sim(M,Px) = [10 - HD(M,Px)], HD: Hamming distance



b. Power-law network, 10000 nodes; $\mu \approx 4$ and $\alpha \approx 0$.



a. Random network, 10000 nodes; $\mu = 4$

Distribution of node degrees in the two networks. μ implies node in-degree and α is the power-law exponent.

This work was partially supported by the Future & Emerging Technologies unit of the European Commission through Project BISON - *Biology-Inspired techniques for Self-Organization in dynamic Networks* (IST-2001-38923).

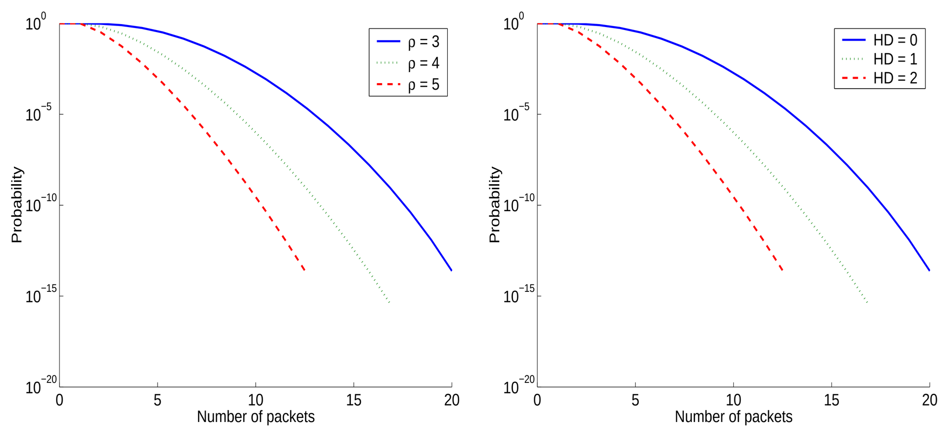
http://rcswww.urz.tu-dresden.de/~ganguly

Proliferation Controlling Function

- The proliferation of message packets at any node A is heavily dependent on the similarity between the message packet (M) and the information profile (P_I) of A.
- Given $p = e^{-HD} \times \frac{\rho}{n}$, HD is the Hamming distance(M,P_I), n represents the number of neighbors the particular node has, ρ represents the proliferation constant; we define $P(\eta)$ - the probability of producing at least η agents during proliferation by the following equation.

$$P(\eta) = \sum_{i=\eta}^{n} \binom{n-1}{i-1} \cdot p^{i-1} \cdot (1-p)^{n-i}$$

- The significance of the above equation is elaborated through the adjoining figures.



ρ varies; $HD = 0$, $n = 20$ HD varies; $\rho = 3$, $n = 20$

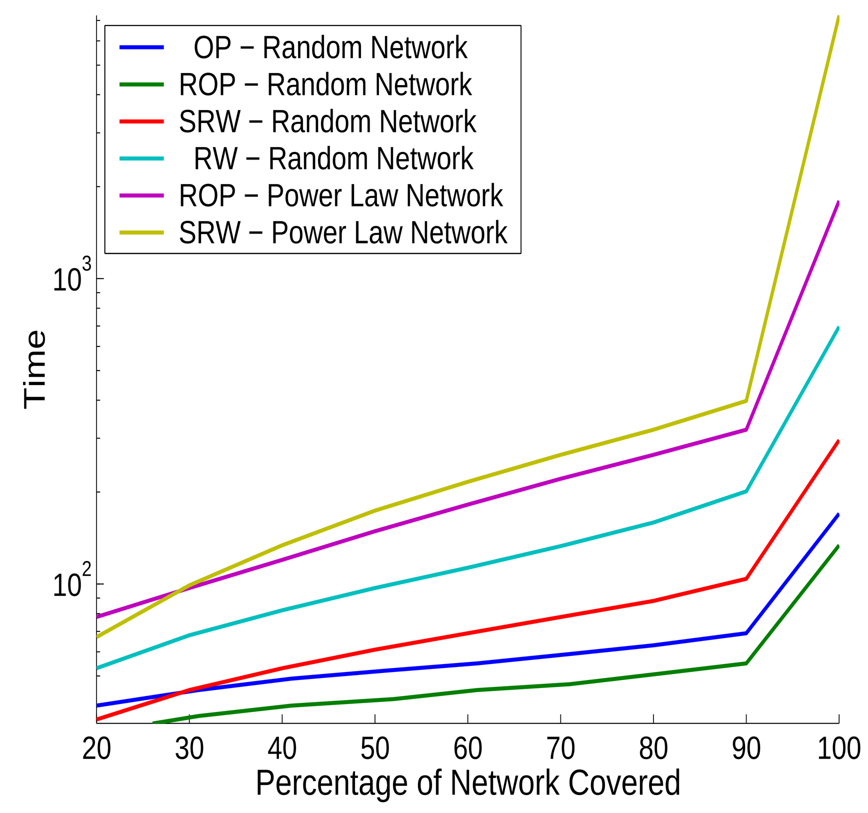
Probability of proliferation of at least η messages. The figures illustrate some commonality. (i) At least one packet necessarily proliferates, that is, the agent itself never get destroyed; (ii) The probability of proliferation of larger numbers of agents exponentially decreases.

Simulation

- Two types of experiments - (a). Experiment Coverage; (b). Experiment Time-step.
- Comparison between different agent behaviors : random walk (RW), self-avoiding random walk (SRW), opportunistic proliferation (OP), restricted opportunistic proliferation (ROP).
- Fairness in 'power' between proliferation and random walk is ensured by keeping the total number of agents roughly the same in all the cases. As is understood, the number of packets increase in proliferation while it remains constant in random walks. So, random walk experiments typically start with higher number of agents.

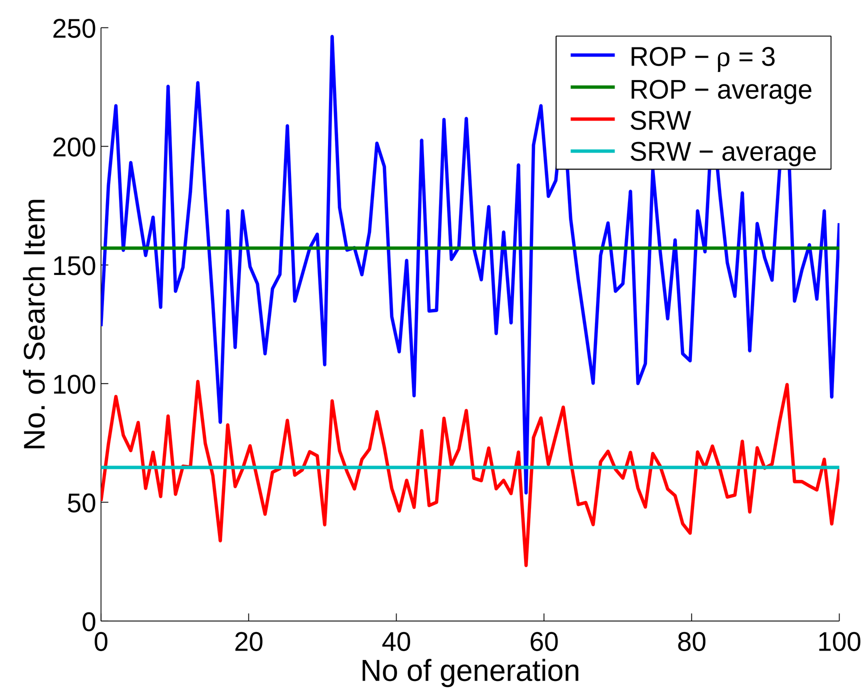
Experiment Coverage

- In this experiment, upon initiation of a search, the search operation is performed till the message agents cover the entire network. The experiment is repeated 500 times on randomly selected initial nodes.
- During the experiment, we collect different statistic at every 10% of coverage of the network that is, we collect statistic at [20%, 30% $\cdots$ 90%, 100%] of coverage of the network.
- The graph plots the % of network covered in the x-axis, while the time taken to cover corresponding % of network is plotted on the y-axis. The proliferation constant - ρ is 3 in all the cases.
- The graphs shows the performance of SRW, and ROP for both random and power-law networks and OP and RW for random network.
- Observations (a) The performance of ROP is best followed by OP, SRW, and RW. (b). Performance is better in random graph than power-law graph.



Experiment Time-step

- Each search is initiated by a peer residing at a randomly chosen node and the number of search items (n_s) found within 50 time steps from the start of the search is calculated.
- The search output (n_s) is averaged over 100 different searches (a generation), whereby we obtain N_s, where $N_s = \frac{\sum_{i=1}^{100} n_s}{100}$.
- The graph shows experiment carried on random network with ROP and SRW algorithm.
- x-axis of the graph shows the generation number while the y-axis represents the average number of search items (N_s) found in the last 100 searches.The proliferation constant - ρ for ROP is set to 3.
- The fluctuations occur due to the difference in the availability of the searched items selected at each generation. However, we see that on the average, search efficiency of ROP is almost 2.5-times higher than that of SRW. (For ROP, the number of hits ≈ 157, while it is ≈ 64 for SRW.)



Discussion and Outlook

In this work, we have produced detailed experimental results showing that agents using the simple immune-inspired concept of proliferation can cover the network more effectively than by mere random walk. This effectivity is demonstrated across the two major types of Internet topologies. This result about agent behavior, we believe is a fundamental result and can be applied beyond the domain of the proposed $p2p$ search application. However, a detailed theoretical analysis to explain these interesting results has to be undertaken in the future to explore the full potential of proliferation algorithms.

Further Reading

- *Design of a Robust Search Algorithm for P2P Networks* :- Niloy Ganguly, Geoff Canright, Andreas Deutsch; 11^{th} *International Conference on High Performance Computing*, Bangalore, India, 19-22 December 2004.
- *A Cellular Automata Model for Immune Based Search Algorithm* :- Niloy Ganguly, Andreas Deutsch; 6^{th} *International Conference on Cellular Automata for Research and Industry, ACRI* October, 2004, Amsterdam, Netherlands.
- *Design Of An Efficient Search Algorithm For P2P Networks Using Concepts From Natural Immune Systems* :- Niloy Ganguly, Geoff Canright, Andreas Deutsch; 8^{th} *International Conference on Parallel Problem Solving from Nature*, Birmingham, UK, 18-22 September 2004.
- *Developing Efficient Search Algorithms for P2P Networks Using Proliferation and Mutation*:- Niloy Ganguly, Andreas Deutsch. *International Conference on Artificial Immune Systems*, Catania, Italy, 13-16 September 2004.