Stable Service Placement on Dynamic Peer-to-Peer Networks: A Heuristic for the Distributed k-Center Problem

Evan A. Sultanik and **William C. Regli**
Drexel University
3141 Chestnut Street,
Philadelphia, Pennsylvania, 19104-2875

Abstract

The proliferation of wireless networks has underscored the need for systems capable of coping with sporadic network connectivity. The restriction of communication to neighboring hosts makes determining the global state especially difficult, if not impractical. This paper addresses the problem of coordinating the positions of an arbitrary number of services, encapsulated by mobile agents, in a dynamic peer-to-peer network. The agents' collective goal is to minimize the distance between hosts and services, even if the topology is changing constantly. We propose a distributed algorithm to efficiently calculate the stationary distribution of the network. This can be used as a hill climbing heuristic for agents to find near-optimal locations at which to provide services. Finally, we show that the agent-based hill climbing approach is temporally-stable relative to the instantaneous optimum.

Introduction

Dynamic, peer-to-peer networks impose numerous constraints that render most centralized algorithms inapplicable. Temporal dynamism implies the need for online approaches, often requiring heuristics and search techniques from the field of artificial intelligence. The proliferation of these networks, due to advances in both peer-to-peer and wireless technology, has created a rich problem space. This paper focuses on the k-center problem of dynamic, peer-to-peer networks.

A *dynamic peer-to-peer network* is a group of nodes that may not always be fully-connected, and whose connectivity may be in constant flux. A *mobile ad hoc network* (MANET) is a specific type of dynamic peer-to-peer network in which nodes may move; the network topology is determined by the nodes' spatial interrelationships. In the context of this paper, a "*mobile agent*" refers to any program capable of halting its execution and migrating to another *host*, at which the program will continue execution. A "*host*," therefore, is any node on the network capable of receiving and executing a program.

Given a fixed network topology, it is relatively straightforward to approximate solutions to problems such as k-center, modulus their complexity. However, distributing topological information throughout these networks is very expensive, since communication is restricted to neighboring nodes (RoyChoudhury, Bandyopadhyay, & Paul 2000). Even in optimal conditions, at least n messages are required to update a network of n nodes when a change in the topology occurs. In highly dynamic networks, such as MANET, maintaining global state information is infeasible.

Many see a solution to the lack of global state information in the paradigm of agency. Specifically, mobile, multi-agent collectives have been found viable to supplant centralized and static-network algorithms for routing (Migas, Buchanan, & McArtney 2003), itinerary optimization (Qi & Wang 2001), information retrieval (Das, Shuster, & Wu 2003), distributed constraint satisfaction (Mailler & Lesser 2004), service discovery (Ratsimor *et al.* 2004), and service-based computing (Kopena *et al.* 2005). In the setting of dynamic, peer-to-peer networks, services can be encapsulated by mobile agents. One advantage to this approach is that services can migrate to portions of the network to optimize stability and bandwidth usage. For example, to optimize stability the service might migrate to nodes of high degree (that are less likely to become disconnected). In order to optimize bandwidth usage, the service might migrate to a center of the network topology.

A problem then arises if multiple homogeneous services coexist on the network: how can the services coordinate to collectively optimize bandwidth? This is an asymmetric variant of the k-center problem. The k-center problem is $\mathcal{NP}$-Hard (Kariv & Hakimi 1979), and it was recently shown that the weighted, asymmetric variant can be approximated in $O(log^{*}n)$ time (Gørtz & Wirth 2003). However, this approximation relies on knowledge of the entire network graph. Also, temporal network dynamism is not taken into account.

It is important to note that services need to migrate to new optimal centers as a result of every topological change. Even if services are optimally positioned, the bandwidth required to migrate to the next instantaneous optimum might outweigh any advantage attained from having an optimal positioning. For example, consider the sequence of topological changes pictured in Figures 1(a)–(c). These topologies were generated in simulation over a period of 10 simulation quanta. Even during this brief period of time, the optimal 3-center cover (depicted by the boxed vertices) varied drastically.

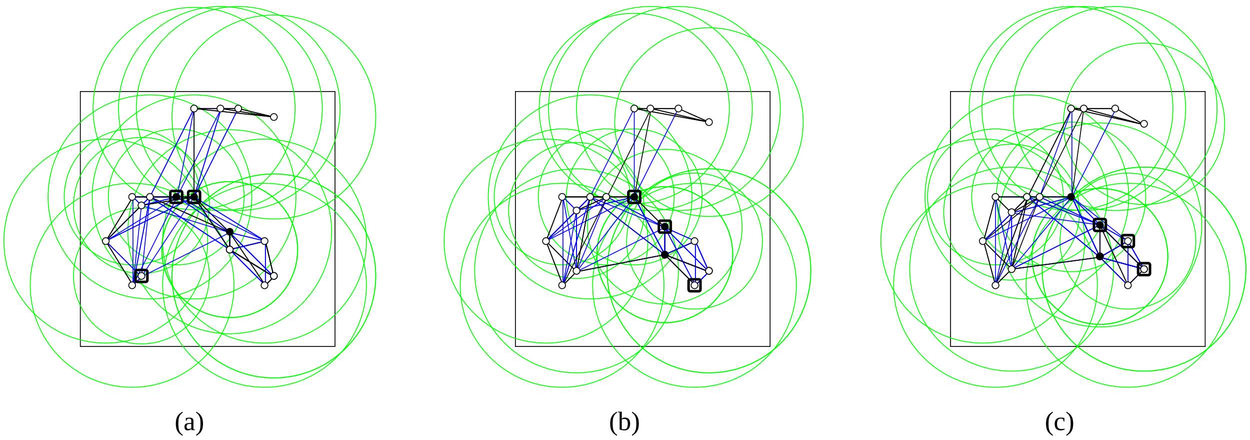

Figure 1: Three snapshots of a simulated MANET of 20 hosts with varying radio ranges. Snapshots (a), (b), and (c) occur chronologically, over a 10 second simulated span. Boxed vertices represent the instantaneous optimal 3-center cover. Filled vertices represent the 3-center cover calculated using the proposed heuristic.

The remainder of this paper proposes a heuristic for service placement in the temporal asymmetric weighted k-center problem. Additionally, all computation is distributed using a multi-agent system; global state information is not required. We empirically validate the heuristic by showing that it is near-optimal. Furthermore, we show the heuristic is temporally-stable relative to the instantaneous optimum.

Formalization

Definition 1. (Vertex Eccentricity)
The *eccentricity* of a vertex v is the maximum length of a shortest path from v to any other vertex.

Definition 2. (Graph Center)
Given a graph $G = \langle V, E \rangle$, the *center* of G is the set of vertices $C \subseteq V$ with minimum eccentricity.

Definition 3. (Covering Radius)
Given a graph, $G = \langle V, E \rangle$ and a set $S \subseteq V$, the *covering radius* of $\langle G, S \rangle$ is the minimum length r such that $\forall v \in V$ there exists a path of length less than or equal to r from v to at least one $s \in S$.

Definition 4. (Asymmetric Weighted k-Center Problem)
Given $G = \langle V, E \rangle$, a directed graph with non-negative edge weights, and positive integers k and r, does there exist a set S such that $S \subseteq V$, $|S| = k$, and the covering radius of $\langle G, S \rangle$ equals r?

Definition 5. (Stationary Distribution)
Given the transition probabilities $P(u, v)$ for a random walk on a graph (i.e. the probability that the walk will visit v immediately after u), the *stationary distribution*, $\pi(v)$, satisfies

$$\pi(v) = \lim_{i \to \infty} f P^i(v),$$

where $f : V \to \mathbb{R}$ is any initial distribution such that

$$\sum_v f(v) = 1.$$

Definition 6. (Agent Location)
Given a set of agents, A, and a set of hosts, H, we define a function, η, mapping an agent to the host on which it is physically located in a given state:

$$\eta : A \to H.$$

Inversely, the function η^{-1} maps a host to the set of agents located on that host.

Problem Statement

Given:

- an agent, $a \in A$;
- the host on which a is located: $h = \eta(a)$;
- the set of neighboring hosts,

$$N = \{h' : (h, h') \in E\} \subseteq H;$$

- the stationary distribution value of h: $\pi(h)$;
- the stationary distribution values of each $h' \in N$,

determine the optimal successor host $h' \in N$ to which to migrate in order to minimize communication cost throughout the network. "Communication cost" entails

- the cost of other agents' communication with a;
- the cost of other agents' communication with services homogeneous to a;
- and the cost of a's migration.

Observations

A graph's *stationary distribution* is directly related to the edge density of the graph and also the visitation frequency of a random walk (Chung 1994). The stationary distribution can also be modeled as a Markov process. Given the graph's adjacency matrix, M, the probability transition matrix, R, and any real number damping factor $\epsilon \in (0,1)$, the stationary distribution will be the vector π equal to the principle right eigenvector of

$$(\epsilon R + (1-\epsilon)M)^T .$$

Knowledge of the current global adjacency matrix, M, is not available on every host in the network. However, we can approximate the stationary distribution by deploying a set of random-walking agents on the network. Having the agents chose successor hosts uniformly will ensure a probability transition matrix, R, such that $(\forall i,j : R_{i,j} = 1)$. A host h can then simply record agent visitation frequency to determine its stationary distribution value $\pi(h)$. We have empirically showed that the center of the graph has the following property (Sultanik & Regli 2004):R

$$C = \left\{ v \in V : \pi(v) = \max_{u \in V} \pi(u) \right\} \qquad (1)$$

However, the authors are not aware of any formal proof of this property. For all intents and purposes, the host with the highest visitation frequency will be the optimal solution to the 1-center problem.

It is important to note that stationary distribution values are *not* guaranteed to increase monotonically along the shortest path from a vertex to a center. Since each host only knows its own visitation frequency, the non-monotonicity of the frequencies means that a hill climbing search for the center may encounter local maxima. However, as we will show, local maxima are not necessarily bad. In point of actuality, the following sections empirically reveal that local maxima in the stationary distribution provide near-optimal center cover. Furthermore, we show that using the random-walking agent visitation frequency as a heuristic provides a temporally-stable cover, relative to optimal.

Approach

The general approach is as follows:

- deploy a set of random-walking agents on the network;
- each host records random-walking agent visitation frequency; and
- each service uses the visitation frequencies as a hill climbing heuristic, stopping at the first maximum found that is not currently inhabited by another service.
- in the case that another homogeneous service is already located at the local maximum, perform a self-avoiding random walk to find a new gradient.

This process is depicted in Figure 2.

The primary advantage of this approach is that it is completely decentralized; global state information is not required.

Empirical Validation

Our approach is validated in MATES (Sultanik, Peysakhov, & Regli 2005), a discrete event simulator. The *City Section Mobility Model* (Camp, Boleng, & Davies 2002) is used for host movement, and link connectivity is determined by the

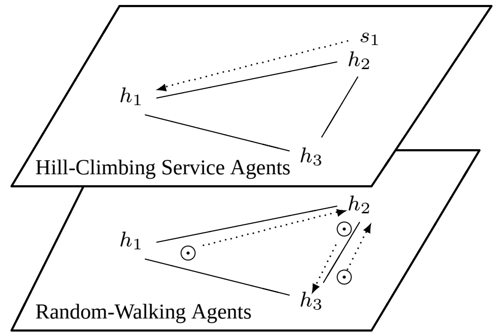

Figure 2: Random-walking agents, $\odot$, determine the stationary distribution of the network, which is used as a hill climbing heuristic for service-providing agents. In this example, service s_1 migrates to h_1 because $\pi(h_1) > \pi(h_2)$.

	Number of Services				
	1	3	5	7	9
% of Optimal	70	70	63	55	50
Stability	0.2	0.5	0.4	0.2	0.2

Table 1: Analysis of the hill climbing approach: mean statistics from Figures 3 and 4.

Euclidean distances between hosts. A series of 30 runs of 5000 simulation quanta (i.e. seconds) were conducted on a network of 20 hosts and 20 random-walking agents. For unique sequence of topology changes in each run, the proposed heuristic is tested in placing 1, 3, 5, 7, and 9 services. Aggregate data collected from the simulation are presented in Table 1.

Optimality

The accuracy of the heuristic is validated by comparison against the instantaneous optimal k-center cover. In the worst case, every host in the network will need to communicate with a service at a given point in time. Therefore, the optimality of a cover is determined by the sum of the distances of hosts to their closest service. Figure 3 presents a running average of the percentage of optimality of the heuristic over a single simulation run.

It is interesting to note that the optimality of the cover seems to increase in an inverse linear relationship with the number of services. The highly correlated nature of the optimality curves pictured in Figure 3 suggest that network topology changes play the major role in determining optimality of the heuristic (no matter the number of services).

Temporal Stability

The *temporal stability* of a cover is the communication overhead due to a topological change. For each simulation run, we compare the temporal stability of the instantaneous optimum against that of the proposed approach. In other words, we are given an optimal cover, S_{opt}, the optimal cover after a topological change, S'_{opt}, the cover calculated using

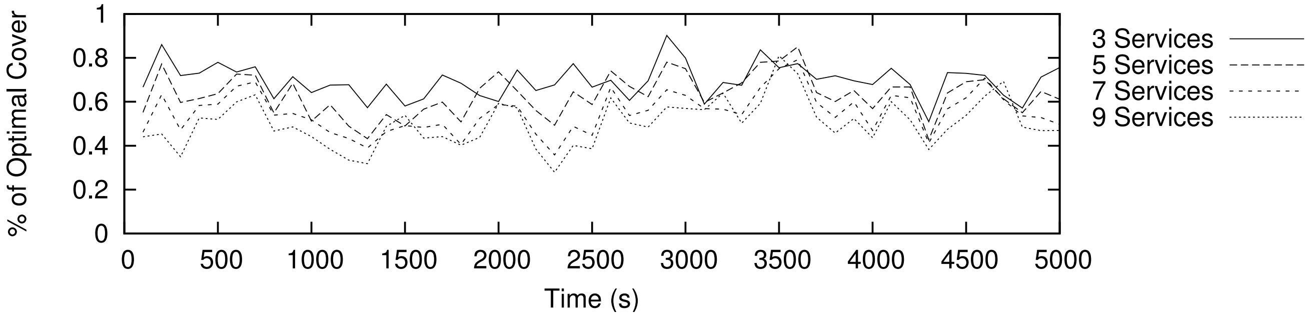

Figure 3: Running average (over 100s) of the percentage of an optimal cover when placing 3, 5, 7, and 9 services in a network of 20 hosts. The higher the percentage, the higher the optimality of the heuristic.

the approach, S_{app}, and the cover calculated after the topological change, S'_{app}. Given the benefit of the doubt, we calculate the optimal route for each service to take through the network in order for the new covers to become S'_{opt} and S_{app} after the topological change. Figure 4 presents a running average of the difference in the communication costs between the optimal cover and the heuristic's cover. Positive values indicate that the optimal cover incurred higher communication overhead than the proposed approach.

Limitations and Future Work

The communication cost of the heuristic could be drastically improved by employing an intelligent algorithm for dealing with conflict. In other words, when the number of servers is high, the contention for local maxima increases drastically. If a service's search encounters a local maxima that is already inhabited, an efficient way to locate a new, unused gradient is needed. One approach to this problem might be using an ant-inspired algorithm in which agents leave a repellant pheromone trail along their chosen gradient.

As mentioned above, to the authors knowledge no formal proof of Equation 1 exists. Analysis of this graph theoretic property might provide insight into the probability the local maxima actually being the global optima. Likewise, analysis might also shed light into how many random-walking agents are required to provide a sufficiently up-to-date estimate of a host's stationary distribution value.

The number of agents required could then be extended as a quality of service metric for the system. In other words, one would ideally have a function mapping a desired visitation frequency and stationary distribution accuracy to the number of random-walking agents required in the network.

The authors also wish to theoretically formulate tight bounds on the optimality of the heuristic, possibly determining if it is an approximation of the distributed k-center problem.

Conclusions

This paper provides a novel approach to service placement on dynamic, peer-to-peer networks. We have shown that in highly dynamic networks, in which the topology changes constantly, maintaining the instantaneous optimal incurs high communication overhead. Knowing the instantaneous optimal in the first place requires knowledge of the global state (including the topology), which is an unreasonable assumption in these networks. The proposed approach is distributed and does not rely on global state information. Furthermore, it was validated in simulation and shown to be near-optimal.

Communication overhead is often not considered in distributed artificial intelligence. This constraint will become increasingly influential as dynamic peer-to-peer networks become more and more prevalent. We hope that our research will further interest in this area, motivating the confluence of active networking and artificial intelligence.

References

Camp, T.; Boleng, J.; and Davies, V. 2002. A survey of mobility models for ad hoc network research. *Wireless Communication & Mobile Computing (WCMC): Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications* 2(5):483–502.

Chung, F. R. 1994. *Spectral Graph Theory*. Number 92 in Regional Conference Series in Mathematics. American Mathematical Society. chapter 1, 14.

Das, S.; Shuster, K.; and Wu, C. 2003. Dynamic information retrieval optimization using mobile agents. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi Agent Systems*, 968–969.

Gørtz, I. L., and Wirth, A. 2003. Asymmetry in k-center variants. In *Proceedings of the 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, Lecture Notes in Computer Science. Springer Verlag, Berlin.

Kariv, O., and Hakimi, S. 1979. An algorithmic approach to network location problems. I. the p-centers. *SIAM Journal of Applied Mathematics* 37:513–38.

Kopena, J. B.; Naik, G.; Peysakhov, M. D.; Sultanik, E. A.; and Regli, W. C. 2005. Service-based computing for agents on disruption and and delay prone networks. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents & Multi-Agent Systems*. (To Appear).

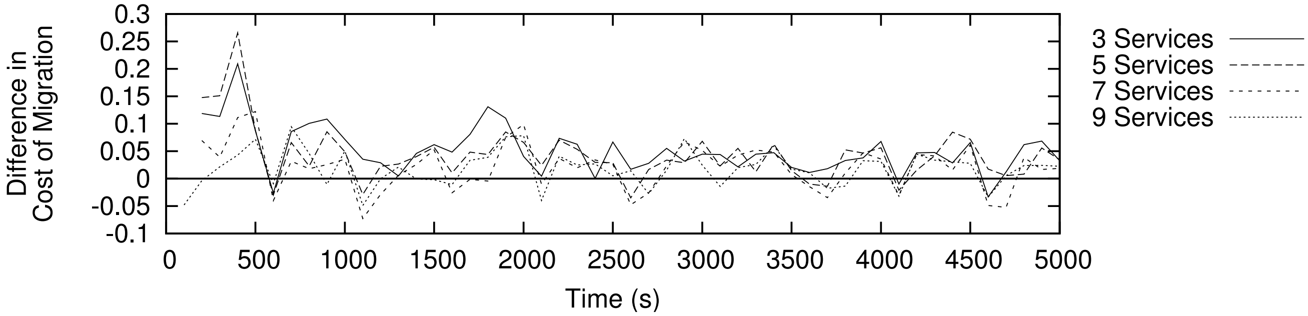

Figure 4: Running average (over 100s) of the difference between the distances optimal centers had to move after a topological change, compared to the distances calculated centers had to move. Positive values indicate that the heuristic is more stable than the optimal positioning. Values along the Y-axis are link connectivity metrics used in simulation; the higher the value, the higher the communication cost.

Mailler, R., and Lesser, V. 2004. Using cooperative mediation to solve distributed constraint satisfaction problems. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems*, 446–453.

Migas, N.; Buchanan, W. J.; and McArtney, K. A. 2003. Mobile agents for routing, topology discovery, and automatic network reconfiguration in ad-hoc networks. In *Proceedings of the 10th IEEE Conference and Workshop on the Engineering of Computer-Based Systems*, 200–206.

Qi, H., and Wang, F. 2001. Optimal itinerary analysis for mobile agents in ad hoc wireless sensor networks. In *Procedings of the International Conference on Wireless Communications*, 147–153.

Ratsimor, O.; Chakraborty, D.; Joshi, A.; Finin, T.; and Yesha, Y. 2004. Service discovery in agent-based pervasive computing environments. *Mobile Networks and Applications* 9(6):679–692.

RoyChoudhury, R.; Bandyopadhyay, S.; and Paul, K. 2000. A distributed mechanism for topology discovery in ad hoc wireless networks using mobile agents. In *Proceedings of the 1st ACM international symposium on Mobile ad hoc networking & computing*, 145–146. IEEE Press.

Sultanik, E. A., and Regli, W. C. 2004. Service discovery on dynamic peer-to-peer networks using mobile agents. In *Proceedings of the Third International Workshop on Agents and Peer-to-Peer Computing*, Lecture Notes in Computer Science. (In Press).

Sultanik, E. A.; Peysakhov, M. D.; and Regli, W. C. 2005. Agent transport simulation for dynamic peer-to-peer networks. In *Proceedings of the Sixth International Workshop on Multi-Agent-Based Simulation*.