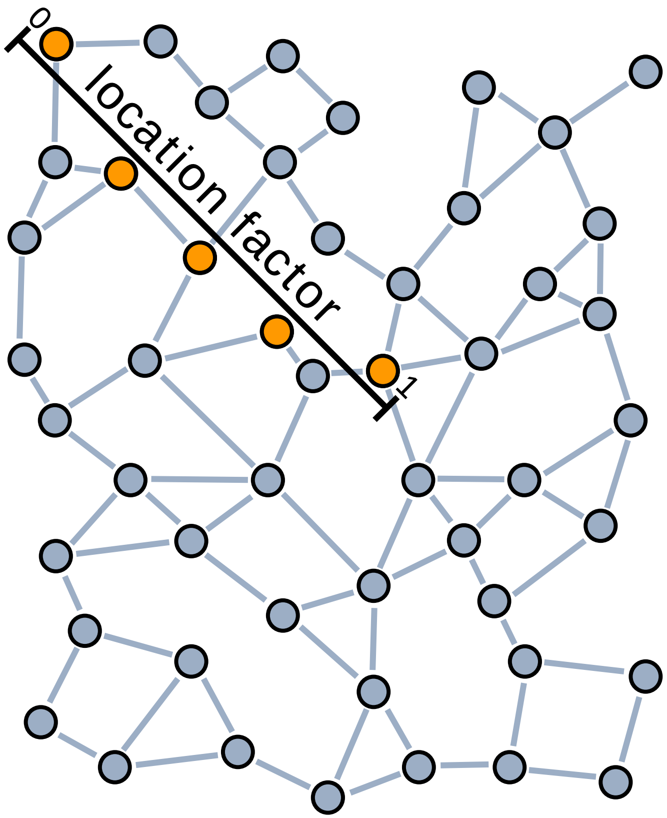
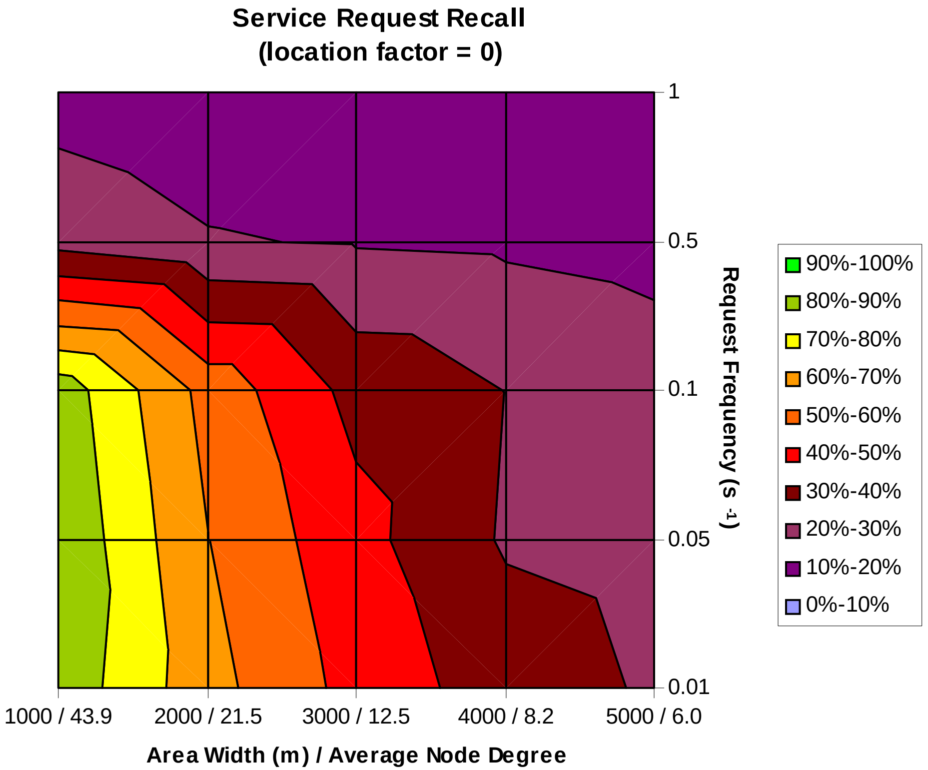
A Survey of Current Directions in Service Placement in Mobile Ad-hoc Networks

Georg Wittenburg and Jochen Schiller
Freie Universität Berlin

Middleware Support for Pervasive
Computing Workshop (PerWare '08)
Hong Kong, China

Motivation


0
location factor
1


Service Request Recall
(location factor = 0)
1
0.5
0.1
0.05
0.01
Request Frequency (s -1)
1000 / 43.9
2000 / 21.5
3000 / 12.5
4000 / 8.2
5000 / 6.0
Area Width (m) / Average Node Degree
90%-100%
80%-90%
70%-80%
60%-70%
50%-60%
40%-50%
30%-40%
20%-30%
10%-20%
0%-10%

Service Placement

- Dynamically adapt location of services to changing network topology and service demand of clients:
  - Decide on suitable number of service instances
  - Identify appropriate nodes to act as servers
  - Migrate and/or replicate service instances
- Minimize cost of providing service to client nodes:
  - Network traffic, radio interference, energy expenditure, ...
  - Availability, latency, ...
- Key questions:
  - *Where* to place service instances?
  - *How many* service instances for cost optimal operation?
  - *When* to adapt current configuration of services?
  - *How* to transfer services between nodes?

Outline

- Background
  - Services
  - Applicability
  - Facility Location Theory
- State of the Art
  - Brief Summaries of Surveyed Approaches
  - Classification by Communication Cost
  - Classification by Context of Research
  - Evaluation
- Conclusion

Services

- Term “service” in this context:
  - Software component executed on one or several nodes
  - Reacts to service requests received from clients through well-defined interface according to service-specific protocol
  - Several different services may be active simultaneously

- Classification of services:
  - Node-specific vs. node-independent
  - Centralized vs. distributed
  - Monolithic vs. composite
  - Message-based vs. streaming

Applicability

- Network properties:
  - Changing topology, i.e. subject to mobility and/or churn
  - Changing regional service demand
  - Limited heterogeneity of devices

- Types of services:
  - Data dissemination or processing:
    - Directory services
    - Multimedia streaming
  - Network topology:
    - Cluster heads
    - Landmark nodes

Facility Location Theory

- Two theoretical problems related to service placement:
  - Common input:
    - Sets of facilities and clients (bipartite graph)
    - Cost functions for service provisioning and new facilities
  - *k*-Median Problem:
    - Place *k* facilities so that total service cost is minimized
    - ➢ Decision on placement of a given number of service instances
      - Special case: Placement of a centralized service ($k=1$)
  - Facility Location Problem:
    - Choose number and location of facilities so that sum of service cost and facility cost is minimized
    - ➢ Decision on number and placement of service instances
- Both problems are NP-hard
- Approximations exist, but commonly require exact knowledge about inputs

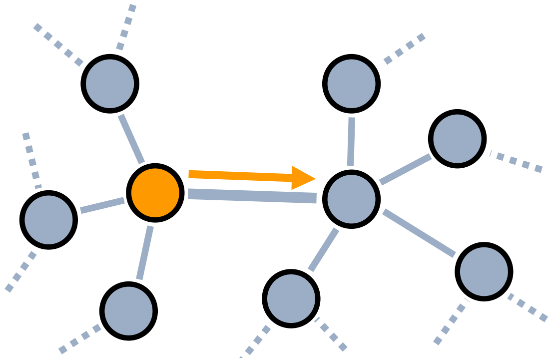
State of the Art

Survey of following publications:

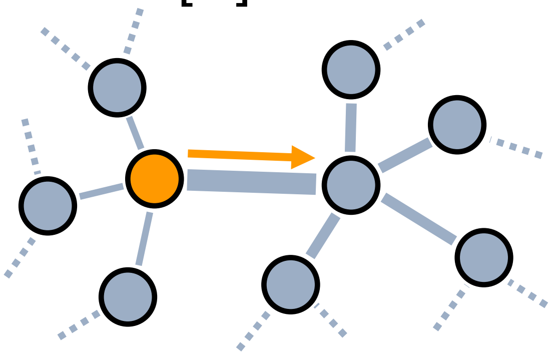
1. P. Bellavista, A. Corradi, and E. Magistretti. REDMAN: An Optimistic Replication Middleware for Read-only Resources in Dense MANETs. *Journal on Pervasive and Mobile Computing*, 1(3):279–310, Aug. 2005.
2. C. Frank and K. Römer. Distributed Facility Location Algorithms for Flexible Configuration of Wireless Sensor Networks. In *3rd IEEE Intl. Conf. on Distributed Computing in Sensor Systems*, Santa Fe, USA, June 2007.
3. T. Furuta, M. Sasaki, F. Ishizaki, A. Suzuki, and H. Miyazawa. A New Clustering Algorithm Using Facility Location Theory for Wireless Sensor Networks. Technical Report NANZAN-TR-2006-04, Nanzan Academic Society, Mar. 2007.
4. D. Krivitski, A. Schuster, and R. Wolff. A Local Facility Location Algorithm for Large-Scale Distributed Systems. *Journal of Grid Computing*, 2006.
5. N. Laoutaris, G. Smaragdakis, K. Oikonomou, I. Stavrakakis, and A. Bestavros. Distributed Placement of Service Facilities in Large-Scale Networks. In *26th Annual IEEE Conf. on Computer Communications*, Anchorage, USA, May 2007.
6. B. Li and K. H. Wang. NonStop: Continuous Multimedia Streaming in Wireless Ad Hoc Networks with Node Mobility. *IEEE Journal on Selected Areas in Communications*, 21(10):1627–1641, Dec. 2003.
7. H. Liu, T. Roeder, K. Walsh, R. Barr, and E. G. Sirer. Design and Implementation of a Single System Image Operating System for Ad Hoc Networks. In *3rd Intl. Conf. on Mobile Systems, Applications, and Services*, Seattle, USA, June 2005.
8. T. Moscibroda and R. Wattenhofer. Facility Location: Distributed Approximation. In *24th ACM Symp. on the Principles of Distributed Computing*, Las Vegas, USA, July 2005.
9. K. Oikonomou and I. Stavrakakis. Scalable Service Migration: The Tree Topology Case. In *5th Annual Mediterranean Ad Hoc Networking Workshop*, Lipari, Italy, June 2006.
10. F. Sailhan and V. Issarny. Scalable Service Discovery for MANET. In *3rd IEEE Intl. Conf. on Pervasive Computing and Communications*, Kauai, USA, Mar. 2005.

Handling a Centralized Service (1)

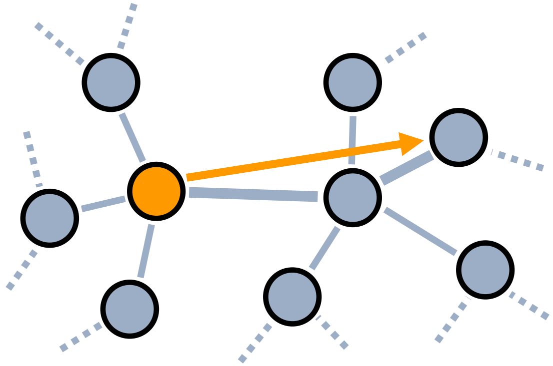
- Migrate service to neighbor that routes most traffic [7]



- Migrate service to neighbor that routes more than half of the traffic [9]

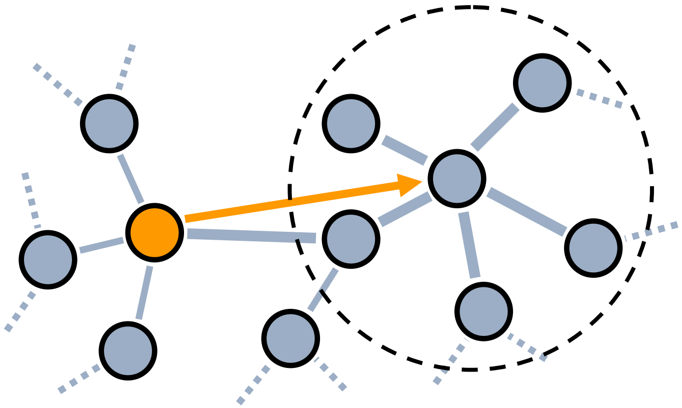


- Migrate service to node that causes most traffic [7]

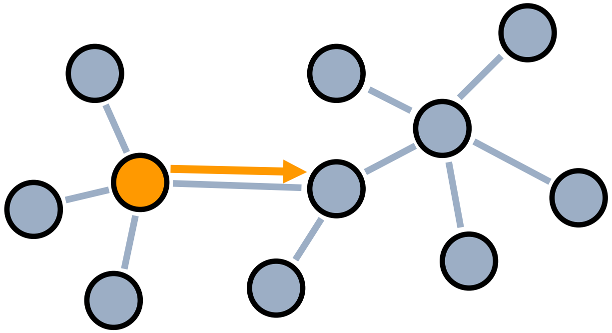


Handling a Centralized Service (2)

- Migrate service to one-hop cluster that causes most traffic [7]

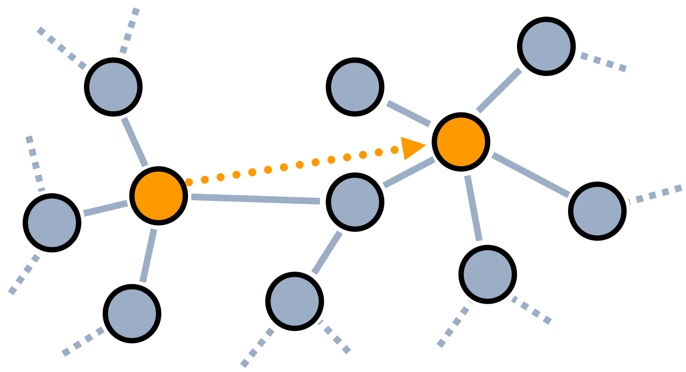


- Migrate service one hop towards most distant node until topological center of network is reached [1]

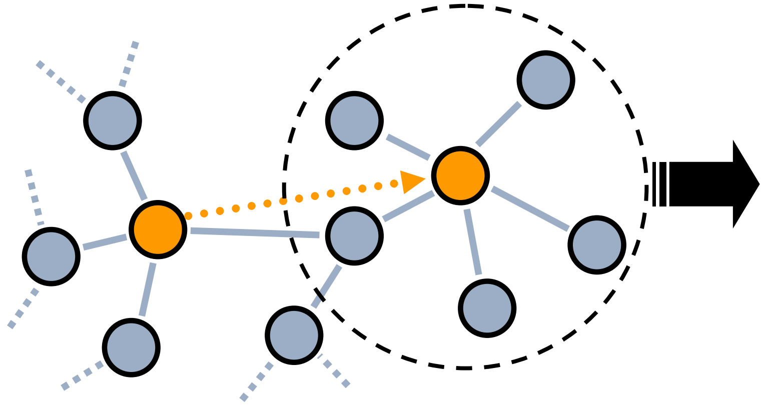


Handling a Distributed Service (1)

- Replicate service maximizing number of new clients [10]



- Replicate service instance per mobility group [6]



Handling a Distributed Service (2)

- Centralized solutions to facility location problem:
  - Centrally collect all information [3]
  - Calculate service demand by local enquiries and estimating demand of distant nodes [5]
    - Demand of distant nodes is mapped to nodes *n* hops away
- Distributed solutions to facility location problem:
  - Iteratively reach agreement between potential servers and clients on optimal mapping [2]
    - Potential servers send advertise to clients
    - Clients reply to most cost-efficient advertisement
    - Server is started if replies match advertisements
  - Distributed hill climbing using majority votes [4]
    - Nodes agree on next step of hill climbing approximation using a low-overhead majority voting primitive
  - Round-based fractional linear program [8]
    - Iteratively approximate parameters of primal and dual problem

Classification by Communication Cost

Supported type of service / **Main protocol mechanism**	Centralized service	Distributed service
Passive monitoring	[9]	
Piggybacking	[7]	[3[1], 6[2]]
Limited broadcasts		[10[3]]
Iterative limited broadcasts	[1]	[2, 4, 5]
Iterative unlimited broadcasts		[8]

[1]Service placement algorithm is run centrally on base station
[2]Places service instance on any node within mobility group
[3]Only supports placement of one service instance per neighborhood

Classification by Context of Research

Two major approaches in service placement research:

- Middleware:
  - Service abstraction in existing component system
  - Service placement as additional feature
  - Tend to consider simple cases (e.g. centralized service) using heuristics
- Facility Location Theory:
  - Networking as application of theory, but centralized approximations do not work in distributed systems
  - Complex service placement with focus on algorithmic correctness and complexity
  - Tend to neglect cost of communication (service discovery, route discovery, shared media access, ...)

Evaluation

- Placement of single centralized service just by monitoring service requests
- Placement of distributed service instances requires active exchange of information via iterative limited broadcasts

- Mostly self-contained studies, few comparisons
- Quantitative evaluation only against systems without placement

- Not explicitly considered so far:
  - Several disjoint services on same network
  - Traffic between service instances
  - Timing of placement decisions
  - Integration with service location or routing protocols

Conclusion

- Service placement is beneficial for certain MANETs.
  - Scenarios with node-independent services
  - Dynamically changing regional network conditions
  - ➢ “Ad-hoc load balancing”

- Current approaches deal with establishing the optimal location and (in part) number of service instances.
  - Tradeoff between quality of approximation and communication overhead

- Critical open question is *when* to adapt configuration.
  - Transient changes in regional demand
  - Overhead of migration/replication and service/route discovery