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

![Diagram showing service request recall with varying location factors.]

- **Location Factor**: 0, 0.25, 0.5, 0.75, 1
- **Request Frequency (s⁻¹)**: 0.01, 0.05, 0.1, 0.5, 1
- **Area Width (m)**: 1000, 2000, 3000, 4000, 5000
- **Average Node Degree**: 43.9, 21.5, 12.5, 8.2, 6.0

Service Request Recall
(location factor = 0)

- 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:


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]
• 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

<table>
<thead>
<tr>
<th>Supported type of service</th>
<th>Main protocol mechanism</th>
<th>Centralized service</th>
<th>Distributed service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive monitoring</td>
<td></td>
<td>[9]</td>
<td></td>
</tr>
<tr>
<td>Piggybacking</td>
<td></td>
<td>[7]</td>
<td>[3(^1), 6(^2)]</td>
</tr>
<tr>
<td>Limited broadcasts</td>
<td></td>
<td></td>
<td>[10(^3)]</td>
</tr>
<tr>
<td>Iterative limited broadcasts</td>
<td></td>
<td>[1]</td>
<td>[2, 4, 5]</td>
</tr>
<tr>
<td>Iterative unlimited broadcasts</td>
<td></td>
<td></td>
<td>[8]</td>
</tr>
</tbody>
</table>

\(^1\) 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
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