Energy Efficient and Adaptive Service Advertisement, Discovery and Provision for Mobile Ad Hoc Networks

Ph.D Presentation

Christopher N. Ververidis

Advisor: Professor George C. Polyzos

Athens University of Economics and Business
Department of Computer Science
Motivation I

• Why/when do we need MANETs?
  – Infrastructure is not always available.
    • Places/situations where it is impossible or impractical to setup an infrastructure
      – Remote and hard to reach areas, Battlefields etc.
  – Infrastructure may be very costly to use.

• What is their purpose?
  – To enable mobile users to exchange data and use each other’s services.
  – Use Cases:
    • Disaster relief teams, Warfront activities, Vehicular networks, Social P2P networking (e.g. gaming)...)
Motivation II

• Solving the problems of connectivity is not enough for the adoption of MANETs.
  – We need efficient Service Discovery protocols too!

• SD protocols for MANETs **have to:**
  1. Cope with the dynamism of a MANET
     – Bandwidth
     – Mobility
     – Channel characteristics
  2. “Respect” the limited device capabilities
     – Energy
     – Memory
     – Storage
Related Work

- **SD in fixed networks**
  - Energy & Bandwidth & Connectivity not a problem
  - **Use of Flooding or Multicasting**
    - No scalability in MANETs (great overhead)
  - **Use of Centralized directories**
    - Single point of failure
    - Assumption of availability
  - **DHTs**
    - Physical proximity not taken into account (Hop count is important)

- **Approaches for SD in fixed networks are not viable in MANETs due to:**
  - Limited device energy
  - Limited bandwidth
  - Distributed nature (No well-known central management points)
  - Unexpected connection loss: channel variability, mobility, switching off
The Problem with Application Layer based SD Protocols

- If Service Discovery is implemented above the routing layer:
  - **Two** message producing processes coexist:
    - one for communicating **service** information among nodes
    - one for communicating **routing** information among nodes
  - A node is forced to perform the battery-draining operation of receiving and transmitting packets **multiple times**
    - IEEE802.11: collisions increase substantially since both processes make extensive use of broadcasting
Viable Solution: Integrated Protocols

• Integration of service discovery and route discovery
  – Viable approach to reduce energy consumption
    • Client side: Reduce overhead for searching for services
    • Server side: Reduce overhead for advertising services

  – Approach (Perkins et al.): Discover services and routes to service providers at the same time.
    • Proactive approaches:
      – DSDV, OLSR etc.
    • Reactive approaches:
      – DSR, AODV, etc.

  – Compared versus application layer SD protocols using global flooding
Our Approach: Hybrid Integrated Protocols

- **Routing Protocols:**
  - Proactive
    - Better for High Call to mobility Ratio
  - Reactive
    - Better for Low Call to mobility Ratio
  - Hybrid
    - Adaptable to any situation

- **We propose 2 hybrid integrated protocols:**
  - The Extended Zone Routing Protocol (E-ZRP)
  - The Adaptive service and route discovery protocol (AVERT)
The E-ZRP protocol

- **E-ZRP operation:**
  - Based on ZRP (Hybrid Protocol)
  - **Piggybacks UUIDs** for services into routing messages
    - Simultaneous route and service advertisement and discovery
  - **E-ZRP proactive part** (Intra-Zone Routing Protocol (IARP))
    - Allows the continuous monitoring of “nearby” services and routes
  - **E-ZRP reactive part** (Inter-Zone Routing Protocol (IERP))
    - Allows the on-demand request for services and routes not found in a node’s proximity (*efficient search – not flooding*)

![Diagram](image.png)

- **Border node**
- **Neighbor node**
- **Zone radius = 2 (hops)**
- **Bordercasting**
Experimental Evaluation I

- **Our goals**
  - To provide an experimental assessment of energy savings obtained by implementing service discovery at the routing layer
  - To provide an analysis of the factors that have an impact on service availability in MANETs.

- **Evaluated:**
  - *Energy consumption, Discoverability*
  - **Service Availability Duration (SAD):** The time elapsed from the first discovery of a provider of a specific service by a node to the time that the last route to any provider of the same service is not available to the node.
    - Why do we care about SAD?
      - Measures the availability of services in a MANET in terms of time in a straightforward manner.

- **Used the Qualnet Simulator**
  - Discrete Event Simulator - Full Stack simulations (detailed model library)
    - Extended with (E-ZRP,AVERT,IZR, SPIZ, APS) ~8000 lines of C code
Experimental Evaluation II

• **Comparison against APS** (*AP*lication layer-based *S*ervice discovery protocol)
  – APS operates similarly to E-ZRP for proactive service discovery, using scoped flooding of service advertisements.
  – Broadcast timers are the same for APS and E-ZRP.
  – APS is as lightweight as possible, using small packets (info only on the originator’s service).
  – APS does not store route information
    • We compare E-ZRP to APS over ZRP
Energy Consumption vs. Service Discoverability (Proactive Part)

- Integrated approaches are more energy efficient due to less messaging overhead
- How much better is E-ZRP vs. APS?
  - Settings
    - 250 nodes, 1000 sec simulation time, RWP (when mobile), unique service/node

<table>
<thead>
<tr>
<th>Broadcast interval</th>
<th>A</th>
<th>200 s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>160 s</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>80 s</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>40 s</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>20 s</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>15 s</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>10 s</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>5 s</td>
</tr>
</tbody>
</table>

Longer intervals
Fewer messages transmitted
Nodes receive less services information
Discovery Delay (Reactive Part)

• **APS imposes significant delays for discovering out of zone services**
  – Each point on the diagram is an average obtained over 20 service discovery requests between different node pairs at the same distance (in hops)

• **Since IERP uses the mechanism of bordercasting, it can efficiently and quickly “scan” distant areas of the network to find the requested service**
  – Unicasting is used

• **APS takes a long time to “scan” the network since it relies on hop-by-hop broadcasting**
  – Broadcasting is used leading to collisions and hence longer intervals have to be used in forwarding
Investigation on Service Availability Duration (SAD)

- **The impact of mobility on SAD**
  - Settings
    - 20 nodes, 4000mx4000m terrain size, 400m range, RWP model
    - Every node hosts 1 out of 3 available service types, 3 hops zone radius
    - Perfect channel
      - receive packet correctly if signal/noise = 1
  - Both protocols perform similarly under a perfect channel (APS being more expensive in terms of energy consumption)
Investigation on Service Availability Duration (SAD)

- **The impact of density on SAD**
  - Settings (same as before) but:
    - High Density: 20 nodes, 2000mx2000m terrain size
    - Low Density: 20 nodes, 4000mx4000m terrain size

- The impact of mobility on SAD is more severe as network density decreases
  - Disconnections happen more easily when increasing speed for nodes moving on a larger terrain (compared to cases of smaller terrains)
Investigation on Service Availability Duration (SAD)

- **The impact of noise on SAD**
  - Settings (same as before) but:
    - 20 nodes, 2000mx2000m terrain size
    - Perfect channel: receive packet correctly if signal/noise = 1
    - Realistic channel: receive packet correctly if signal/noise = 10

- Lower SAD for APS vs. E-ZRP.
  - Using a protocol for route discovery and also a separate application layer-based protocol for service discovery collisions are more probable.
From E-ZRP to AVERT

- **E-ZRP can be tuned optimally if the call to mobility ratio is known a priori.**
  - Nodes need pre-configuration before joining the network.
- **The same zone radius should be set for all nodes.**
  - Optimal if:
    - The MANET’s call to mobility ratio remains the same.
    - All parts of the network have the same call to mobility ratio.
- **When call to mobility ratio is unknown/unpredictable** we propose the use of AVERT (Adaptive service and route discovery protocol)
AVERT

- **AVERT is based on the Independent Zone Routing Framework**
  - IZR, an improved version of ZRP, allows every node to use its own zone size, which is also dynamically adjusted.
  - The routing control traffic of IZR is decreased by 60% compared to ZRP.

- **Extensions:**
  - The extensions to IZR are essentially the same with those done to ZRP.
  - AVERT employs an additional *mechanism for adapting the frequency of sending proactive traffic* (route and service advertisements).
Adaptation mechanism for AVERT

- **Broadcasting Frequency Optimizer:** Searching for the optimal frequency for sending proactive traffic (NDP and IARP packets)
  - **Motivation:** Nodes not engaged in service invocation, discovery or provision (as clients, as providers or as intermediates) can decrease the rate of sending NDP and IARP packets to conserve energy.

- **Operation:**
  - BFO runs periodically on every node.
  - Periodically nodes monitor the DATA traffic seen locally.
    - If DATA traffic decreases increase the broadcast interval by T.
    - If DATA traffic increases decreases the broadcast interval by T.
Performance Evaluation of AVERT

- **Evaluation of AVERT against:**
  - Plain IZR (Samar et al.) – but extended to support service discovery.
  - SPIZ (Noh et al.)
    - proposes that the ATE algorithm of IZR is modified such that service providers increase their zone radius when the popularity of their service increases.

- **Tuning IZR and SPIZ for best performance:**
  - Assuming the metric: % of Completed Services/ Total Energy Expended
    - Max when Broadcast Interval = 100s (maximum permissible) for IZR and SPIZ
    - This results in unacceptable success ratios (52%-65% of the maximum achievable)
      - could be even worse if we allowed intervals>100s ----→ unusable

- We define the following metric (service efficiency $\sigma$):

  \[
  \sigma = \text{Percentage of Completed Services} \cdot \frac{\text{Completed Services}}{\text{Total Energy}}
  \]

- Optimization of broadcast intervals according to $\sigma$ allows protocols to settle at acceptable success ratios (65%-91% of the maximum achievable).
Setting the T parameter of AVERT

- **Simulation Settings**
  - Scenario 1: **High** Client to Server Ratio (5.6 to 1)
  - Scenario 2: **Medium** Client to Server Ratio (1 to 1)
  - Scenario 3: **Low** Client to Server Ratio (0.17 to 1)
  - Lowest allowable broadcast interval = 10s
  - Maximum allowable broadcast interval = 100s

Optimal performance is reached using **Small values of T**

(More fine-grained adaptation)
Comparing AVERT to IZR and SPIZ

- SPIZ and IZR must have pre-configured broadcast intervals.
  - AVERT can improve performance up to 35% compared to SPIZ and IZR (when the max permissible broadcast interval is 100s).
Profit Maximization for Service Provision in MANETs
Concept and Motivation

- **Differentiation factors** for service provisioning over MANETs from service provisioning over fixed networks:
  - Service provision in MANETs is **opportunistic**
    - There are no fixed, well-known service providers.
    - Any node (individual) can be a service provider for her own benefit and for as long as she participates in the MANET, or for as long as she desires to be a service provider.
  - More unreliable and error-prone communications.
  - **Server capacity** is much more constrained
    - Communication costs are significant due to devices’ energy constraints.
    - Client selection is a major issue especially if servers want to maximize their profit.
  - **Solutions for server profit optimization must be computationally efficient** due to the processing constraints of mobile devices.
System Description

• **Basic Assumptions:**
  – Clients pay only for the service as long as the connection to the server is *active* (e.g. $/second).
  – Servers can be *cooperative* or *non-cooperative*
    - In case of cooperative servers, the servers are synchronized and can communicate via a side channel.
    - An incentives scheme is used for packet forwarding.

• **Operation**
  – Periodically clients broadcast service requests.
  – Periodically servers select the clients that are to be served.
GAP Problem

• We model the problem of maximizing server profits as a Generalized Assignment Problem (GAP):

• GAP definition:
  – There are $n$ items $x_1$ through $x_n$ and $m$ bins.
  – Each item has a weight $a_{ij}$ and a value $c_{ij}$.
  – Every bin has a capacity $b_i$.
  – The problem is to find the optimal assignments of items into the bins such that:
    • The capacity constraints of the bins are not violated.
    • The total value obtained is maximized.

• This is directly applicable to our problem of mobile server profit maximization if we assume:
  – Items are clients
  – Bins are mobile servers with serving capacities $b_i$.
  – $a_{ij}$ is the amount of resources consumed at mobile server $i$ if client $j$ is to be served.
  – $c_{ij}$ is client $j$’s payment to mobile server $i$. 
GAP is not enough

- **Solving the classic GAP formulation leads to optimality if:**
  - a pay-in-advance model is assumed (unfair for the client in case of disconnection)
- **If a pay-as-you-go model is assumed, the Classical GAP does not lead to optimal selection:**

Let $p_{ij}$ be the portion of service received by client $j$ when served by server $i$.

Assume that solving the classical GAP gives an allocation in which:

- client $l$ is allocated to server $k$,
- client $n$ is not allocated to server $k$,
- $a_{kn} = a_{kl}$, $c_{kn} \leq c_{kl}$ and
- $p_{kn} \geq p_{kl}$ such that $c_{kn} * p_{kn} > c_{kl} * p_{kl}$

Then the solution of GAP is not optimal in terms of the final amount to be paid to the server $k$. 
Enhanced GAP (E-GAP)

- We propose an enhanced GAP (E-GAP) model of the problem with estimations on the proportion of service to be delivered from server \( i \) to client \( j \) (\( p_{ij} \)).

Instead of
\[
\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} c_{ij}
\]
we use
\[
\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} c_{ij} p_{ij}
\]

\( p_{ij} \) can be an estimation of the connectivity time/ serving period

We could also replace
\[
\sum_{j=1}^{n} x_{ij} a_{ij} \leq b_i
\]
by
\[
\sum_{j=1}^{n} x_{ij} a_{ij} p_{ij} \leq b_i
\]

but this would imply that the server may be over-provisioning and leave some demand unsatisfied.
Estimation of Connectivity

- Computing $p_{ij}$: $p_{ij} = \frac{\min\{\text{Serving period duration, } F_x \text{ between client } j \text{ and server } i\}}{\text{Serving period duration}}$

- Curve Fitting
- Relationship of expected connectivity duration to speed, density, number of hops:

$$F_x = (a_x \cdot \ln(Speed) + b_x) \cdot \ln(Density) + (c_x \cdot \ln(Speed) + d_x)$$

- Using $F$ a server may predict how much profit can be obtained from any given client and make near optimal selections.
Simulation Results
(cooperative servers)

- **Settings**
  - 2 servers, 20 clients, $a_{ij}=c_{ij}=1$ unit, $b_i$ is 5 or 25
  - Terrain size varies: 1250m x 1250m, 1500m x 1500m, 1750m x 1750m, 2000m x 2000m, 2500m x 2500m
  - RWP mobility model with constant speed: 3.5m/sec, 5m/sec, 7m/sec, 15m/sec
Simulation Results
(non-cooperative servers)

- Sets with common clients decrease the total profit obtained.
  - More evident in dense scenarios where the probability to select the same client is larger compared to sparse scenarios.

![Graph showing the effect of terrain size on percentage loss for non-cooperative servers compared to cooperative servers.](image-url)
Summary and Conclusions

• Proposed 2 hybrid integrated protocols (E-ZRP, AVERT)
  – E-ZRP:
    • Investigated the benefits of implementing service discovery at the routing layer instead of the application layer.
    • Proposed the Service Availability Duration metric
      – Evaluated the impact of mobility, density and noise on Service Availability.
  – AVERT:
    • High capability of adaptation to MANET conditions
      – More autonomous & flexible (zone adaptation + broadcast frequency optimization)
    • Investigated the performance of AVERT versus alternatives.

• Proposed a model for solving the problem of profit maximization from service provisioning in MANETs that takes into account client-server connectivity.
  – Proposed a method for estimating connectivity duration between clients and servers.
  – Investigated the performance of the connectivity aware method against a connectivity unaware method for various settings of mobility and density.
Future Work

• Interoperability of Service Discovery approaches
  – Multiple protocols exist (for heterogeneous environments and devices)
    • Transparent approaches -> require protocol translators
    • Translation Gateways -> not common for MANETs
  – Need for scalable approaches

• Benchmarking and analytical models for Service Discovery
  – Multiple parameters often neglected by existing approaches
    • Node mobility, density, radio link behavior
Publications - I

Journal Article

Magazine Article

Refereed Conference and Workshop Papers
Refereed Conference and Workshop Papers (continued)


Book Chapters


Papers with Poster Presentations


Thank you!

Christopher N. Ververidis

Mobile Multimedia Laboratory
Department of Computer Science
Athens University of Economics and Business
chris@aueb.gr