Ad Hoc & Sensor Wireless Networks, Vol. 1, pp. 01–24
Reprints available directly from the publisher
Photocopying permitted by license only

A Framework for Routing in Large Ad-hoc Networks with Irregular Topologies

MARC HEISSENBÜTTEL, TORSTEN BRAUN,
DAVID JÖRG AND THOMAS HUBER*

Institute of Computer Science and Applied Mathematics,
University of Bern, Switzerland
E-mail: {heissen, braun, joerg, thuber,}@iam.unibe.ch

Received: May 1, 2005, In Final Form: September 14, 2005.

In this paper, we consider routing in large wireless multihop networks with possibly irregular topologies. Existing position-based routing protocols have deficiencies in such scenarios as they always forward packets directly towards the destination. This greedy routing frequently fails and costly recovery mechanisms have to be applied. We propose the Ants-based Mobile Routing Architecture (AMRA) for optimized routing, which combines position-based routing, topology abstraction, and swarm intelligence. AMRA routes packets along paths with high connectivity and short delays by memorizing past traffic and by using ant-like packets to discover shorter paths. The geographic topology abstraction allows AMRA to cope with high mobility and large networks. Simulative evaluation indicate that AMRA finds significantly shorter paths with only marginal overhead compared to other position-based routing protocols.

Keywords: Ad-hoc networks, routing, swarm intelligence

1 INTRODUCTION

Routing in wireless multihop networks has generated a lot of interest and a large number of routing protocols have been proposed. The routing protocols have to cope with the special characteristics of these networks such as a highly

---

* The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

dynamic topology and limited resources in terms of battery capacity, bandwidth, and computational power. Existing routing algorithms can be broadly classified into topology-based and position-based routing protocols.

Topology-based routing protocols either establish routes on-demand, e.g., AODV [1], DSR [2], or proactively maintain hop-by-hop information at nodes such as OLSR [3], TBPRF [4] similar to existing protocols for the Internet. The control traffic induced by proactive protocols is substantial because update messages are propagated throughout the network for any changes in the topology. Reactive protocols employ a kind of flooding to acquire and maintain routes on-demand. The overhead induced degrades the overall network performance and is costly into terms of used network resources.

Position-based routing protocols on the other hand make use of location information in the forwarding decision, e.g., GFG/GPSR [5], GOAFR [6], BLR [7]. A node obtains its position e.g. by GPS and periodically transmits hello messages to announce this information to its neighbors. The position of the destination may either be given implicitly or can be requested from a location service such as GLS [8]. Routing is performed solely based on this position information. Packets are sent to a neighboring node, which reduces the distance to the destination. In turn it forwards them to one of its neighbors and so on until the packets reach the destination. If this greedy forwarding fails because no neighbor is closer to the destination, a recovery mechanisms is applied. Unlike topology-based protocols, position-based protocols require only little control traffic, do not need to maintain paths, and are nearly stateless as they only have to maintain neighbor information. Thus, they are generally considered as scalable and more robust to changes in the network topology than topology-based protocols. These characteristics make them the preferred choice for large and highly dynamic networks. However, their performance and behavior is far from optimal in topologies where greedy routing fails and packets are routed in recovery mode.

Lately, several routing protocols based on ant colony optimization have been proposed. These routing protocols try to adapt the problem solving abilities of ants for routing in networks. Current traffic conditions and link costs are measured by transmitting ”artificial ants”, i.e., special control packets, into the network, which mark the traveled path with an ”artificial pheromone”. The laying of the pheromone by real ants is modeled by increasing the probability of links in the routing table. Ant colony optimization has been successfully adapted several times for routing in fixed wired networks. However, they are hardly applicable in wireless multihop networks and are not able to cope with their salient characteristics.

In this paper, we introduce the Ants-based Mobile Routing Architecture AMRA that combines position-based routing, topology abstraction, and swarm intelligence. AMRA is tailored for large networks of possibly tens of thousands of nodes with irregular topologies where routing along the line-of-sight towards the destination is not possible due to obstacles like lakes and mountains. In such scenarios, AMRA tries to find useful paths over intermediate positions such that packets can be solely forwarded in greedy mode and avoid the shortcomings of routing in recovery mode. Topology abstraction is used to provide in a transparent manner an aggregated and static topology. On this topology, a routing protocol based on ant colony optimization determines good paths on a large scale. Topology abstraction is also the key to make ants-based routing scalable. Position-based routing is then applied to forward the packets physically along the selected paths.

The remainder of this paper is organized as follows. Section 2 gives a short introduction of the paradigm for ant colony optimization and gives a survey on related work. In Section 3, the protocol architecture AMRA is presented in detail. The protocols are evaluated and simulation results are given in Section 4. Finally, Section 5 concludes the paper.

2 RELATED WORK

As AMRA is based on position-based routing and ant colony optimization, we discuss in more detail related work in these fields. First, we review several proposed ant-based routing protocols. Afterwards, position-based routing protocols are discussed.

2.1 Ants-Based Routing

The principle of ant colony optimization was applied to various optimization problems such as the traveling salesman problem, the graph coloring problem, and vehicle routing problems. Recently, several routing protocols have been proposed inspired from social insects behavior for fixed, wired communication networks and for ad-hoc networks, e.g., AntNET [9] and ABC [10], CAF [11]. The ant colony principle is applied as follows to determine shortest paths in the network. Current traffic conditions and link costs are measured by transmitting artificial ants, i.e., special control packets, into the network, which mark the traveled path with an “artificial pheromone”. The laying of the pheromone is modeled by increasing the probability for the traveled links in the routing table. The pheromone may be a measure for any metric under consideration such as delay, bandwidth, and hop count. Other ants are more attracted to higher pheromone concentrations, i.e., they will follow the higher quality paths with a higher probability reinforcing the pheromone trails even more. Consequently, more and more subsequent ants choose these paths. On the other hand, the pheromone on the other paths will no longer be reinforced, decays, and eventually the trail will vanish.

Lately, various routing protocols have been introduced based on ant colony optimization for wireless multihop networks. They are either designed for small networks with only few nodes or employ mechanisms as proposed for other topology-based routing protocols to operate efficiently.

2.1.1 Ant-Colony-Based Routing Algorithm (ARA)

Ant-Colony-Based Routing Algorithm (ARA) was proposed in [12]. The routing algorithm is similar to other conventional topology-based routing protocols such as AODV [1]. Ants are only emitted on demand, i.e., if a node has to send a packet to a destination for which it does not have a path. A node broadcasts a forward ant which is flooded throughout the network. Each intermediate node stores an entry in the routing table for the forward ant. This entry contains the ant's source address, the previous hop, and a pheromone value that depends on the number of hops to the source node. When the destination node receives a forward ant, it creates a backward ant. The backward ant returns in opposite direction over the path taken by the forward ant. Like the forward ants, the backward ants create entries in the routing table at intermediate nodes and, thus, establish a bidirectional path between source and destination nodes. Data packets are used afterwards to maintain the paths established by the ants. When a node relays a data packet, it increases the pheromone values for the source and destination node of this packet over the previous and next hop respectively, thus, strengthening the path in both directions.

2.1.2 Termite

Termite routing algorithm as presented in [13] follows most closely the ant colony optimization. Each node maintains a routing table tracking the amount of pheromone on each outgoing link for all known destinations. When a packets arrives at a node, the pheromone for the source of the packet is incremented by a constant value. Each entry in the pheromone table is periodically multiplied by a decay factor. Due to mobility, it may happen frequently that a node does not have a pheromone entry for a destination, thus route request and route reply packets have to be introduced again similar as in AODV [1]. A certain number of route request packets are sent when a node needs to find a path to an unknown destination. The packets perform a random walk and lay down pheromone on the followed trail. The route request packets are forwarded until a node is found which contains some pheromone for the requested destination or the destination itself. This node issues a route reply packet, which is routed back to the originator of the route request. On its way, the route reply packets adds pheromone at the nodes towards its own source.

2.1.3 AntHocNet

AntHocNet[14] is similar to ARA, but additionally introduces a proactive component. The routes are also only set up reactively if needed. A forward ant is broadcasted by the source and finds multiple paths to the destination. A backward ant traveling back to the source establishes the paths towards the destination by updating the entries in the routing tables. After the route setup phase, data packets are then forwarded probabilistically over available links for load balancing. Unlike ARA, AntHocNet periodically transmits also ants during the data session. The ants also follow the pheromone trails but have a small probability to be broadcasted at intermediate nodes. Thus, these ants explore paths around the existing ones and are mainly used to look for path improvements.

ARA, Termite, and AntHocNet, are topology-based protocols where routes are established on-demand. Therefore, they also have the same characteristics of other topology-based protocols such as large control traffic overhead and, thus, are not suited for large networks with highly dynamic topologies as considered in this paper. They mainly make use of the ant colony optimization to improve the resilience and reliability of paths or to improve existing paths compared to other topology-based protocols.

2.2 Position-Based Routing

Basically all position-based routing protocols route packets directly towards the destination and if routing fails, they apply a recovery mechanism. To the best of our knowledge, Terminode routing [15] as discussed below is the only exception where packets are not routed directly towards the destination, but over so called anchor points. As mentioned in the introduction, position-based protocols have numerous advantages and are the preferred choice for routing in large and dynamic networks. Unfortunately, position-based protocols also suffer from several drawbacks, especially in large networks.

- Routing a packet along the line-of-sight between the source and destination may often not be possible in realistic networks. In such scenarios, the performance of position-based routing protocols may degrade severely as greedy routing fails and the recovery mechanism has to be applied. The followed path may then be suboptimal as shown in an example in Fig. 1.
- Each packet is sent completely independently of all others, e.g., if greedy routing fails and packets are forwarded in recovery mode along a very long path even though a much shorter exists, all subsequent packets will follow the longer path. The protocols have no way to adapt and to learn from experiences.
- Packets are routed solely based on location information and other criteria like delay, link capacity, and current traffic load are not taken into account. Even when routing along a straight line to the destination

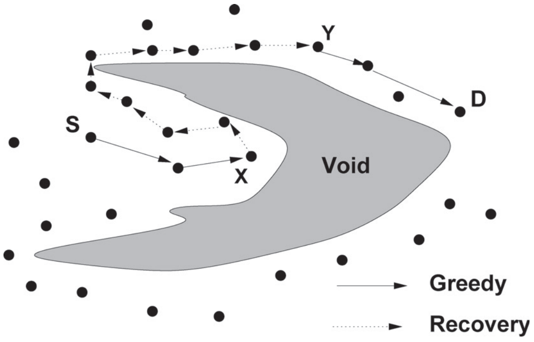

FIGURE 1
Suboptimal path taken by position-based routing protocol.

is possible, it may be advantageous to take other paths to avoid areas with congestion and high delays.

We briefly describe in the following a standard position-based protocol and also the Terminode routing protocol as its objective is similar to AMRA. Like AMRA, it incorporates features to cope efficiently with irregular topology. An overview of other existing position-based protocols can be found, e.g., in [16] and [17].

2.2.1 Greedy and Face Routing

Perhaps the most cited position-based routing protocols is GPSR [18] which is however only an extension of the earlier published GFG [5] with MAC-layer enhancements. Thus, in this paper we refer to these algorithms together as GFG/GPSR. A packet is routed in a greedy mode towards the position of the destination. Each node selects the node among all its neighbors that is geographically closest to the packet's destination. This process is repeated until the packet reaches the destination. If a node does not have any neighbor closer to the destination, the packet is forwarded in recovery mode, i.e., the packet is routed according to the right-hand rule to recover from the local minimum. The right-hand rule is a well-known concept for traversing mazes. To avoid loops, the packet is routed in recovery mode on the faces of a locally extracted planar subgraph, namely the Gabriel graph. As soon as the packet arrives at a node closer to the destination than where it entered recovery mode, the packet switches back to greedy routing, cf. Fig. 1. It was shown that GFG/GPSR guarantees delivery for static and connected networks. If however nodes are mobile, packets may still loop in the network.

The problem of suboptimal routing in recovery mode was already addressed previously. GOAFR as proposed in [6] optimizes GFG/GPSR and avoids routing beyond some radius in recovery mode by branching the graph within an ellipse of exponentially growing size. The objective of GOAFR is however different from that of AMRA. In GOAFR packets are still routed directly towards the destination and if greedy routing fails, it reduces the number of hops that packets are routed in recovery mode. This is unlike AMRA that tries to avoid to route in recovery mode at all by routing over intermediate positions.

2.2.2 Terminode Routing

To the best of our knowledge, Terminode Routing [15] is the only position-based routing protocol that does not always forward packets directly towards the destination, but allows forwarding the packet along an anchored path to circumvent voids. In order to optimize routing in case of voids in the network topology, Terminode routing finds a list of anchor points. Therefore, each node maintains a list of nodes, called friends, to which it maintains a good path. Friends may be distributed all over the network and do not need to be in the vicinity of a node. To find an anchored path to the destination, a node asks its friends that in turn ask their friend and so on. The found anchor points are added as a loose source path route to the header of the data packets. Nodes forward packet to the next anchor point as indicated in the header. When the packet is received at a node close to the anchor point, the entry is removed and the packet routed to the next anchor point. Only if there are no more anchor points in the header, the packet is routed towards the position of the destination. If a packet gets close to the destination, the packet is routed with Terminode local routing. Terminode local routing is used to deal with inaccurate position information of the destination because it may have moved since the last location update. When nodes broadcast periodically hello messages, they not only include their position but also the positions of all known one hop neighbors. Consequently, each node is aware of its two hop neighborhood and a data packet can still be delivered, if the destination node has moved from the destination position as indicated in the packet header.

3 ANTS-BASED MOBILE ROUTING ARCHITECTURE AMRA

In this section, we first give an overview of the whole AMRA architecture consisting of three independent protocols. Then, we describe each of the three protocols separately. Finally, we show how they interact to route packets efficiently to the destination in large scale ad-hoc networks with irregular topologies. We assume that the overall node distribution in the network remains quite static and only varies slowly over time. For most realistic scenarios, this is reasonable assumption as nodes are typically located in towns and on/along streets. We will also study the performance of AMRA if this is not the case and the node distribution changes abruptly.

We model large-scale mobile ad hoc networks as a set of wireless nodes distributed over a two-dimensional area. Nodes are aware of their absolute geographical position by means of GPS and are able to determine other nodes position accurately enough through a location management scheme [8].

3.1 Overview

AMRA is a two-layered framework with three independent protocols rather than a single routing protocol. The two protocols used on the upper layer are called Topology Abstracting Protocol (TAP) and Mobile Ants-Based Routing Protocol (MABR). StPF (Straight Packet Forwarding) is situated on the lower layer and acts as an interface for MABR to the physical network.

TAP is the key to make routing scalable and provides in a transparent manner an aggregated and static topology with fixed “logical routers” and fixed “logical links” to MABR. A logical router represents a fixed geographical area. Thus, mobile nodes within a logical router are situated close together sharing similar routing information and have a similar view on the network topology on a large scale. A logical link represents a path along a straight line to another logical router over possibly multiple physical hops. The actual routing protocol MABR operates on top of this abstract topology and thus does not have to cope with changing topologies. MABR maintains probabilistic routing tables at logical routers and is responsible for determining logical paths on this abstract topology. Data packets are routed based on these probabilistic routing tables between logical routers over logical links. They increase the probability of the followed path depending on the encountered network conditions. Furthermore, ”artificial ants” packets are transmitted periodically to explore new paths. Unlike data packets, these packets are routed purely position-based directly towards their randomly chosen destination. Eventually, the best paths will emerge and MABR is able to circumvent areas with bad or no connectivity, i.e., data packets will always be routed over logical links with high connectivity such that greedy routing is possible. StPF is a position-based routing protocol and then responsible to physically forward packets over the logical link determined by MABR to the next logical router.

3.2 Topology Abstraction TAP

TAP is used to supply in a transparent manner an aggregated and static topology with fixed ”logical routers” and fixed ”logical links”. Logical routers are fixed geographical areas of equal size arranged in a grid to cover the whole area. Unlike in a cellular network where regular hexagons are typically used, we use squares for simplicity reasons. Depending on its current position, each node is part of one specific logical router. A

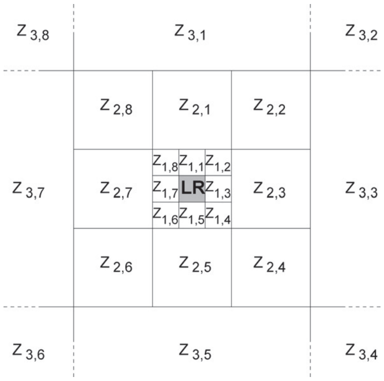

FIGURE 2
Zones relative to logical router in the center.

node can easily detect based on its position, when it crosses the border of the current logical router and then automatically becomes a member of the new logical router. All nodes located within a logical router have the same logical view on the network. Nodes within a logical router corporate in specific routing control tasks such as the emitting of ants. However, each node maintains its own routing table and does never share with or transmit any routing information to its neighbors.

In order to scale to large networks, each logical router groups other logical routers into zones $Z_{i,j}$ as shown in Fig. 2. The zone size increases exponentially with the distance to the center router and allows covering large areas with few zones. The reason is that in the view of a fixed node, close-by nodes that move some distance may be located in an entirely different direction, whereas the same movement of a node far away only marginally affects the direction. It is important to notice that the view of zones is relative. Each router resides in the center of its own zone model. That means that the view of a node changes when it moves to another logical router. A specific fixed geographical position may belong to a certain zone at a given moment and belong to another zone when the node has moved. To simplify addressing, each logical router is identified by the geographical coordinates of its center. This geographical identification simplifies routing with StPF, which uses position information for routing over logical links.

Each logical router maintains a set of outgoing logical links to all its adjacent logical routers. A logical link represents a path along a roughly straight line to a distant logical router over possibly multiple physical hops. TAP is the key to make routing scalable and find good paths on a large scale in the network, i.e., by routing over logical links and not directly towards the destination.

TABLE 1
Routing Table.

	LL_1	LL_2	...	LL_8	μ_d
$Z_{1,1}$	0.9	0.1	...	0	257
$Z_{1,2}$	0.05	0.9	...	0	504
$Z_{1,3}$					
...					
$Z_{2,1}$	0.8	0.1	...	0.1	1045
$Z_{2,2}$	0.125	0.125	...	0.125	0
...					
$Z_{3,1}$	0.15	0.8	...	0	5348
...					

3.3 Mobile Ants-Based Routing MABR

The actual routing protocol MABR operates in the upper layer on top of the abstract topology provided by TAP and thus does not have to cope with frequent changing topologies inevitable in mobile networks. It determines over which logical links, i.e., intermediate positions, packets should be forwarded to circumvent voids in the network topology. These logical links may just lead in the opposite direction of the final destination, e.g., in cases the routing along a line of sight towards the destination is not possible. The important point is that these logical links should be chosen by MABR in such a way that the packet can be routed always greedily to the next logical router along this logical link.

Each node maintains a probabilistic routing table, which depends on its current view on the network, its past locations, and overheard packets. Consequently, routing tables are generally also slightly different for nodes within the same logical router. The zones and the logical links are organized in rows and columns respectively. More precisely, there is a column for the eight unidirectional outgoing logical links and a row for each zone. The entries $P_k^{i,j}$ indicate probabilities to choose the logical link LL_k for destination coordinates in zone $Z_{i,j}$. If location information is provided by GPS, nodes are normally also synchronized. Thus, we can determine the average delay $\mu_d^{i,j}$ of packets received from zone $Z_{i,j}$. The average delay is stored to judge the goodness of the paths taken of incoming packets from the respective zone. If nodes are not synchronized, the average $\mu_d^{i,j}$ can also be calculated based on hop counts of packets or we can sum the encountered delays in the intermediate nodes to estimate the end-to-end delay. At the beginning, all entries are initialized to 0.125 and μ_d to 0. An exemplary routing table is given in Table 1 with the entries $P_k^{i,j}$.

The size of the routing table is in the order of some hundred bytes even for very large networks due to the exponential growing size for farther zones. It requires 9 columns to store the 8 logical links and the average delay and a multiple of eight rows for the zones. Assuming a typical logical router size of $250\,m \times 250\,m$, 11·8 rows would be sufficient to cover the whole globe. A further advantage is that these routing tables never have to be transmitted and only kept in a node's memory.

Packets, data and ants, are marked with their source coordinates, the last visited logical router, and a time stamp. Furthermore, nodes operate in promiscuous mode such that the routing table is also updated for all overheard packets to expedite the dissemination of routing information. When a node receives a packet, it first determines from which zone $\mathrm{Z}_{\mathrm{i,j}}$ the packet originates with respect to its own current view on the network. That means that packets update the routing tables at nodes in the opposite direction that they travel, i.e., towards their sources. The delay d of the packet is used to update the average delay $\mu_d^{i,j}$ for the respective zone in the following way.

$$\mu_d^{i,j} \leftarrow \mu_d^{i,j} + n(d - \mu_d^{i,j})$$

The difference of the delay of the packet and the average delay $d - \mu_d^{i,j}$ is weighted with a factor n. Thus, we have an exponential weighted moving average where the weight of a past sample decreases exponentially fast with n. Based on the goodness of the trip taken of the current packet with respect to the average, a factor r is calculated which affects the amount of pheromone laid down, i.e., by how much the probability is increased.

$$r = \begin{cases} \frac{\mu_d^{i,j}}{3 \cdot d} & : \quad \frac{\mu_d^{i,j}}{3 \cdot d} < r_C \\ r_C & : \quad \text{otherwise} \end{cases} \tag{1}$$

The worse the delay d of the current packet is compared to the average $\mu_d^{i,j}$, the smaller the factor r and the less the probability will be increased. $0 < r_C \leq 1$ is the ceiling of r and limits the amount of pheromones one packet can deposit on a logical link. The last logical link LL_k over which the packet was forwarded is now updated, i.e., the logical link to the previously visited logical router of the packet, with respect to zone $\mathrm{Z}_{\mathrm{i,j}}$. The important idea here is to increase in this way only the probability for logical links along which packets could be routed in greedy mode. $P_k^{i,j}$ is recalculated as follows.

$$P_k^{i,j} \leftarrow P_k^{i,j} + (1 - P_k^{i,j}) \cdot r^2 \tag{2}$$

In this way, small probabilities are increased quicker than already large probabilities for a given r. The probability of the other seven logical links

$l \neq k$ for that zone $Z_{i,j}$ are decreased by

$$P_l^{i,j} \leftarrow P_l^{i,j} - P_l^{i,j} \cdot r^2$$

such that the sum of all logical links in a row to a certain zone remains 1.

Unlike other ants-based routing algorithms, pheromones do not need to decay with time as the overall distribution of the nodes remains the same. A node adapts the pheromone values in its routing table only if it moves to a new logical router in order to reflect the change in its view on the network as follows.

$$P_k^{i,j} \leftarrow P_k^{i,j} + \left(0.125 - P_k^{i,j}\right) \cdot \frac{1}{3^i} \tag{3}$$

All entries asymptotically approach a probability of 0.125. A uniform distribution of 0.125 for all links indicates that links do no longer have pheromone trails and no link is favored over another for a given destination. The factor $\frac{1}{3^i}$ is larger for smaller i, thus, the probabilities $P_k^{i,j}$ for closer zones $Z_{i,j}$ approach 0.125 faster. The dominator is the distance to the zones which increases exponentially with 3^i, i.e the pheromone values decay more rapidly for closer zones. The reason is as already discussed before that if a node moves a fixed distance, a close-by destination may turn out in a completely different direction which requires to decrease the pheromones more quickly than for a distant destination. MABR updates the routing tables identically for ants and data packets, they differ only in their forwarding policy as described below.

3.4 Straight Packet Forwarding StPF

Finally, the physical forwarding process along the logical links selected by MABR is accomplished by StPF, which can be basically any position-based routing protocol. Because many such position-based routing protocols have already been proposed and analyzed in the literature, we did not design a new protocol for StPF. We instead use the perhaps best known position-based protocol GFG/GPSR [5] as StPF, see Section 2.2.1. Basically, any other position-based routing protocol may be applied as well, such as GOAFR [6], BLR [7].

3.5 Routing of Data Packets and Ants

Data packets are routed based on these probabilistic routing tables between logical routers over logical links. The first node within a logical router that receives the packet determines the next logical hop. Therefore, it determines to which zone a packet should be routed from the destination coordinates as given in the packet header. The node then selects the logical link with the highest probability to this zone to forward further the packet. Furthermore, if desired load balancing can be achieved easily by selecting a logical link proportionally among all possible logical links. Thus, data packets are routed logical-hop by logical-hop over the logical links, i.e., from one logical router to one of its adjacent logical routers and so on. If none of the eight logical links for the destination zone has a probability above a threshold $Prob\_Thres$, the data packet is routed purely geographically directly towards the final destination. This is a reasonable heuristic decision if no useful routing information is available. In scenarios with high mobility and where a lot of nodes have not very accurate routing tables because the time is often too short for the best paths to emerge, we encountered that packets may be temporarily routed forward and backward between logical routers. To mitigate this effect, a packet must never be sent back to the last visited logical router or one of its two adjacent logical routers. The packet is routed over the logical link with the highest probability among the remaining five possible logical links. In order to avoid that a packet loops over several logical links, a packet is sent purely position-based if it does not arrive within three times the expected average for this zone.

Ants are used to explore new emerging paths and find shorter paths. Unlike data packets, ants are solely routed by StPF, i.e., they always head directly towards the destination and are only diverted if voids in the routing topology cause them to be routed in recovery mode. In order to control the ant generation rate, nodes cooperate in emitting ants and detecting new paths. A node only emits an ant packet, when it did not detect an ant with source coordinates within its current logical router for $t_{EmitAnts}$, i.e., each logical routers generates an ant every $t_{EmitAnts}$. In cities with a high node density and a large number of nodes in a logical router area, a node only rarely has to transmit an ant packet. The destination of the ant is chosen uniformly randomly over the whole simulation area. Ants are either routed according to the left or right-hand rule in recovery mode if they reach a dead end and greedy routing fails. The reason for routing ants with both rules is that ants otherwise could miss shortest paths.

3.6 Example of Routing with AMRA

Until now, we only discussed the protocols separately and only roughly described their interaction. In this section, we study how they corporate by means of an example depicted in Fig. 3. The zones of the network are sketched in the view of node *S*. Nodes in $Z_{3,3}$ have previously transmitted ants, or also data packets, which either were destined for the logical router in which *S* is located or just pass through *S* for a more distant destination. Exemplarily, the path taken of ants emitted by *D* are shown. Ants routed by the right-hand and left-hand rule enter the logical router of *S* over the logical links LL_1 and LL_6, respectively. *S* determines that these ants origin from a source in zone $Z_{3,3}$ according to its current view. The delay of ants arriving over LL_1 is much shorter than of ants LL_6 due to the shorter

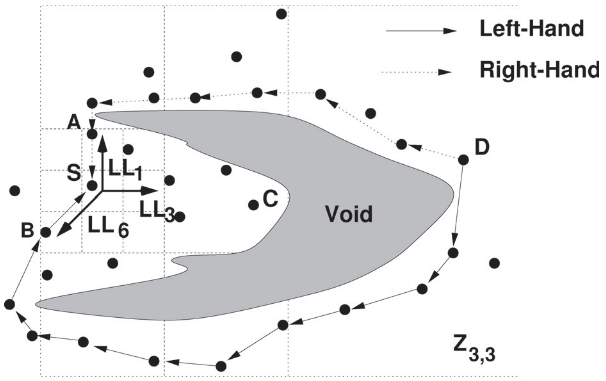

FIGURE 3
Ants laying trails over LL_1 and LL_6 to $Z_{3,3}$.

path. Consequently, the probability of LL_1 is higher than of LL_6 for zone $Z_{3,3}$ in the routing table at node *S*. Furthermore, as no packets enter over one of the other six logical links, their probability is close to zero. Not only node *S* but all intermediate nodes which forward or overhear the ants update their routing table according to their view on the network.

When *S* now has to send packets to any node located in zone $Z_{3,3}$, it routes the packets over LL_1, i.e., to node *A* in Fig. 3. Thereby, it is irrelevant whether *S* is the source of the packet or just any intermediate node which forwards the packet. *A* in turn forwards the packets then over one of its logical links with the strongest pheromone and the packets travel the indicated path to *D*. *S* does not forward packets for *D* towards *C* over LL_3. The probability of LL_3 is close to zero because no packets originating from zone $Z_{3,3}$ have been received from this direction. In this way, AMRA avoids to forward packets into directions which have no, or only a very small, probability for the packets to arrive directly. On the other hand, if a purely position-based routing protocol is used such as GFG/GPSR, packets for *D* are routed first towards *C* where greedy routing fails as no neighbor is located closer to *D* as already depicted in Fig. 1. The hop count is not only increased by the longer path but also through a property of the recovery mode, more precisely by the Gabriel Graph, which yields shorter links than potentially available. As GFG/GPSR is memoryless, subsequent packets are routed exactly over the same longer suboptimal paths. Even if *S* did not have any useful entry in its routing table, AMRA would simply forward the packet without any delay and the next node which has an entry for the destination zone forwards the packet along the best logical links. The probability that nodes do not have entries is very small because very distant zones are proportional large to ensure that packets from this zone are overheard.

TABLE 2
Parameters.

Parameter	Value
Simulation Time	$900\,s$
Traffic Start	$120\,s$
Traffic End	$880\,s$
Traffic Type	Constant Bit Rate
Traffic Rate	4 Packets/s
Packet Size	64 Byte
MAC Protocol	802.11
Bandwidth	2 Mbps
Transmission Range	$250\,m$
Confidence Interval	95%

4 EVALUATION

We implemented and simulated AMRA in the Qualnet network simulator and compare its performance with two other position-based protocols, namely GFG/GPSR and Terminode routing (The code of the Terminode routing protocol is a courtesy of the authors of the Terminodes protocol [15].) We first describe the general simulation scenario. We conducted several simulations with large and irregular network over a wide range of conditions and present the obtained results afterwards. Simulation parameter are set to the values as given in Table 2, if not noted otherwise. The parameters of AMRA are set as follows. The logical router size was set to twice the transmission range, i.e $250\,m \times 250\,m$. In this way, nodes within the same logical router overhear the same packets and thus have similar routing tables. Furthermore, we set the probability threshold for a link $Prob\_Thres$ to 0.2. The time $t_{EmitAnts}$ after which a node creates an ant, when it did not detect that an ant was emitted from any node within its current logical, was set to $5\,s$. The parameters for the pheromone laying function are set to $r_C = 0.8\,s$ and $n = 10$, i.e., approximately the last ten samples contribute to the average $\mu_d^{i,j}$.

4.1 Large Network with Irregular Topology

To simulate realistically large networks with irregular topologies, we use the restricted random waypoint mobility model, which was first introduced in [15]. The model defines rectangular city areas and highways connecting cities. In reality, most people often move within relatively small geographical areas and only rarely travel long distances to other cities. On the other hand, some people may travel frequently also over long distances between cities such as commuters, express agents, and truck drivers. The restricted random

TABLE 3
Parameters for large networks.

Parameter	Value
Number of Nodes	500 (Commuters: 300 Ordinary nodes:200)
Prob to change city	Commuters: 0.9 Ordinary nodes: 0.1
Size of Area	$3000\,m$ x $2500\,m$
Number of Cities	4
Size of Cities	$1000\,m$ x 1000
Number of Highways	3
Min Speed in City	$1\,m/s$
Max Speed in City	$15\,m/s$
Pause Time	Commuters: $1\,s$ Ordinary nodes: $30\,s$
Min Speed on Highway	$10\,m/s$
Max Speed on Highway	$30\,m/s$

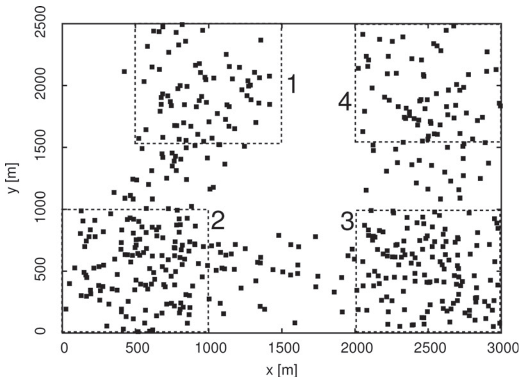

FIGURE 4
Snapshot of restricted random waypoint mobility model.

waypoint mobility model tries to capture this behavior by introducing two kinds of nodes that correspond to the rarely and frequently traveling people called commuters and ordinary nodes respectively. These two kinds of nodes differ in their frequency to move to another city and their pause time between the trips. Node movement within a city is according to the standard random waypoint mobility model and defined by a minimum and maximum speed and a pause time. Nodes move to one of the cities connected via a highway with a certain probability. The node speed on the highway is higher than for trips within the same city. The parameter values used for this restricted mobility model are given in Table 3. In Fig. 4, we see a horseshoe-like topology with the 4 cities and the three highways connecting cities 1 and 2, 2 and 3, 3 and 4 and a typical node distribution after some simulation time.

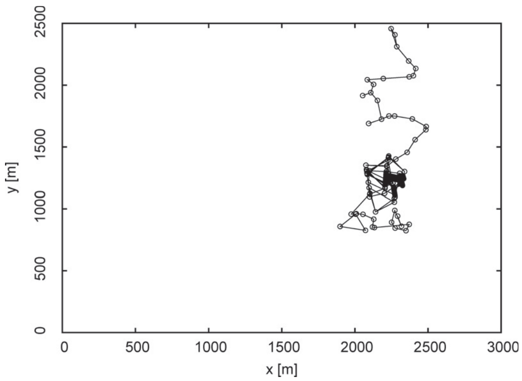

FIGURE 5
Looping packet on the highway.

4.1.1 Simulation with Mobility

The results of the simulations with mobility were found to be very disappointing. Even when varying, the number of ants, the size of the logical router, etc. AMRA was only able to deliver around 10% of the packets. The reason for the poor performance was found when we simulated GFG/GPSR in the same scenarios. Surprisingly, also GFG/GPSR achieved only a delivery ratio of around 15%. The same experiments without mobility yielded a delivery ratio of almost 100% as expected because GFG/GPSR guarantees delivery in static networks. Further analysis revealed that the reason is the high mobility on the highway that causes packets to loop frequently. Nodes keep track of their neighbor positions obtained by hello messages. If mobility is high, these stored positions do not correspond to the actual position and wrong forwarding decisions are taken. Some packets may recover from the loop, others do not. Consequently, the queues of the nodes get filled up and start to drop packets. Furthermore, packets are also dropped because the TTL-field expires if they are caught in a loop for a while. An exemplary path of a packet routed by GFG/GPSR in one of the simulations is shown in Fig. 5.

Therefore, it is not surprising that AMRA performs poorly as it uses GFG/GPSR to physically forward packets. Terminode routing suffers from the same problem as packets are also routed by GFG/GPSR between the anchor positions. The reason why in [15] Terminode was able to deliver much more packets in a similar scenario is because first there are stationary nodes distributed all over the area and second the minimum speed for nodes on the highway was set to $1\,m/s$. Thus, there were always slow moving or stationary nodes on the highway which build a backbone for routing.

Due to these observations, we conclude that position-based routing protocols which rely on neighbor information for forwarding are not able to operate in highly dynamic networks. The problem may also exist in

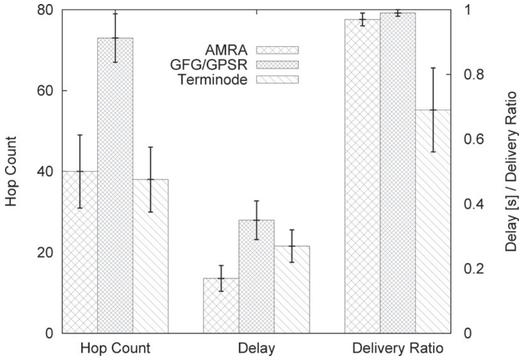

FIGURE 6
Results in a large static network.

networks with frequently changing topology due to sleep cycles of nodes such as in sensor networks. The outdated neighbor information leads to many wrong routing decisions. In such scenarios, position-based protocols which route packets without having knowledge about the neighborhood and do not require transmission of hello messages may be more appropriate. They were shown to be almost unaffected even by very high mobility, e.g., cf. [7].

4.1.2 Simulation without Mobility

As it was not possible to obtain meaningful simulation results with mobility, we evaluated the performance of the protocols in static irregular networks. Nodes only move at the beginning in order to obtain a typical node distribution. After a certain time, nodes are stopped and remain static during the rest of the simulation. It is uninteresting to have source and destination nodes close to each other during the data transmission, which may happen if they are randomly chosen. Therefore, we choose the source and destination nodes from city 1 and 4 only. The results of these simulations are given in Fig. 6.

The hop counts for AMRA and Terminode routing are in the order of 40 whereas GFG/GPSR required almost 80 hops. Unlike GFG/GPSR, AMRA and Terminodes routing forward packets not directly towards the other city but along the intermediate positions which indicate the path over cities 2 and 3 instead. Along this path, packets can be routed most of the time in greedy mode.

As they are able to deliver data packets with significant less hops, the delay is accordingly also shorter. The delay of AMRA is approximately $0.18\,s$ and is about half of GFG/GPSR $0.35\,s$. The reason for the longer delay of $0.28\,s$ for Terminode is the large size of the headers as already mentioned before. The Terminode local routing requires adding all known

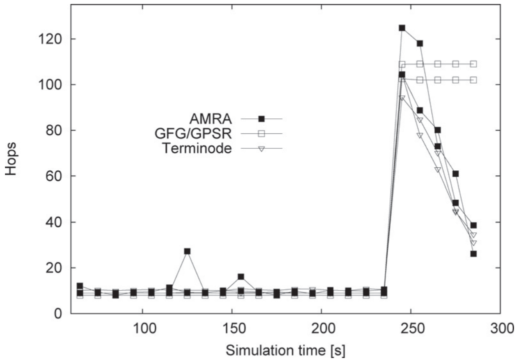

FIGURE 7
Reaction to removal of a highway.

positions of neighbors in the periodically transmitted hello messages by each node. The hello messages can easily grow to a size of several hundred bytes in dense networks leading to network congestion.

Terminode routing performs poorly and only delivers 70% of the packets. This is again due to increasing control traffic which congests links. GFG/GPSR and AMRA on the other hand are able to deliver almost 100% of the packets. Only few packets drops were observed due to collisions with hello messages and other data packets. The reason that AMRA has a slightly lower delivery ratio is because of the probabilistic pheromone trails. These may cause packets to loop sometimes over several logical routers and packets are dropped because of an expired TTL-field.

4.1.3 Radical Topology Changes

Until now, we assumed that the overall node distribution in the network is approximately constant even though individual nodes may be highly mobile. Considering a realistic scenario, it may happen that the distribution also changes, e.g., there are only few cars on a highway during the night, but there is a very high node density in the morning during the rush hour. Therefore, we also want to assess the ability of the investigated protocols to adapt to radical topology changes. We evaluated two scenarios where a highway was inserted and removed respectively between city 1 and 4 after 240 seconds of simulation time. The result for the highway removal scenario are depicted in Fig. 7 where each graph refers to a specific simulation run. The hop count sharply increases for GFG/GPSR to more than 100 hops and remains high as the protocol has no way to adapt to the new situation and learn shorter paths. The hop count of AMRA and Terminode routing also increases for a short time to approximately the same values as GFG/GPSR . However within few ten seconds of simulation time, the values decrease again to approximately 30 hops when AMRA learned the

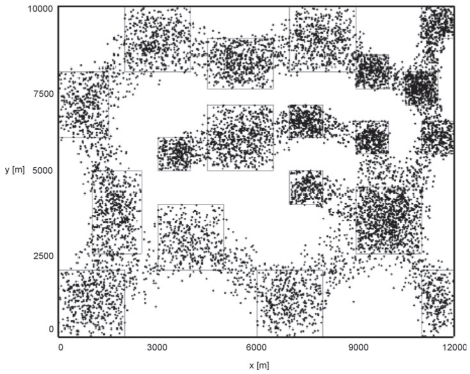

FIGURE 8
Complex network with 10000 nodes and 19 cities.

new topology and pheromone trails were established between cities 1 and 4 over cities 2 and 3. The graph for the highway insertion scenario is not shown. GFG/GPSR routed packets immediately over the shorter path because all packets are routed directly towards the destination. AMRA and Terminodes were also able to adapt very quickly within a few seconds to the new topology.

4.2 Very Large Network

Unfortunately, we were not able to conduct simulations, not even static, with thousand of node and a highly complex network topology with Qualnet. Thus, we also implemented and simulated AMRA in a simple Java simulator which allows such simulations also under mobility. The Java-simulator implements the same functionality as Qualnet such as CBR traffic and the restricted random waypoint mobility. However, it does not account for any physical propagation medium properties or MAC layer functionality. Therefore, packets cannot be dropped due to collisions or congestion and packets do not experience delay. In order to asses the performance of AMRA, the hop count was used. We belief that a hop count metric is a good representative for the delay as CSMA based MAC protocols such as IEEE 802.11 have high cost for acquiring the medium. The Terminode routing protocol is a highly complex protocol. Therefore, we did not implement Terminode routing but used additionally to GFG/GPSR a shortest path algorithm for comparison and to assess the goodness of paths chosen by AMRA.

We simulated 10000 mobile nodes in an area of $10000\,m \times 12000\,m$ with 19 cities and 19 interconnecting highways. The nodes move again according to the restricted random waypoint mobility model with the same parameters as before except that the size of the cities varies between $1000\,m \times 1000\,m$ and $2000\,m \times 2000\,m$. A snapshot of this network is depicted in Fig. 8.

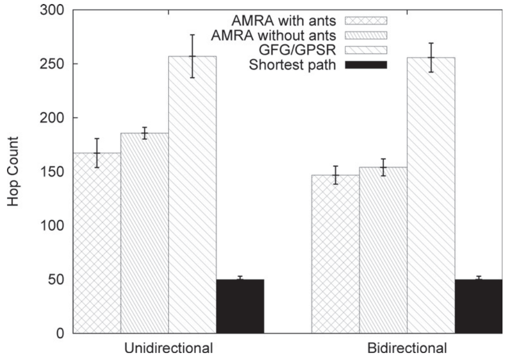

FIGURE 9
Hop count in a very large network.

We are aware that this scenario may be unrealistic. However, there may exist similar network topologies in other situations, e.g., a sensor networks distributed over a large area may have a similar highly irregular topology.

There are 200 traffic sources in this simulation. We wanted to study the effect of a varying number of ants and also if traffic is sent unidirectional and bidirectional. It is common for a lot of applications to have bidirectional traffic flows or sometimes simply because TCP is used as Transport protocol. Thus, for the bidirectional simulations, we had 100 pairs of nodes and nodes transmit data to their respective peer. Once AMRA was simulated with no additional ants transmitted to explorer shorter path and once with 500 ants transmitted per second in the whole network. In Fig. 9, we see the average hop count of AMRA, GFG/GPSR, and the shortest path algorithm. Obviously, the hop count of GFG/GPSR and the shortest path algorithm does not change for uni- and bidirectional traffic as the traffic in the opposite direction has no influence. On the other hand, AMRA benefits from bidirectional traffic and the hop count drops about 10% compared to pure unicast transmissions, independent whether ants are transmitted or not. The reason is that packets update the routing tables in the opposite direction of their trip, i.e., towards their source. Thus, if data packets flow in both directions, they can help each other directly to find shorter paths. The ratio of AMRA to GFG/GPSR is approximately 250 to 150 hops in this scenario. Furthermore, we can observe that AMRA without ants performs only marginally worse than with ants. The 200 sources generate already enough traffic in the network such that node can maintain useful path information on the large scale to distant areas. Source and destination nodes may be temporarily close to each other such that greedy routing of GFG/GPSR is able to deliver packets. During this time, GFG/GPSR will perform equal or even slightly better than AMRA. Considering this fact, one might expect a even higher performance gain of AMRA over GFG/GPSR when source and destination nodes are far apart such that simple greedy routing will not succeed.

5 CONCLUSIONS

In this paper, we proposed the AMRA architecture which makes use of topology abstraction, ant colony optimization, and position-based routing. Unlike conventional position-based routing protocols such as GFG/GPSR, AMRA does not route packets greedily towards the destination if not possible but over intermediate positions which yields more optimal paths as routing in recovery mode can be avoided. The actual forwarding of the packets between these intermediate positions is still accomplished by a position-based routing protocol such that AMRA retains their advantages such as low control traffic and statelessness. The paths over the intermediate positions are determined by data packets reinforcing their traveled path by making use of a paradigm from ant colony optimization. AMRA shows equal performance as GFG/GPSR in simple networks (not shown in this paper). Designed for large and irregular topologies, AMRA performs superior in such scenarios and routes packets over paths that are 50% and more shorter than of GFG/GPSR. This performance comes at a certain cost, but which can be kept reasonable small. Compared to conventional position-based routing, AMRA has to maintain a small routing table at nodes limited to some hundred bytes only and few additional control packets are transmitted to detect new paths. If enough data traffic is transmitted, no control packets are required at all. The AMRA architecture is designed such that the individual protocols could be replaced with relative small costs. Instead of using a MABR based on ant colony optimization, DSR [2] could also be used to find paths on the abstract topology. Similarly, the more sophisticated GOAFR [6] protocol could be used instead of GFG/GPSR for the physical forwarding. The gain may be however limited as GOAFR is superior to GFG/GPSR for scenarios where packets are routed frequently in recovery mode, what is exactly avoided by the use of AMRA. An further important finding of this paper is that GFG/GPSR and, thus, all position-based routing protocols which require neighbor knowledge and transmit hello messages, are not able to operate in highly dynamic scenarios where neighbor information is outdated quickly. Thus, this is an indication that GFG/GPSR may also not be appropriate for sensor networks where network topology changes frequently due to sleep cycles of nodes.

ACKNOWLEDGMENT

Many thanks go to Dr. Ljubica Blažević of EPF Lausanne for the code of the Terminode routing protocol. Many thanks go also to Aaron Zollinger and Regina O'Dell of ETH Zürich who kindly provided the code that served as the basis for Java simulator used in this thesis.

REFERENCES

[1] Perkins, C. E. and Royer, E. (2003). "Ad hoc on-demand distance vector (AODV) routing," RFC 3561, Internet Engineering Task Force IETF. [Online]. Available: http://www.ietf.org/rfc/rfc3561.txt

[2] Johnson, D. B., Maltz, D. A. and Broch, J. (2001). "DSR: The dynamic source routing protocol for multihop wireless ad hoc networks," in *Ad Hoc Networking*. Addison-Wesley, ch. 5, 139–172.

[3] Clausen, T. and Jacquet, P. (2003). "Optimized link state routing protocol (OLSR)," RFC 3626, Internet Engineering Task Force IETF. [Online]. Available: http://www.ietf.org/rfc/rfc3626.txt

[4] Ogier, R., Templin, F. and Lewis, M. (2004). "Topology dissemination based on reverse-path forwarding (TBRPF)," RFC 3684, Internet Engineering Task Force IETF. [Online]. Available: http://www.ietf.org/rfc/rfc3684.txt

[5] Bose, P., Morin, P., Stojmenovic, I. and Urrutia, J. (1999). "Routing with guaranteed delivery in ad hoc wireless networks," in *Proceedings of the 3th International ACM Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM '99)*, Seattle, USA, 48– 55.

[6] Kuhn, F., Wattenhofer, R. and Zollinger, A. (2003) "Worst-case optimal and average-case efficient geometric ad-hoc routing," in *Proceedings of the 4th ACM International Symposium on Mobile and Ad Hoc Networking and Computing (MobiHoc '03)*, Annapolis, Maryland, USA, 267–278.

[7] Heissenbüttel, M., Braun, T., Bernoulli, T. and Wälchli, M. (2004). "BLR: Beacon-less routing algorithm for mobile ad-hoc networks," *Elsevier's Computer Communications Journal (Special Issue)*, *27*, 11, 1076–1086.

[8] Li, J., Jannotti, J., De Couto, D. S. J., Karger, D. R. and Morris, R. (2000). "A scalable location service for geographic ad-hoc routing," in *Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM '00)*, Boston, USA, 120–130.

[9] Di Caro, G. and Dorigo, M. (1998). "AntNet: Distributed stigmergetic control for communications networks," *Journal of Artificial Intelligence Research*, *9*, 317–365.

[10] Schoonderwoerd, R., Holland, O., Bruten, J. and Rothkrantz, L. (1997). "Ant-based load balancing in telecommunications networks," *Adaptive Behavior*, *5*, 2, 169–207.

[11] Heusse, M., Snyers, D., Guerin, S. and Kuntz, P. (1998). "Adaptive agent-driven routing and load balancing in communication networks," *Advances in Complex Systems*, *1*, *2* & *3*, 237–254.

[12] Güneş, M., Kähmer, M. and Bouazizi, I. (2003). "Ant-routing-algorithm (ARA) for mobile multi-hop ad-hoc networks - new features and results," in *Proceedings of the 2nd Mediterranean Workshop on Ad-Hoc Networks (Med-Hoc-Net'2003)*, Mahdia, Tunesia.

[13] Roth, M. and Wicker, S. (2003). "Termite: Ad-hoc networking with stigmergy," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'03)*, San Francisco, USA.

[14] Di Caro, G., Ducatelle, F. and Gambardella, L. (2005). "AntHocNet: An adaptive nature-inspired algorithm for routing in mobile ad hoc networks," *European Transactions on Telecommunications, Special Issue on Self-Organization in Mobile Networking*, *16*, 2.

[15] Blazevic, L., Giordano, S. and Boudec, J. Y. L. (2001). "Self organized terminode routing simulation," in *Proceedings of ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile systems (MSWiM 2001)*, Rome, Italy.

[16] Mauve, M., Widmer, J. and Hartenstein, H. (2001). "A survey on position-based routing in mobile ad-hoc networks," *IEEE Network*, *15*, 6, 30–39.

[17] Stojmenovic, I. (2002). "Position based routing in ad hoc networks," *IEEE Communications Magazine*, *40*, 7, 128–134.

[18] Karp, B. and Kung, H. T. (2000). "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM '00)*, Boston, USA, 243–254.