

Towards the Design of an Energy-efficient, Location-aware Routing Protocol for Mobile, Ad-hoc Sensor Networks

ARIS PAPADOPOULOS
ap7@doc.ic.ac.uk

JULIE A. McCANN
jamm@doc.ic.ac.uk

Department of Computing
Imperial College London
180 Queen's Gate, London SW7 2AZ

Abstract. Developments in wireless, mobile communications combined with advancements in electronics have contributed to the emergence of a new class of networks: Wireless ad-hoc sensor networks. Tiny, smart, network-enabled sensing nodes can be deployed to construct sensor fields that form the infrastructure for various self-adaptive and autonomic applications. In this paper we identify the requirements and properties that still need to be addressed and discuss possible approaches that could be adopted in the design of efficient routing protocols for such networks.

1. Introduction

Recent developments in wireless, mobile communications combined with the constant advancements in electronics that enable the integration of complex components into smaller devices, have contributed to the emergence of a new class of wireless ad-hoc networks: Sensor networks. Typically a sensor board consists of a number of sensors of different modalities which, when combined with a microprocessor and a low-power radio transceiver, forms a smart network-enabled node. The on-board sensors may be motion detectors, thermistors, light sensors, microphones, accelerometers, magnetometers, humidity and barometric pressure sensors, GPS receivers etc. A sensor network may deploy a huge number of nodes depending on the nature of the application. Such applications include medical services, battlefield operations, crisis response, disaster relief, environmental monitoring, premises surveillance, robotics and more.

Sensor networks are also inherent in the concepts of smart dust [1] and ubiquitous computing [2]. Smart dust technology concerns the design and implementation of networks consisting of tiny, invisible sensing grains that aim to be untraceable in practice. Currently, smart dust motes scale down to 1mm^2 and consist of microfabricated sensors, a photodetector, passive and active optical transmitters, a signal processing unit, a solar cell and a thick-film battery. On the other hand, ubiquitous computing concerns the building of intelligent environments. By placing a processor behind virtually every object, the computers are drawn out of their racks to be seamlessly integrated with the physical environment and form a ubiquitous infrastructure that will monitor and/or support every human activity from the simplest to the most complex one.

In many ubiquitous computing applications there is no fixed, backbone infrastructure to support the nodes and therefore the network must be self-adaptive and autonomous, in essence *autonomic*. In the majority of the protocols presented in this paper, a mobile, ad-hoc, wireless network consisting of homogeneous nodes of equal capabilities is assumed¹. In this sense, a sensor network has obvious similarities with a traditional ad-hoc network. For the rest of this survey we will use the abbreviation MANETs when referring to traditional mobile, ad-hoc networks, WSNs to denote mobile, ad-hoc, sensor networks and the term *mote* to specifically refer to a WSN's node. However there are some vital differences between WSNs and MANETS, which are outlined as follows:

- Energy efficiency/longevity: Sensor nodes power capacity is restricted because sensor motes may be smaller, battery powered and be required to stay alive for longer periods without any support.
- Scalability: Sensor networks are typically denser and require a larger number of nodes. Some projects claim to attempt to deploy billions of devices, including passive ones.

- Mobility: Motes are characterised by higher mobility and topology changes are more frequent than in MANET nodes.
- Fault-tolerance: Although MANETs are designed to be fault tolerant, extra care should be taken regarding sensor networks. This is because the latter is expected to be able to function even after a large number of node failures, which could be a result of their limited power capability and extended life specification.
- Identification: In contrast to MANETs, schemes that make no use of unique IDs are preferred due to the large number of nodes and the applications, which typically require data multicasting rather than end-to-end communication, therefore avoiding IP usage.
- Cross-layer design: Application-level decisions directly influence the design of the substrate layers. For example, different routing protocols may be required according to whether the application is demand or event-driven.

We believe that most of the current protocols designed specifically for WSNs fail to exploit one of the most interesting and distinguishing characteristics of the networks' nature: the ability of the nodes to have some knowledge over their own and their neighbours' positions. When used appropriately, this feature can save energy, which is the key issue for the design of efficient protocols. Although several such methods are proposed for routing messages over MANETs, more work is required towards a protocol that can serve the WSNs needs as presented and distinguished previously and, at the same time exploit location awareness to a satisfactory degree.

The rest of this paper is organised as follows: Section 2 briefly presents a WSN architecture and application design issues that are directly linked to the underlying network-layer. In section 3, we classify and introduce some basic MANETs routing techniques that could prove to be of help towards the design of a WSNs routing protocol. Section 4 presents algorithms specifically designed for routing over WSNs and in section 5 we briefly overview algorithms proposed for energy consumption minimisation over MANETs. In section 6 we consider some location and position aware routing for traditional ad-hoc networks that could form the basis for a new energy-efficient, scalable, location-aware protocol for WSNs. We conclude the paper in section 7 extracting the final conclusions and proposing specific research directions.

2. Sensor Networks Architecture and Design Aims

A sensor field consists of up to several thousands of densely deployed, networked, mobile, sensing nodes. A distinguished node, usually referred in the literature as the *sink*, is responsible for gathering the data collected by the other nodes and forwarding it to the external, fixed infrastructure for further processing or forwarding. The sink may be no different from any other node of the WSN and therefore carry all relevant restrictions, or also may be fixed.

According to the most prominent power attenuation model [3,4] when a node s transmits to a node r with power P_s , the power at the point where r lies will be:

$$P_r = \frac{P_s}{\|s, r\|^k}$$

where $\|s, r\|$ is the Euclidean distance between the source and the receiving node, and k is the *distance power gradient*. In the real world, it holds that $2 \leq k \leq 6$ according to the topology of the space, but for an even, flat surface it stands that $k \approx 2$.

As energy consumption is proportional to the square distance between the communicating nodes for the two dimensional Euclidean space, multi-hop forwarding is preferred over direct transmission (like

¹ When this does not apply, it will be stated explicitly.

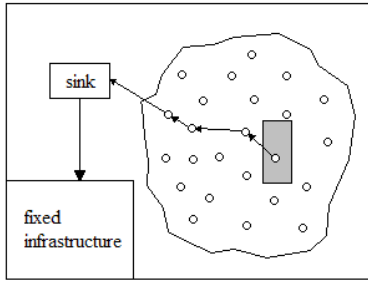


Figure 1: Multi-hop routing from the area of interest (shaded rectangle) to the sink in a sensors field

driven or hybrid. In the first case, data that are collected by the sensors, flow continuously towards the sink. In an event-driven application, data are collected and sent to the sink when an event of interest occurs. In the case of a demand-driven application, data are sent to the sink as a response to an explicit request that is pushed into the network in the form of a query. Finally a hybrid model can combine a number or all of the above methods. At this point, the routing protocol design is inevitably linked to the application level. A simplified overview of a sensor field is provided in Figure 1, in which one can see the sink, area of interest and multi-hop data dissemination.

Furthermore, a routing protocol must take into consideration that the application's nature is not end-to-end, in contrast with most MANET applications. For example a number of nodes may be able to gather information from the area of interest. In the case of a demand-driven application this means that more than one node must receive the query. The challenge in this case is that flooding the network with the query messages results in unwanted energy waste, and therefore it should not be considered as a solution. Instead there should be a way to target the request to specific nodes that could have the potential to return relevant information. To this end, location-aided routing i.e. directing the packets appropriately, using exact position or approximate location information, seems to be a feasible approach. This is because such algorithms tend to choose the shortest path to the destination and avoid flooding, which results in low control packet overhead and energy savings.

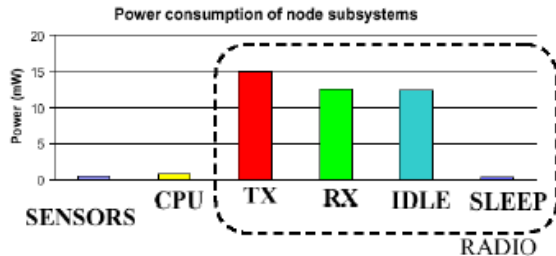


Figure 2: Energy consumption during sensing, processing and executing the various communication operations as reported in [5]

Probably the most critical design consideration of such an application is how processing is distributed over the WSN. As transmitting 1 bit of data proves to be much more expensive than processing it [6] (see figure 2), computation should be pushed inside the network in order to minimise transmissions. The question that arises is how this is possible and what does processing refer to when discussing sensor networks. First, it would be possible that a receiving node could send raw data to the sink, leaving the processing and analysis of the data to the fixed infrastructure. Instead, processing the signals on-board and responding in only with relevant to the query information is much more efficient as it saves energy from transmitting uninteresting data. In this case, sensor nodes are required to run signal-processing algorithms in order to recognise event and source types. This in fact has little to do with information routing. However data concerning a single event that are sensed by multiple nodes could be aggregated inside the network so that the information that will be finally routed back to the sink will be more accurate and the number of transmissions will be decreased. A WSN routing protocol should consider the data aggregation capabilities of the nodes.

traditional MANETs). Therefore data is routed back to the sink through a series of links between neighbouring nodes that may have no knowledge of the future or even current topology of the network due to its vast number of nodes and their high mobility.

A WSN application may be continuous, event-driven, demand-

According to the previous discussion a routing protocol for a demand-driven sensor application should include the following three phases:

1. Query forwarding to the appropriate nodes
2. Data aggregation
3. Data routing back to the sink

To the best of our knowledge no algorithm has been proposed adopting this holistic approach and incorporating methods for all the above stages.

3. Routing Protocols for Traditional MANETs

This section briefly reviews the main routing algorithms that have been proposed for traditional MANETs in order to clarify why they are not appropriate for WSNs.

3.1 Table-Driven Routing

Table-driven routing is also called *proactive* and *precomputed* routing. Algorithms that fall into this category are Destination-Sequenced Distance-Vector Routing (DSDV) [7], Clusterhead Gateway Switch Routing (CGSR) [8] and Wireless Routing Protocol (WRP) [9]. They all store routing tables at every node, recording the paths to all other nodes of the network and use periodic broadcasts to keep routing information up to date.

These protocols are not applicable to WSNs for a number of reasons. The most important is that they generate a great number of routing messages traffic in order to keep all nodes informed of all possible routes to all possible destinations. This traffic is generally too heavy for energy-limited motes. Considering in addition the high mobility and large scale of the WSNs, it is possible that, even if the energy restrictions were neglected, the paths to the distant motes would almost never be valid as they would change very fast.

3.2 On-Demand Routing

On-demand or otherwise *reactive* routing does not maintain routing information for all possible destinations at all nodes. Instead, it performs a *route discovery* procedure only when the source needs to send a message to a destination. The main representatives of this class are Dynamic Source Routing (DSR) [10], Ad-hoc On-Demand Distance Vector Routing (AODV) [11], Temporally Ordered Routing Algorithm (TORA) [12], Associativity Based Routing (ABR) [13] and Signal Stability-Based Adaptive Routing (SSA) [14]. Although on-demand routing protocols are a better approach to WSN routing optimisation because the routing traffic they create is much lighter and the storage space they require considerably smaller, they still suffer important disadvantages. For example, DSR assumes networks, which are small in diameter and have moderate mobility. A first observation is that the inherent route request phase inevitably enforces a delay in the actual packet transmission. In addition, in a highly mobile environment where links are established and disabled rapidly, messages may be lost on their way to the destination. Finally the dynamic nature of WSNs can locally create heavy traffic of bursty nature resulting in waste of energy and bandwidth.

4. Routing Protocols for WSNs

In response to the new design needs imposed, a new class of protocols, developed specifically for WSNs has appeared. The main representatives are Directed Diffusion [15], Low-Energy Adaptive Cluster Hierarchy (LEACH) [16], Sensor Protocols for Information via Negotiation (SPIN) [17], Geographical and Energy Aware Routing (GEAR) [18], Energy-Aware Data-Centric Routing (EAD) [19], SPEED [20] and Berkeley's Energy-Aware Routing [21]. We review them below.

4.1 Directed Diffusion

Directed Diffusion is certainly one of the most interesting protocols. Data is defined as pairs of attributes and values. A query is expressed by the sink as four such pairs: the type of the event of interest (e.g. detection of a human), the rate at which the sink requests to receive data (e.g. every 100ms), the duration of time which the sink will be interested in listening for data (e.g. 10s) and the area of interest in which the target event happens (e.g. a rectangle of specified GPS coordinates). Such a query message is named an *interest*.

The sink initiates the protocol by flooding the network with periodic broadcasts of an interest with a high interval value, i.e. requesting data at low rate. An *interest cache* is maintained on each mote. The cache stores the received interest itself, a timestamp, the duration and several *gradients* up to the number of the neighbours. The gradients point to the neighbour from which the interest is received and the interval value requested. Upon receiving an interest, a node updates its interest cache and decides on whether to forward it to its neighbours based on if it has seen the same interest recently.

The Directed Diffusion protocol is data-centric and highly adaptive as it selects empirically low delay paths based on local interactions. This also implies that the non-end-to-end approach is adopted and that there is no need for global IDs throughout the network. Simulations comparing the protocol to flooding and omniscient multicast on the basis of the *average dissipated energy* and *average delay* metrics showed substantial improvements. However, while mobility is indirectly addressed by the second metric, scalability remains an issue in that performance simulations involved up to 250 nodes only.

4.2 Low-Energy Adaptive Cluster Hierarchy (LEACH)

LEACH organises the network into clusters. Each node decides on whether to become a clusterhead according to a certain probability that is specified a priori. Clusterheads process tasks requiring higher energy. LEACH addresses this by adopting a mechanism of changing clusterheads after specified intervals of time in order to spread battery usage fairly over the network. After clusterheads are elected, nodes select the cluster in which to affiliate by choosing the clusterhead with which they need the minimum energy to communicate.

LEACH differs from the rest of the protocols described here as it adopts direct instead of multi-hop transmission. However, its weaknesses are quite obvious. The authors have made a number of assumptions, which do not match typical WSN architectures and requirements. For example they show in some cases that direct transmission becomes more efficient than multi-hop. Adopting their idea and using its formalisation one can easily see that multi-hop routing is more efficient for a network of 10m diameter, consisting of just 10 nodes. Again there is no evidence that this protocol would scale for networks larger than 100 nodes and the authors neglect the fact that for larger diameters the base station may be out of the clusterheads transmission range. Furthermore the fact that the percentage of clusterheads must be determined a priori requires some knowledge of the network from beforehand, diminishing the self-adaptive property of the protocol.

4.3 Sensor Protocols for Information via Negotiation (SPIN)

SPIN is designed to address three deficiencies of flooding: *Implosion*, *overlap* and *resource blindness*. Implosion refers to the waste of resources taking place when a node forwards a message to a neighbour although the latter may have already received it from another source. Overlap occurs when two nodes sense the same region and produce and push into the network the same results. Resource blindness denotes the incapability of the protocol to adapt the node's behaviour according to its power status.

There are two versions of the protocol. SPIN-1 is a simple 3-stage handshake protocol. A node advertises its data (DATA) by broadcasting meta-data (ADV), which may include node-ID. A receiving node that has not seen the descriptive advertisement before, sends back a request (REQ) and then the sending node replies with the DATA. SPIN-2 simply incorporates an energy threshold below which a node will not participate in the handshaking. This way SPIN-2 implements negotiation and resource adaptation.

SPIN was tested on a 25-node testbed inside a rectangle of 40x40m² therefore scale and mobility performance cannot be ascertained.

4.4 Geographical and Energy Aware Routing (GEAR)

GEAR exploits location information to forward messages to specific areas of interest. It makes local optimisation decisions considering the cost to send a message from every node to the destination based on locally available information and has the ability to learn efficient routes and bypass *holes* i.e. states where all neighbouring nodes are further away from the target than the current transmitter.

The core assumptions are that the network is immobile and that each node has some way to know its own and neighbours' positions and energy levels.

Therefore the most important limitation of GEAR is that it can only be used for static networks. And while such a protocol could be extended by addressing mobility with a series of static configurations, this is not feasible for GEAR. This is because of the oscillations caused when a node "falls" into a hole. For a large network, it would take a considerable period of time for the oscillation to converge, time that simply is not available in a mobile environment.

4.5 Energy-Aware Data-Centric Routing (EAD)

As can be seen in figure 2, the only communication state that does not consume significant energy is the sleep mode. In this mode however, while the mote is able to sense the environment, it cannot be used to route messages. EAD is a heuristic that builds a tree with maximum number of leaves (this problem is actually NP-hard), rooted at the sink. This way a virtual backbone infrastructure that is responsible for routing the messages is formed by the non-leaf nodes while the leaf nodes can enter the sleep mode and turn their electronics on, only when events that should be reported happen.

All nodes that receive from leafs participate in local competitions. Only the winners in each region will eventually become non-leaf nodes. To serve fairness, the nodes with a reasonable energy level should become non-leaf nodes and remain alive to route messages. To achieve this the time that a node waits until it transmits should be shorter as the energy level of the nodes increases.

As the nodes that form the virtual backbone will execute operations of high power demands, they will be drained out of energy soon unless some mechanism responsible for the broadcast tree maintenance is enabled. In fact EAD adopts the LEACH's rotation method.

EAD introduces two interesting concepts: *neighbouring broadcast scheduling* and *distributed competition* among neighbours. While its concept fits quite well in the data-centric network, as the non-leaf nodes are also responsible for data aggregation, it does not address mobility at all. The problem is that, as EAD designers say, the data-transmit phase is substantially longer than the initialisation. In a highly mobile environment this means in practice that before a round is completed, the tree may already be dissolved. We will briefly discuss some algorithms proposed for minimum energy routing over MANETs based on a pre-computed hierarchy in the next section. In this sense EAD could fit in the next section although there are some substantial differences.

4.6 SPEED

SPEED's main aim is to maintain a uniform delivery speed across the whole network. At the same time it adopts a stateless, real time, architecture that performs location-aware forwarding and constitutes one of its advantages. It contains the following components (summarised in figure 3):

- An API that provides one receiving and three sending services, supporting multicast (geocast), anycast and unicast operations.
- A neighbour beacon exchange scheme that neighbouring nodes use to exchange position information. This is similar to the typical hello message exchanging.
- A single-hop delay estimation scheme, which is used to provide the metric with which network congestion is measured. The delay is calculated as the period between the message transmission and the acknowledgment reception by subtracting the processing time of the receiver.

- The Stateless Non-deterministic Geographic Forwarding (SNGF)

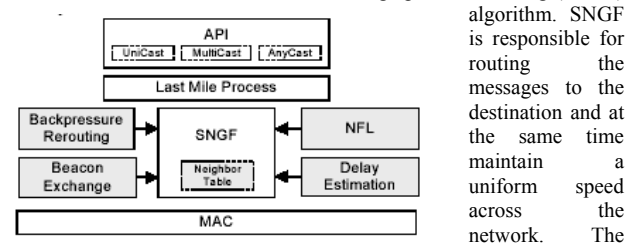


Figure-3: SPEED's architecture (from [20]).

The Stateless Non-deterministic Geographic Forwarding (SNGF) algorithm. SNGF is responsible for routing the messages to the destination and at the same time maintain a uniform speed across the network. The second task is aided by additional components discussed later. SNGF makes local routing decisions by

considering as candidates only those nodes among the neighbours that are closer to the destination than the sending node. Further, out of the resulting set, the nodes with higher *relay speeds* have higher possibilities to receive the message. The relay speed is defined as the advance in distance towards the destination over the hop delay as it is calculated by the previously mentioned component.

- The Neighbourhood Feedback Loop (NFL), which is activated if no neighbouring node, is able to maintain a desired relay speed, which is set as a threshold. A *miss* is considered to take place when either a message is delivered at a speed lower than the threshold or if a collision occurred. Using feedback from the MAC layer, NFL computes the *relay ratio* based on the miss ratios and forwards its value to SNGF which in turn decides whether to forward the message. Using a randomly generated number and the relay ratio, SNGF decides whether to forward the message after all. NFL is a way to relief congestion by introducing a non-deterministic decision on whether to forward a message towards a heavy-traffic area.

- A Back-Pressure Rerouting mechanism that enables nodes to be notified about congested areas using special control messages that are pushed backwards inside the network. Due to the real-time requirement SPEED uses greedy localised routing. The most important side-effect of this design choice is that there is no guaranty that a path to the destination will be actually discovered. Void avoidance is achieved with a backpressure method similar to the one used for congestion.

- The Last Mile Process, which supports multicasting or anycasting upon the message arrival to the destination area.

SPEED introduces an architecture that covers many aspects of the WSNs framework quite successfully. It incorporates stateless, real-time, location-aided routing which enables message delivery to specific areas without unnecessary flooding. It still remains a question though whether its main aim, i.e. to maintain the delivery delay below some certain level, is more important for WSN applications than mobility and scalability, two aspects that are shadowed by the uniform speed specification. It is clear that SPEED becomes more efficient as the message rate increases while for low rates it was less efficient than the protocols with which it was compared (see [20]). This probably makes it more suitable for a continuous WSN where traffic is heavier and spans constantly large parts of the network.

4.7 Berkeley's Energy-Aware Routing

Energy-Aware Routing is inspired by Directed Diffusion. It consists of three phases: The setup or interest propagation, the data communication and the route maintenance phase. In contrast with Directed Diffusion though, it is *destination initiated*.

The destination node floods the network with a request to the direction of the source (see section 6 for methods to achieve directed routing). An intermediate node forwards the message only to the neighbouring nodes that are closer to the source than itself. Each receiving node keeps a table with the cost of the path from the destination to itself, which is calculated as the cost propagated by the sending neighbour plus an energy metric value that corresponds to the link between them. High costs are discarded and not forwarded with the table.

Each node assigns a forwarding probability to each one of its neighbours, which is inversely proportional to its corresponding cost. The average cost that a node needs to pay in order to forward a message is calculated over all its neighbours and forwarded as well.

When the request reaches the source, the data communication phase is entered with the source picking a neighbour according to the probabilities stored in the forwarding table. The data message is forwarded towards the destination in the same manner. Route maintenance is achieved by infrequent localised flooding on behalf of the destination.

The protocol is claimed to result in more uniform energy consumption across the network due to the probabilistic nature of the path selection. Although similarity to Directed Diffusion is claimed, its description refers to an end-to-end communication model rather than the data-centric model that lies in the core of the Directed Diffusion concept. Also the cost propagation introduces an additional control message overhead that could not be affordable in a highly mobile environment. According to these observations one could argue that the algorithm could be useful for specific scenarios, but definitely not as the generic solution on which a complex WSN could rely on.

5. Minimum Energy Broadcast Routing

There is a class of algorithms designed to tackle the Minimum Energy Broadcast Routing problem (MEBR) over MANETs. These algorithms are not directly applicable to the problem discussed here but they introduce some concepts that could prove to be useful to this end. The main methods that are proposed for the MEBR is the well-studied Minimum Spanning Tree (MST), the Broadcast Incremental Power (BIP) and the Adaptive Broadcast Consumption (ABC).

Next we introduce the MEBR problem. The aim is to deliver a message to all nodes of a MANET using the minimum overall energy. A key assumption is that transceivers have the capability of adjusting their transmission range so as to save energy when possible. Also, the algorithms address the static model which in turn could be extended in the case of traditional MANETs to the dynamic one, where nodes are mobile.

- **MSTs** were first proposed for this purpose. An MST builds a tree rooted at the source node, which spans the whole network at the minimum total cost. In this model each edge weight corresponds to the energy required by the wireless link and is typically proportional to the square distance between the communicating nodes. While this seems a suitable method, in reality it is not due to the nature of wireless communications. The key observation is that when a source sends to a destination, all nodes that are inside the transmission range receive the message. This important property is not reflected in the MST approach since each transmission is represented by an edge. In this way the fact that a transmission to a more distant node also covers closer nodes is not modelled because not only the edge that represents the further transmission but also the rest of the edges to all closer nodes contribute to the total cost.

- In **BIP** [3] this property of the wireless environment is identified and the aim is to find an assignment of ranges that minimises the total energy consumption over the MANET. The algorithm begins by the source and gradually adds nodes to the set of the covered nodes. At each step the next node to be added is selected according to the *incremental cost* which should be minimum. At each step the algorithm examines whether it is cheaper to connect the next node to the last added one or to increase the range of any of the already covered nodes.

- **ABC** [4] builds on BIP. It adopts the *circle* concept which means that the connectivity modelling is based on circles that represent ranges rather than on edges. ABC is more efficient than BIP as it performs an in-depth discard of all pre-existing circles that are not useful after a new coverage distribution, while at the same time it has the same time complexity namely $O(n^3)$. The high adaptation of ABC is achieved by maintaining two invariants: First, every node which belongs to the covered set must, at any point, be covered by at least one circle and second, every node in the same set must admit an induced path back to the source node.

As mentioned previously, these methods cannot be considered for WSNs mainly because they build a hierarchy, which is based on knowledge of the whole topology, which is assumed to be static. This initialisation phase is followed by the actual data communication phase when the routing over the built hierarchy takes place. On the other hand, they provide a very good insight on the nature of the wireless communication and the energy conservation concept.

6. Location-Aided Routing

In this section we briefly present the basics of location-aided or position-based routing through methods proposed for traditional MANETs. We identify and distinguish between the following general categories:

- Greedy forwarding,
- Greedy routing with guaranteed delivery,
- Hierarchical routing and
- Geocasting and restricted flooding.

6.1 Greedy Forwarding

This first category includes algorithms to select the next neighbour to which the message will be forwarded among all in-range nodes. They can be considered to belong to the MAC layer and they are based on a local criterion assuming though that the destination position is known. These methods could be used when a WSN protocol needs to make local decisions.

Most Forward within Radius (MFR) [22] forwards the packet to the node that is closer to the destination in an attempt to minimise the number of hops. MFR assumes fixed range which results in

deficiencies when a node is able to adjust its transmission radius. In the latter case Nearest with Forward Progress (NFP) [23] could be a better solution as it selects the nearest neighbour which is closer to the destination. Random Progress Method (RPM) [24] routes the message with equal probability towards a neighbour that enables forward progress. Further, Compass Routing selects the neighbour which is closer to the direction of the final destination. Finally in Geographic Distance Routing (GEDIR) [25], forwarding is similar to MFR with the addition of a termination criterion which applies when the neighbour selected is the one from which the message was forwarded. Intermediate Node Forwarding (INF) [26] is proposed as part of Grid routing (discussed later).

6.2 Greedy Routing with Guaranteed Delivery

The main problem inherent to the methods discussed in the previous paragraph is that the message is not guaranteed to be delivered to the final destination as the local nature of the forwarding decision may lead to a dead-end path. A way to overcome this problem is to forward the message to the neighbour with the smallest negative progress when no node with positive progress exists but this may induce loops into the path.

Face, Greedy-Face-Greedy (GFG) [27] and the Greedy Perimeter Stateless Routing (GPSR) [28] use planar graph traversal to overcome the dead-end problem. They construct a planar graph (i.e. a graph whose edges intersect only at the nodes) and route the messages on faces (i.e. polygons bounded by the graph's edges) that are progressively closer to the destination by using the right-hand rule (i.e. each node forwards the message over the edge counterclockwise about itself, from the edge that it received it). This scheme guarantees delivery if there is a path in the original graph. GFG and GPSR exploit the planar graph traversal only when the message encounters a dead-end by switching from the "greedy" to the "face" or "perimeter" mode.

To enable routing decisions locally, additional information is included in the message such as the location where it switched to the perimeter mode, the point at which it entered the current face, the first edge it traversed on current face, the packet mode itself etc. Of course this is an overhead to be considered when discussing WSNs.

6.3 Hierarchical Routing

Hierarchical routing organises the network into some hierarchy in order to decrease its complexity and increase scalability and mobility. The most important representatives of this class are the Scalable Location Update-based Routing Protocol (SLURP) [29], Terminode [30,31,32] and Grid [26,33] routing.

All these protocols incorporate a distributed location service that disseminate location information across multiple nodes that form a hierarchy. SLURP uses k square subregions, Terminode exploits Virtual Home Regions (VHR) defined by a position and a radius and Grid Location Service (GLS) forms a hierarchy of squares so that order- n squares contain exactly four order- $(n-1)$ squares. In order for a source to send a message to a destination, it first queries a location server node that belongs to the hierarchy about the destination's position.

Another common characteristic is that long distance routing is achieved using some greedy approach and when the message reaches the destination region it is sent to the specified node using one of the on-demand or table-driven protocols discussed earlier. In this context SLURP exploits MFR and DSR, Terminode uses Anchored Path Geodesic Packet Forwarding (AGPF) towards a destination which is defined using a Location Dependent Address (LDA) and Grid combines geographic forwarding [38] with a modified version of DSDV.

While these design choices constitute the protocols that are very scalable, their implementations do not match the lightweight protocol requirement imposed by WSNs' energy and storage limitations. SLURP for example uses at least nine different control packets and maintains four data structures on each node.

6.4 Geocasting and Restricted Flooding

Geocasting is deployed to deliver messages to all nodes inside a specified area. Among other techniques, restricted or partial flooding can be used for this. Restricted flooding isolates and floods only a part of the whole network towards the destination area. The most important protocols that fall in this category are the Location-Based Multicast (LBM) [34], which is an extension of Location-Aided Routing (LAR)

[36] and Distance Routing Effect Algorithm for Mobility (DREAM) [35].

- **LAR** exploits restricted flooding to perform unicasting. Assuming that the source knows the destination's position x_0 at time t_0 and it has an estimation of its speed u (e.g. the average or maximum speed among the mobile nodes), it initiates a route request at $t_1 > t_0$. The *expected zone* i.e. the zone in which the destination node is expected to be, is defined as the circle of radius $u(t_1 - t_0)$, centred at x_0 . In LAR-1 the *request zone* is set to be the smallest rectangular, which includes the source node and the expected zone. The source initiates the route request by sending a message that contains the coordinates of the rectangular. The nodes that are inside the request zone place their IDs in the message and forward it. The ones outside the zone discard it. When the destination receives the request, it replies with a message containing its location, a timestamp and the route. The reply is routed by reversing the path inside the request just like in DSR. In LAR-2 the request zone is defined implicitly. An intermediate node that receives the request, forwards it only if its distance from the destination is smaller than the one of its predecessor from the destination. To achieve this, nodes forward their locations together with the request. **LBM** extends LAR to perform geocasting in the obvious way. In [37] the authors identify the possibility of not having any neighbour in range inside the *forwarding zone* (as the request zone is renamed) causing LBM to fail and propose a Voronoi diagram based method to overcome it.

- In **DREAM** each node transmits control packets periodically in order to inform the network of its location. The frequency of these transmissions is proportional to its speed so as to optimise each rate according to the individual mobility. The concept of the *distance effect* is also introduced. According to this, a *life time* is associated with each control message in order to model the fact that two mobiles that are far apart see each other moving slower than if they were closer, causing the message to be discarded after a specified number of hops. The majority of the messages are *short lived* while other *long lived* ones are sent rarely and traverse the whole network. Assuming an average speed u (or even a speed probability function) for the nodes and knowledge of a destination's position x_0 at time t_0 , the source can route a message at time t_1 towards the disk with radius $u(t_1 - t_0)$ centred at x_0 . Each node forwards the message only to the neighbours that lie inside the area bounded by the disk and its two tangents that cross at the point where the source lies at t_1 .

7. Conclusion

In this paper we have identified the design requirements for an efficient routing protocol for WSNs as imposed by the motes' energy and storage restrictions. We have also presented a survey and classification of existing algorithms proposed for routing over MANETS with and without position-awareness, routing over WSNs and broadcasting using the least possible energy, summarising the most interesting characteristics of each of the protocols and commenting on their applicability on WSNs.

We conclude that the existing protocols address selectively specific requirements and match only some of the design aims we have distinguished. There is certainly not a protocol that supports mobility, scalability, data aggregation and serves energy efficiency and fault tolerance at the same time.

We believe that location awareness is the only difference between traditional MANETS and WSNs that can actually be used for optimisation reasons. There is certainly a trade-off between scalability, mobility and location-awareness in the sense that as the networks' size and mobility increases, the location information bulk becomes larger and the necessary updates more frequent. In this context, an adaptive lightweight location service supporting routing over a hierarchy that enforces balanced energy conservation across the whole network to ensure prolonged lifetime is needed and would be considered as a breakthrough development in this research area.

Acknowledgements

We would like to thank Naranker Dulay for his kind suggestions.

References

- [1] J.M. Kahn, R.H. Katz, K.S.J. Pister, "Next century challenges: Mobile networking for Smart Dust", Proc. MOBICOM, 1999, Seattle, 271-278

- [2] M. Weiser, "The computer for the 21st century", Scientific American, September 1991. 94-104
- [3] J.E. Wieselthier, G.D. Nguyen and A. Ephremides, "On the construction of energy-efficient broadcast and multicast trees in wireless networks", Proc. IEEE INFOCOM, March 2000, 585-594
- [4] R. Klasing, A. Navarra, A. Papadopoulos, S. Perennes, "Adaptive Broadcast Consumption (ABC), a new heuristic and new bounds for the minimum energy broadcast routing problem", Proc. IFIP-TC6 Networking, May 2004, to appear
- [5] D. Estrin et. al, <http://nesl.ee.ucla.edu/tutorials/mobicom02>
- [6] G. J. Pottie and W. J. Kaiser. "Embedding the internet: wireless integrated network sensors", Communications of the ACM, 43(5):51-58, May 2000
- [7] C. Perkins and P. Bhagwat. "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers", In Proc. of the ACM SIGCOMM, October 1994.
- [8] C-C. Chiang, "Routing in clusterhead multihop, mobile wireless networks with fading channel", Proc. IEEE SICON, April 1997, 197-211
- [9] S. Murphy, J.J. Garcia-Luna-Aceves, "An Efficient Routing Protocol for Wireless Networks", ACM Mobile Networks and Applications Journal, Nov. 1996, 183-197
- [10] D. Johnson and D. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks in Mobile Computing", T. Imielinski and H. Korth eds., Kluwer Academic, 1996, 153-181
- [11] C. E. Perkins, "Ad-hoc on-demand distance vector routing," in Proc. MILCOM '97 panel on Ad Hoc Networks, Nov. 1997
- [12] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks", IEEE Infocom, 1997
- [13] C.-K. Toh, "A novel distributed routing protocol to support ad-hoc mobile computing," Wireless Personal Communication, Jan. 1997
- [14] R. Dube et al. "Signal stability-based adaptive routing (SSA) for ad-hoc mobile networks," IEEE Pers. Commun., February 1997
- [15] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks", Proc. ACM MOBICOM, Aug. 2000
- [16] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocols for Wireless Microsensor Networks", In Proc. of Hawaiian Int'l Conf. on Systems Science, January 2000
- [17] W.R. Heinzelman, J. Kulik and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks", Proc. MOBICOM, 1999, Seattle, 174-185
- [18] Yan Yu, Ramesh Govindan and Deborah Estrin, "Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks", UCLA Computer Science Department Technical Report UCLA/CSD-TR-01-0023, May 2001
- [19] A. Boukerche, X. Cheng, and J. Linus, "Energy-aware Data-centric Routing in Microsensor Networks", Proc. ACM MSWiM, 2003, 42-49
- [20] T. He, J. A. Stankovic, C. Lu, and T. F. Abdelzaher, "SPEED: A Stateless Protocol for Real-Time Communication in Sensor Networks", ICDCS, 2003
- [21] R.C. Shah and J.M. Rabaey, "Energy aware routing for low energy ad hoc sensor networks", In Proc. IEEE Wireless Communications and Networking Conference (WCNC), March 2002
- [22] H. Takagi and L. Kleinrock, "Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals", IEEE Transactions on Communications, Vol. COM-32, No. 3, March 1984, 246-257
- [23] T.C. Hou and V.O.K. Li, "Transmission range control in multihop packet radio networks", IEEE Trans. on Communications COM-34(1), 1986, 38-44
- [24] R. Nelson and L. Kleinrock, "The spatial capacity of a slotted ALOHA multihop packet radio network with capture", IEEE Transactions on Communications, 32, 6, 1984, 684-694
- [25] I. Stojmenovic and X. Lin, "GEDIR: Loop-free location based routing in wireless network's", IASTED Int. Conf. on Parallel and Distributed Computing and Systems, Nov. 1999
- [26] D. S.J. De Couto and R. Morris, "Location proxies and intermediate node forwarding for practical geographic forwarding", Technical Report MIT-LCS-TR824, MIT Laboratory for Computer Science, June 2001
- [27] P. Bose, P. Morin, I. Stojmenovic and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks", 3rd int. Workshop on Discrete Algorithms and methods for mobile computing and communications, August 20, 1999
- [28] B. Karp and H.T. Kung, "Greedy perimeter stateless routing (GPSR) for wireless networks," in Proc. ACM/IEEE MOBICOM, 2000
- [29] S.-C. M. Woo and S. Singh, "Scalable routing protocol for ad hoc networks", Wireless Networks, vol. 7, Sept. 2001, 513-529
- [30] L. Blazevic, L. Buttyan, S. Capkun, S. Giordano, J. Hubaux, and J. Le Boudec, "Self-organization in mobile ad-hoc networks: the approach of terminodes", IEEE Communications Magazine, 2001
- [31] L. Blazevic, S. Giordano, and J. Le Boudec, "Self organized terminode routing", Technical Report DSC, Swiss Federal Institute of Technology, 2000
- [32] L. Blazevic, S. Giordano, and J.-Y. Le Boudec, "Self Organized Terminode Routing Simulation", In Proc. ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile systems, MSWiM, 2001
- [33] D.S.J. De Couto, J. Jannotti, D. Karger, J. Li, and R. Morris. "A scalable location service for geographic ad hoc routing", In Proc. ACM/IEEE MobiCom, August 2000
- [34] Young-Bae Ko and Nitin H. Vaidya, "Geocasting in mobile ad hoc networks: Location based multicast algorithms", In Proc. IEEE Workshop on Mobile Computing Systems and Applications, February 1999
- [35] S. Basagni, I. Chlamtac, V.R. Syrotiuik, and B.A. Woodward, "A Distance Routing Effect Algorithm for Mobility (DREAM)", MOBICOM'99, 1999
- [36] Y.-B. Ko and N. H. Vaidya. "Location-aided routing (LAR) in mobile ad hoc networks", In ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom'98), October 1998
- [37] I. Stojmenovic, "Voronoi diagram and convex hull based geocasting and routing in wireless networks", SITE, University of Ottawa, TR-99-11, Dec. 1999
- [38] G. Finn, "Routing and Addressing Problems in Large Metropolitan-scale Internetworks", ISI Research Report ISI/RR-87-180, University of Southern California, March 1987



Aris Papadopoulos received his BSc from the Physics Dept. of the Univ. of Athens (2001) and MSc from the Dept. of Computer Science of King's College, Univ. of London (2002). From then on he worked as a Research Assistant at the Univ. of Athens (Electronics Laboratory) and at INRIA Sophia Antipolis (Project MASCOTTE). From November 2003 he is a PhD student within the Distributed Software Engineering group of the Dept. of Computing at Imperial College London.



Julie A. McCann received her PhD doctorate in 1992 in Ireland, and then joined City University, Computer Science Dept. where she became the leader of the Systems Architecture Research Centre, managing many Operating System, Database and Distributed systems research projects. She is currently a Senior Lecturer at Imperial College London where she lectures in Database Systems Operating Systems. Her research lies in adaptive systems architectures for improved quality of service and performance. Julie is a Chartered Engineer and a Member of the British Computer Society.