A Taxonomy of Routing Protocols for Wireless Sensor Networks

Gergely Ács, Levente Buttyán
Laboratory of Cryptography and Systems Security (CrySyS)
Department of Telecommunications
Budapest University of Technology and Economics, Hungary
{acs, buttyan}@crysys.hu

January 12, 2007

Abstract

Wireless sensor networks are large scale networks consisting of a large number of tiny sensor nodes and a few base stations, which communicate using multi-hop wireless communications. The design of energy efficient routing protocols for such networks is a challenging task, which has been in the focus of the sensor network research community in the recent past. This effort resulted in a huge number of sensor network routing protocols. The proposed protocols show a high variety, which stems from the diverse requirements of the various envisioned application scenarios. In this work, we propose a taxonomy of sensor network routing protocols, and classify the mainstream protocols proposed in the literature using this taxonomy. We distinguish five families of protocols based on the way the next hop is selected on the route of a message, and briefly describe the operation of a representative member from each group.

1 Introduction

Sensor networks are composed of resource constrained sensor nodes and more resourced base stations. All nodes in a network communicate with each other via wireless links, where the communcation cost is much higher than the computational cost. Moreover, the energy needed to transmit a message is about twice as great as the energy needed to receive the same message. Consequently, the route of each message destined to the base station is really crucial in terms network lifetime: e.g., using short routes to the base station that contains nodes with depleted batteries may yield decreased network lifetime. On the other hand, using a long route composed of many sensor nodes can significantly increase the network delay.

Unfortunately, some requirements for the routing protocols are conflicting. Always selecting the shortest route towards the base station causes the intermediate nodes to deplete faster, which results in a decreased network lifetime (if we measure the network lifetime by the time that lasts until the first node dies in the entire network). At the same time, always choosing the shortest path may result the lowest energy consumption and lowest network delay globally. Ultimately, the routing objectives are tailored by the application; e.g., real-time applications require minimal network delay, while applications performing statistical computations may require maximized network lifetime. Hence, different routing mechanisms have been proposed for different applications [1]. These routing mechanisms primarily differ in terms of routing objectives and routing techniques, where the techniques are mainly influanced by the network characteristics.

In this paper, we propose a taxonomy of sensor network routing protocols, and classify the mainstream protocols proposed in the literature using this taxonomy. We distinguish five families of protocols based on the way the next hop is selected on the route of a message, and briefly describe the operation of a representative member from each group.

2 Taxonomy of routing protocols

In order to select the most suitable routing mechanism for a sensor application, we have to classify all routing protocols according to a well-defined taxonomy. Using this classification, all protocols become comparable for an application designer. As a part of this taxonomy, we define a system model that describes the network and operational characteristics of the routing protocols, and an objective model that describes the routing objectives of the protocols. Furthermore, the system model encompasses the definiton of the network model as well as the operation model of routing protocols.

2.1 Network model

This model desribes the characteristics of a network that can be divided into two groups: the characterisitcs of base stations, and the characteristics of sensor nodes.

2.1.1 Base station

It is commonly agreed that the base station is a powerful device with unconstrained energy supply and computational capacity. However, the following characteristics of a base station may severely influance the operation of a routing protocol:

Number: The number of the base stations can be one or more than one. In most practical applications, the increased number of base stations provides more robust data gathering, and may also decrease the network delay. However, the typical number of the base stations is one. If only one base station is presented (and there is no need for explicit communication between sensor nodes), the destination node for all messages is the same, while in case of multiple base stations, the destination node can differ for some messages.

Mobility: During the routing process, the base station can be fixed (stationary) or mobile. In some applications, where the number of base stations is too small to ensure acceptable network delay and robustness, the base station supports mobility during data gathering. This property of the base station severely affects the routing protocol, since all nodes in the network field cannot be continously aware of the current position of the base station, and the routing mechanism needs to find the mobile base station in the field. Moreover, the routing topology created by a routing protocol may heavily vary in time that causes extra overhead in the network layer. Some routing protocols cannot be employed with mobile base stations, while others tolerate limited mobility.

Presence: The base station can be either continously or partially presented during the routing process. In the latter case, the routing protocol must support the temporary lack of a base station (e.g., the base station is switched off for a certain amount of time due to maintenance reasons), since a missing base station cannot definitely mean a failure. Thus, the messages should not be dropped or rerouted rather their delivery should be delayed.

Coverage: Many routing protocols assume that base station can cover the whole network field by its power range. In these networks, the base station can reach every other nodes, if there are no obstacles in the field. Therefore, there is no need for routing between the sensor nodes and the base station in such cases. However, we note that it is not a reasonable assumption in most practical applications; a more realistic assumption would be that the base station can communicate with only some nodes in its close vicinity.

2.1.2 Sensor nodes

In most sensor networks, sensor nodes are homogeneous tiny devices with constrained energy supply and computational capabilities. In addition, we assume that all sensor nodes are stationary. The following characteristics of sensor nodes may differ for some networks. Hence, they can influance the protocol operation.

Deployment: Sensor nodes can be deployed in either a deterministic or a random fashion. When the nodes are deployed along a road-side, or in a metro-station, the deployment is rather deterministic than random. In these cases, the routing protocol should adapt to the fixed network topology. However, numerous routing protocols proposed so far assume that the deployment is random (e.g., the nodes are dropped out from a helicopter).

Transmission power: The transmission power can be either dynamically adjustable or fixed. In the latter case, each sensor node transmits each message using the same energy level. In the former case, every node can calculate what energy level should be used to transmit a message to a neighboring node. This energy level may be inversely proportional to the cost assigned to the neighboring node.

Coverage: It is commonly assumed that a sensor node cannot reach all nodes in the network field. A routing protocol can require large transmission power per node in order to get a fully connected network. However, it can only be beneficial in small-sized networks due to the large energy consumption and interference range.

Addressing: The task of routing in sensor networks is to deliver the queries coming from the base station to the sensor nodes which have the requested data (in case of query-driven routing protocols, see later), and to return the requested data to the base station. Accordingly, we can distinguish the addressing method of queries and responses:

- Query-addressing: All routing protocols which use query dissemination in the networks employ data-based (What is the average temperature?), or location-based addressing (What is the average temperature in location (x, y)?).
- Response-addressing: The response is either returned on the reversed path which the query traversed, or it is routed back purely based on location information. In the former case, neighboring nodes use locally (or globally) unique identifiers to identify the neighbor from which they received the query, and which is further used to forward the reply towards the destination.

MAC interface: The data-link layer can be responsible for neighbor discovery (where the neighbor definition is protocol-dependent). In addition, it also need to perform the calculation of cost values (where the cost definition is also protocol-dependent). Some routing protocols are integrated with the data-link layer in order to achieve better performance in terms network delay and energy consumption (cross layer design). However, the data-link layer is generally not responsible for these tasks. Thus, the routing protocol itself must calculate its own neighbor list and the costs of neighbors.

2.2 Operational model

The operational model describes the main orthogonal operational characteristics of a routing protocol.

Communication pattern: A routing protocol can support the communication from sensor nodes to sensor nodes, from base stations to sensor nodes, as well as from sensor nodes to base stations.

- Node-to-Node: This communication pattern is not typical for sensor networks, only those protocols support that inherently which were primarily proposed for ad hoc networks but can also be used in sensor networks. Generally, there is no need for this kind of communication in sensor networks.
- Node-to-Base station: This pattern need to be supported in order to route responses back to the base station. This communication pattern is typically reverse-multicast (many-to-one), a.k.a. convergecast, which means that every sensor node is able to send (directly or indirectly) a message to any base station. If there are multiple base stations or only one node is responsible for gathering and transmitting the sensed data to the base station, this pattern can also be unicast.
- Base station-to-Node: This pattern need to be supported in order to route requests originated from the base station to sensor nodes. This communication pattern is typically anycast (one-to-many), which means that any sensor node which has the requested data can respond to the query. If some nodes are uniquely identified in the network (by their ids, locations, etc.), then multicast (one-to-many) and unicast (one-to-one) patterns can also be supported. All these patterns mean that the base station(s) are able to send (directly or indirectly) a message to any sensor nodes.

Hierarchy: Employing hierarchical routing protocols, a hierarchy level is assigned to each node, and a node only forwards those messages that are originated from a lower-level node. Optionally, a node aggregates incoming data and forwards this aggregated data to upper-layer nodes. The base station can be found on the top of the hierarchy. The hierarchy construction can be dynamic or static. Using dynamic construction, the role of the aggregator is rotated, and all nodes that are selected an aggregator will forward all data to their aggregator. The aim of forming this hierarchy is to prolong the network lifetime. Using non-hierarchical protocols, each sensor node can accept all messages coming from any other sensor nodes for further aggregation and forwarding. Thus, any sensor node can behave as an aggregator node in non-hierarchical architectures.

Delivery method: In most routing protocols, a node selects only a single path towards the base station, and the only instance of a message (single/single) is forwarded along this single path. However, a node may also select multiple paths, and the node forwards either the single instance of a message on a deterministically or randomly chosen single path (multiple/single) or one copy of a message per path (multiple/multiple).

Computation: Each sensor node selects the next-hop towards the base station either by itself in a decentralized manner, or every node sends its neighbor list to the base station and the base station computes the next-hop for all nodes in the network in a centralized manner. Although the centralized computation gives optimal solution, it may yield heavy network communication, which is only tolerable in small-sized networks with fixed network topology. Using decentralized or centralized computations, all nodes only store the identification of their neighbors, where the neighbor definition is protocol-dependent. In the simplest case, a node A considers another node B as a neighbor, if A receives a routing message sent by B. In other cases, nodes discover their neighbors by broadcasting simple HELLO messages in a certain energy level. A node A considers another node B as a neighbor, if A receives a HELLO message sent by B.

Next-hop: A common characteristic of all protocols is that each node selects its next-hop (for the query and/or the response) towards the destination based on locally stored information, which may include the routing costs, next-hop identifiers, etc. The next-hop can be selected by

- randomly among all neighbors (probabilistic)
- inferring routing information from the sensed data that is carried by the message (content-based)
- using the stored routing control information (control-based)
- using a hierarchical-based scheme (hierarchical)
- using geographic positions (location-based)
- broadcasting the message and the neighbors decide whether to re-broadcast the message (broadcast-based)

If both queries and responses are routed in a location-based or broadcast-based manner, a node is typically required to store only negligible amount of routing information like the positions of neighbors or its own routing cost. Routing protocols belonging to this group are *stateless* protocols. On the other hand, if the queries or responses are routed in a probabilistic, hierarchical, content-based or control-based manner, a node may need more extensive processing or storage resources. These routing protocols are also referred as *stateful* protocols.

Reporting model: The reporting model describes *what* initiates the data reporting process. In this sense, we distinguish time-driven, query-driven, and event-driven protocols.

- Time-driven: Employing a time-driven routing protocol, a sensor node is triggered in specific moments, when it should perform its measurement task and forwards the measurement to its next-hop neighbor. These activations can be periodic or one-shot in time. Short periods may cause more traffic in the network, and the quality of routing in terms of energy efficiency becomes a crucial concern. Time-driven sensors may be pre-programmed, or the reporting schedule may come with explicit queries. Furthermore, a time-driven routing protocol can support the reporting of
  - complex (the reported data has several atomic components, e.g., temperature and humidity) or simple (atomic) data (e.g., only temperature is reported)[1],
  - aggregated or non-aggregated data,
  - replicated (more than one sensor can provide the requested information) or unique data (only one sensor can provide the requested information).
- Query-driven: In most sensor applications, the base station disseminates its query in the network, while the sensor nodes try to resolve this query, and they may send a response back to the base station. Hence, the task of a query-driven protocol is to route the queries to the measurement area, and to route back the response to this query. Furthermore, a query-driven routing protocol can support the reporting of
  - complex or simple (atomic),
  - aggregated or non-aggregated,
  - replicated or unique data.
- Event-driven: A sensor node sends a measurement towards the base station only if a given event occurs (e.g., the temperature falls below a certain threshold). Furthermore, an event-driven routing protocol can support the reporting of
  - complex or simple (atomic),
  - aggregated or non-aggregated,
  - replicated or unique data.

Most routing protocols belong to multiple reporting models.

2.3 Routing objectives

Some sensor applications only require the successful delivery of messages between a source and a destination. However, there are applications that need even more assurance. These are the real-time requirements of the message delivery, and in parallel, the maximization of network lifetime.

[1]Here, we assume that the sensors may responsible for different sensing tasks.

Non-real time delivery: The assurance of message delivery is indispensable for all routing protocols. It means that the protocol should always find the route between the communicating nodes, if it really exists. This correctness property can be proven in a formal way, while the average-case performance can be evaluated by measuring the message delivery ratio.

Real-time delivery: Some applications require that a message must be delivered within a specified time, otherwise the message becomes useless or its information content is decreasing after the time bound. Therefore, the main objective of these protocols is to completely control the network delay. The average-case performance of these protocols can be evaluated by measuring the message delivery ratio with time constraints.

Network lifetime: This protocol objective is crucial for those networks, where the application must run on sensor nodes as long as possible. The protocols aiming this concern try to balance the energy consumption equally among nodes considering their residual energy levels. However, the metric used to determine the network lifetime is also application dependent. Most protocols assume that every node is equally important and they use the time until the first node dies as a metric, or the average energy consumption of the nodes as another metric. If nodes are not equally important, then the time until the last or high-priority nodes die can be a reasonable metric.

Protocol	Network model										Objectives
	Base station				Sensor nodes						
	Num.	Mobility	Presence	Coverage	Deploy.	T. power	Coverage	Addressing		MAC	
								Query	Response		
Rumor Routing [6]	One	Fixed	*	Partially	Random	*	Partially	Data	ID	*	Non real-time
MCFA [9]	One	Fixed	*	Partially	*	*	Partially	X	X	*	Non real-time
Energy Aware Routing [14]	More	Fixed	*	Partially	*	Adjustable	Partially	Data	ID	Needed	Non real-time, lifetime
Directed Diffusion [3]	More	Limited	Continous	Partially	*	*	Partially	Data	ID	*	Non real-time
GBR [21]	More	Limited	Continous	Partially	*	*	Partially	Data	ID	*	Non real-time, lifetime
LEACH [2]	One	Fixed	Continous	Full	Random	*	Partially	Data	ID	*	Non real-time, lifetime
TEEN [4]	One	Fixed	Continous	Partially	Random	*	Partially	Data	ID	*	Real-time, lifetime
APTEEN [5]	One	Fixed	Continous	Partially	Random	*	Partially	Data	ID	*	Real-time, lifetime
PEGASIS [8]	One	Fixed	Continous	Partially	Random	*	Full	Data	ID, Location	*	Non real-time, lifetime
ACQUIRE [15]	More	Limited	Continous	Partially	*	*	Partially	Data	ID	*	Non real-time
IDSQ/CADR [7]	More	Fixed	*	Partially	*	*	Partially	Data	ID, Location	*	Non real-time, lifetime
Geographic Routing (guaranteed delivery) [20, 17, 19, 22, 23]	More	Mobile	*	Partially	*	*	Partially	Location	Location	*	Non real-time
GEAR [16]	More	Limited	*	Partially	*	*	Partially	Location	Location	*	Non real-time, lifetime
MECN [11]	One	Fixed	*	Partially	*	Adjustable	Partially	X	Location	*	Non real-time, lifetime
TTDD [10]	More	Mobile	*	Partially	*	*	Partially	Location	Location	*	Non real-time, lifetime
SAR (DAM) [13]	More	Limited	*	Partially	*	*	Partially	X	ID	*	Non real-time, lifetime
HPAR [12]	One	Fixed	*	Partially	*	Adjustable	Partially	*	*	*	Non real-time, lifetime
SPEED [18]	More	Fixed	Continous	Partially	*	*	Partially	Location	Location	Needed	Real-time, lifetime

Table 1: Network and objective model

Protocol	Operational model										
	Next-hop		Communcation pattern			Hierarchy	Delivery	Computation	Reporting model		
	Query	Response	N2N	N2BS	BS2N				Time-driven	Query-driven	Event-driven
Rumor Routing [6]	Random	Random	X	Rev. M.	Anycast	No	Single / Single	Decent.	X	Non-agg., Simple	X
MCFA [9]	X	Broadcast	X	Rev. M.	X	No	Single / Single	Decent.	*	X	*
Energy Aware Routing [14]	Broadcast	Random	X	Rev. M.	Anycast	No	Multiple / Single	Decent.	*	*	X
Directed Diffusion [3]	Broadcast	Content	X	Rev. M.	Anycast	No	Multiple / Multiple	Decent.	*	*	X
GBR [21]	Broadcast	Content	X	Rev. M.	Anycast	No	Multiple / Multiple	Decent.	*	*	X
LEACH [2]	*	Hierarchical	X	Rev. M.	Anycast	Yes	Single / Single	*	*	*	X
TEEN [4]	*	Hierarchical	X	Rev. M.	Anycast	Yes	Single / Single	*	X	*	*
APTEEN [5]	Hierarchical	Hierarchical	X	Rev. M.	Anycast	Yes	Single / Single	*	*	*	*
PEGASIS [8]	*	Hierarchical	X	Unicast	Anycast	Yes	Single / Single	*	*	*	X
ACQUIRE [15]	Random	Controlled	X	Rev. M.	Anycast	No	Single / Single	Decent.	X	Non-agg.	X
IDSQ/CADR [7]	Content	Contr., Loc.	X	Rev. M.	Anycast	*	Single / Single	Decent.	X	*	X
Geographic Routing (guaranteed delivery) [20, 17, 19, 22, 23]	Location	Location	Unicast	Unicast	Unicast	No	Single / Single	Decent.	Non-agg., Unique, Simple	Non-agg., Unique, Simple	Non-agg., Unique, Simple
GEAR [16]	Location	Location	Unicast	Rev. M., Unicast	Anycast	No	Single / Single	Decent.	Non-agg.	Non-agg.	Non-agg.
MECN [11]	X	Location	X	Rev. M.	Anycast	No	Single / Single	Decent.	*	X	*
TTDD [10]	Hierarchical	Hierarchical	X	Rev. M.	Anycast	Yes	Single / Single	Decent.	Non-agg., Unique	Non-agg., Unique	X
SAR (DAM) [13]	X	Hierarchical	X	Rev. M.	Anycast	Yes	Single / Single	Decent.	X	X	Simple
HPAR [12]	Hierarchical	Hierarchical	X	Rev. M.	*	Yes	Single / Single	Cent.	*	*	*
SPEED [18]	Location	Location	Unicast, Anycast, Multicast	Unicast, Anycast, Rev. M.	Unicast, Anycast, Multicast	No	Single / Single	Decent.	Non-agg.	Non-agg.	Non-agg.

Table 2: Classification of routing protocols with respect to operational model.

3 Protocols

Table 1 overviews the network model and routing objectives of the most significant routing protocols, while Table 2 describes the operational model of the same protocols. We merged some protocols into a single row which have identical system model and routing objectives (we put a star after their common name). The content of each cell is explained in Section 2, where a single star in a cell means an arbitrary value (i.e., the protocol supports all values of the cell). In the followings, we distinguish routing protocol families based on how a protocol selects a next-hop on the route of the forwarded message. We also briefly describe the operation of a significant protocol from each family.

3.1 Content-based routing protocols

These protocols determine the next-hop on the route purely based on the query content. This type of routing protocols fits the most to the architecture of sensor networks, since the base station do not query specific nodes rather it requests only for data regardless of its origin. From this family of protocols and paradigms, we briefly describe Directed Diffusion [3], which is considered to be the basis of other protocols like GBR [21], and Energy Aware Routing [14].

Using Directed Diffusion, the base station initially floods the network with an interest, which contains attribute-value pairs describing the requested data. Upon reception of an interests, each sensor node sets a gradient towards the sender node. If a node receives the same interest from different neighbors, then the node can set multiple gradients, which correspond to the same interest, pointing to different neighbors. The neighbors are differentiated by locally unique identifiers. The sources forward the data along their gradients that is followed by the intermediate nodes up to the base station along the route. If there are more gradients at a node for the same interest, then the node forwards one copy of the message for each neighbor along each gradient. A gradient defines the requested data at each sensor node in conjuction with the next-hop towards the base station for which a message, that contains the requested data, should be forwarded. Each gradient is weighted proportionally to the amount of data that is allowed to traverse the gradient. After a while, the base station selects the route with the best quality and increases the weight of the gradients along the route (positive re-inforcement), whereas it decreases on the others (negative re-inforcement). Intermediate nodes may aggregate the received data, and forward this aggregated data along the corresponding gradients with a rate that is proportional to the weight of the gradient. The base station periodically re-sends the interests along the used routes in order to keep alive the gradients of intermediate nodes. In this way, the base station keeps the empirically best routes and eliminates the routes that have worse quality. Optionally, all nodes can use cache techniques in order to achieve shorter response time and increase robustness.

The paradigm fits well for tracking applications, and it only requires the usage of some local addressing method to distinguish the one-hop neighbors of a node. One main drawback of Directed Diffusion is that it consumes significant network resources until the selection of the empirically best route.

3.2 Probabilistic routing protocols

In order to aid load-balancing and increased robustness, the next-hop on the route can be selected in a random fashion among all neighbors. These protocols assume that all sensor nodes are homogeneous and randomly deployed. We overview the operation of Energy Aware Routing [14] protocol as follows.

The main objective of Energy Aware Routing is to prolong the network lifetime by aiding load-balancing. This on-demand protocol is destination initiated, which means that the destination initiates the construction of the routing topology. Using this routing protocol, sensor nodes randomly select the next-hop neighbor for each message to be forwarded. The probability of selecting a certain neighbor is inversely proportional to its cost. This cost of a neighbor depends on the residual energy of the node, and the energy needed to transmit a message to this node. The neighbor list and their corresponding cost values are provided by the MAC protocol. The protocol saves 21.5% more energy and prolongs the network lifetime by 44% compared to Directed Diffusion (if we measure the network lifetime by the time until the first node runs out of its energy supply).

The routing protocol consists of three phases:

1. Setup-phase
2. Data communication phase
3. Route maintenance

Setup-phase. Initially, the destination disseminates a request message in the network using a controlled flooding technique. By this message, each node determines all routes with their costs towards the destination.

1. The destination floods the network in the direction of the source node with a request message containing a cost value initially set to 0: $Cost = 0$

2. Every intermediate node forwards the requests for those neighbors, which are closer to the source, but farther from the destination than the sender of the request. Formally, node N_i sends a request to N_j, where N_j is a neighbor of N_i, only if the following equations hold: $d(N_i, N_S) \geq d(N_j, N_S)$, $d(N_i, N_D) \leq d(N_j, N_D)$, where $d(N_i, N_j)$ denotes the distance of N_i and N_j, and N_S, N_D are the identifiers of the source and the destination, resp.

3. Upon the reception of the request, N_j calculates the cost of the route from N_j to the destination in the following way: $C_{N_j,N_i} = Cost + Metric(N_j, N_i)$, where $Metric(N_j, N_i)$ denotes the metric between nodes N_j and N_i (see later).

4. The requests with too high costs are silently dropped by N_j. Only the requests with low costs are considered, and the corresponding neighbor is added to the routing table of N_j: $FT_j = \{i | C_{N_j,N_i} \leq \alpha \cdot (\min_k C_{N_j,N_k})\}$

5. N_j assigns a probability to each neighbor in its table FT_j: $P_{N_j,N_i} = \frac{1/C_{N_j,N_i}}{\sum_{k \in FT_j} 1/C_{N_j,N_k}}$

6. N_j calculates the average cost of all routes, that are represented in FT_j, towards the destination: $Cost(N_j) = \sum_{i \in FT_j} P_{N_j,N_i} \cdot C_{N_j,N_i}$

7. The cost value of the request to be re-broadcast is set to this average cost: $Cost = Cost(N_j)$, and the request is re-broadcast according to Step 2.

The metric used in Step 3 is calculated as follows: $C_{i,j} = e_{i,j}^{\alpha} \cdot R_i^{\beta}$, where $C_{i,j}$ denotes the cost metric between node N_i and node N_j, $e_{i,j}$ denotes the energy consumed by transmitting a message from N_i to N_j, and R_i is the residual energy level of N_i normalized to the initial energy level. α and β are tunable parameters.

Data communcation phase.

1. The source sends the message to one of its neighbors, where the neighbor is randomly selected with a probability that is equal to the probability assigned to the corresponding neighbor in the routing table of the source.

2. Every intermediate node selects a next-hop for the message in the same way as the source. Namely, it selects a neighbor from its routing table with a probability that is equal to the probability assigned to the neighbor, and forwards the message to this selected neighbor.

3. This process repeats until the message reaches the destination (base station).

Route maintenance phase. The destination (base station) infrequently updates the routing table of all nodes by flooding the network with new requests.

The drawbacks of the protocol are its complex addressing method, and the increased communication overhead during the setup phase compared to Directed Diffusion.

3.3 Location-based routing protocols

These protocols select the next-hop towards the destination based on the known position of the neighbors and the destination. The position of the destination may denote the centroid of a region or the exact position of a specific node. Location-based routing protocols can avoid the communication overhead caused by flooding, but the calculation of the positions of neighbors may result extra overhead. The local minimum problem is common for all decentralized location-based routing protocols: it might happen that all neighbors of an intermediate node are farther from the destination than the node itself. In order to circumvent this problem, every protocol uses different routing techniques. Here, we describe the operation of GEAR (Geographical and Enegy Aware Routing) [16], which is a sensor-specific location-based routing protocol.

Employing GEAR, a sensor node sends the request to only one neighbor towards the destination. Thus, the initial flood of the query is avoided and the protocol saves more energy than Directed Diffusion. The response can be routed in the same way as the query is routed, or some different mechanism may be used for response routing similar to Directed Diffusion. Each node maintains an estimated and a learned cost value for each destination. The learned cost value is used to circumvent holes in the network, and it is considered as a refinement of the estimated cost value. If there are no holes along a route (every intermediate node has

a neighbor that is closer to the destination than the node itself), then the learned cost value equals to the esimated cost value for the destination at a node. The protocol consists of two phases:

1. Forwarding the messages towards the target region
2. Disseminating the message within the target region

Forwarding the messages towards the target region. An intermediate node N selects the next-hop from those neighbors that are closer to the destination than N. The next-hop neighbor must have the minimal learned cost value among the closer neighbors. If such neighbor does not exist, then N selects the neighbor which has the minimal learned cost value among all neighbors. Initially, the learned cost value equals to the estimated cost value for all nodes, where the latter one can be computed using the following formula: $c(N_i, R) = \alpha d(N_i, R) + (1-\alpha)e(N_i)$, where $d(N_i, R)$ is the distance between neighbor N_i and the centriod of the target region, $e(N_i)$ is the normalized residual energy of N_i, and α is a tunable parameter. Every intermediate node calculates its own estimated cost, and broadcasts its cost value. Thus, every node will be aware of the estimated cost of its neighbors. Initially, the learned cost value for destination R is $h(N_i, R) = c(N_i, R)$. After N sends the message to its selected neighbor N_{min} (which has the minimal learned cost value for R), N updates its own learned cost: $h(N, R) = h(N_{min}, R) + C(N, N_{min})$, where $C(N, N_{min})$ denotes the cost of transmitting a message from N to N_{min} in terms of energy, distance, normalized residual energy, or a combination of these metrics. Therefore, if the path from N to R is composed of n nodes, the learned cost value of the path converges to the real cost of the path within n steps. Additionally, all nodes broadcast their own learned costs infrequently. Hence, nodes can circumvent any holes in the network using learned cost values with appropriate update techniques.

Apart from prolonging the network lifetime, GEAR successfully deliver even 80% more messages than other location-based routing protocols like GPSR [17].

Disseminating the message within the target region. When a message reaches the target region, the nodes inside this region employ either controlled or recursive flooding technique in order to disseminate the message inside the region. Controlled flooding is suggested to be used if nodes are not densely deployed. In high-density networks, recursive geographic flooding is more energy effcient than restricted fooding. In that case, the region is divided into four subregions and four copies of the message are created. This splitting and forwarding process continues until the regions with only one node are left.

3.4 Hierarchical-based routing protocols

In case of hierarchical protocols, all nodes forward a message for a node (also called aggregator) that is in a higher hierarchy level than the sender. Each node aggregates the incoming data by which they reduce the communication overload and conserve more energy. Therefore, these protocols increase the network lifetime and they are also well-scalable. The set of nodes which forward to the same aggregator is called cluster, while the aggregator is also referred as clusterhead. Clusterheads are more resourced nodes, where resource is generally means that their residual energy level is higher than the average. The reason is that they are traversed by high traffic and they perform more computation (aggregation) than other nodes in the cluster. Hierarchical routing is mainly two-layer routing where one layer is used to select clusterheads and the other layer is used for routing. In the followings, we overview the operation of LEACH (Low Energy Adaptive Clustering Hierarchy) protocol [2], which has been served as a basis for several other routing protocols like TEEN [4], APTEEN [5], PEGASIS [8], etc.

In LEACH, nodes dynamically form a cluster in a distributed manner. Clusterheads are elected randomly, and this role is dynamically rotated in order to aid load-balancing. Each clusterhead aggregates all data coming from its cluster, and the aggregated data is forwarded directly to the base station. Therefore, LEACH assumes that all nodes in the network are able to reach the base station directly. LEACH uses a TDMA/CDMA MAC to reduce inter-cluster and intra-cluster collisions. However, data collection is centralized and is performed periodically. Therefore, this protocol is most appropriate when there is a need for constant monitoring by the sensor network. Simulations showed that only 5% of nodes need to act as clusterheads in order to minimize energy consumption. The operation of LEACH consists of two phases; setup phase and steady state phase. These phases periodically repeats after each other, a consecutive setup and steady state phase is also called a round during the protocol run.

Setup phase. In the setup phase, clusters are created and clusterheads are selected. A given fraction of nodes, denoted by p, declare themselves as clusterheads at the beginning of each round in the following way (from simulation results, p typically equals to 0.05). A sensor node n chooses a random number between 0 and 1. If this number is greater than a threshold denoted by $T(n)$, then node n declares itself as a clusterhead,

where $T(n) = \frac{p}{1-p\cdot(r \mod 1/p)}$ if $n \in G$. In the formula of $T(n)$, r denotes the round number, and G is the set of nodes which have not been selected as a clusterhead in the last $1/p$ rounds. Afterwards, the newly selected clusterheads advertise their clusterhead status in a message broadcast in a certain energy level. Every node except the clusterheads decide on the cluster to which they want to belong to. This decision is based on the signal strength of the advertisement. The non-clusterhead nodes inform their selected clusterheads that they will be a member of the cluster. Afterwards, a clusterhead creates a TDMA schedule and assigns a time slot to each node when it can transmit. This schedule is broadcast to all members in the cluster.

Steady state phase. In the steady state phase, the cluster-members forward all measured data to their clusterheads. After receiving all data, a clusterhead forwards the aggregated data to the base station. At the end of the phase (after a certain time determined a priori), the network is switched to setup phase again in order to create new clusters. The duration of the steady state phase is longer than the duration of the setup phase in order to minimize overhead.

A disadvantage of the protocol is that it is not applicable to networks deployed in large areas, since all nodes must be able to reach the base station directly. Another drawback is that the protocol assumes that all cluster-members continously report data to their clusterhead. Furthermore, it might happen that the elected clusterheads will be concentrated in one part of the network. Hence, some nodes will not have any clusterheads in their vicinity. Moreover, the protocol assumes that all nodes have the same initial energy level. All despite, the main problem might be the extra overhead caused by dynamic clustering.

3.5 Broadcast-based routing protocols

The operation of these protocols is very straightforward. Each node in the network decides individually whether to forward a message or not. If a node decides to forward, it simply re-broadcasts the message. If it declines to forward, the message will be dropped. To the best of our knowledge, the only representative of this protocol family is called MCFA (Minimal Cost Forwarding Algorithm) [9].

The main advantage of MCFA is that nodes do not store any information about their neighbors, only their own cost. The protocol consists of two phases. In the first phase, each node calculates its own cost which is eventually the cost of the minimal costed route from the node to the base station. In order to perform this cost calculation, the base station floods the whole network with a request message with a cost field C initialized to 0. Initially, all node costs are set to infinity. A node N_i, which receives the request message from node N_j, delays the re-broadcasting with a time proportional to $\alpha \cdot C_{N_i,N_j}$ (in order to choose the correct α value for the link between N_i and N_j the protocol should take into account the link delay, and link reliability), where C_{N_i,N_j} denotes the cost of the link between N_i and N_j (energy consumption, delay, etc.). Afterwards, N_i updates the cost field in the request message: $C = C + C_{N_i,N_j}$, sets its own node cost to C, and finally re-broadcasts the updated request message. After re-broadcasting, N_i declines to re-broadcast any further request messages. In [9], the authors proved that in ideal cases (when α is sufficiently large) each node re-broadcast only one request message, which contains the minimal cost of the node.

In the second phase, all nodes are able to forward any messages towards the base station in the following way. The source node N places its own cost C_N into the message to be sent, and finally broadcasts the message. A node M receiving this message checks whether $C_N - C_{N,M} = C_M$. If it holds, then M is on the minimal costed route between the base station and node N, so it re-broadcasts the message after placing $C_{N,M}$ into the message. Otherwise, M drops the message. Every subsequent node can check whether it lies on the minimal costed route or not based on the cost values inferred from the message.

A drawback of the protocol is that *every* node receiving a message must perform extra computation in order to determine whether it is on the minimal route or not.

4 Summary

In this work, we classified routing protocols mainly proposed for wireless sensor networks. In Section 2, we presented a novel taxonomy of sensor routing protocols. As a part of this taxonomy, we defined a network model, an operational model, and various routing objectives. In Table 1 and 2, we listed the most significant sensor routing protocols according to this taxonomy, where Table 1 contains the classification of routing protocols based on the network model and routing objectives, while Table 2 describes the operational characteristics of the same protocols. In Section 3, we divided all protocols into five groups according to the applied next-hop selection mechanism. Finally, we briefly described the operation of a representative sensor routing protocol from each of these groups.

5 Acknowledgements

The work described in this paper is based on results of IST FP6 STREP UbiSec&Sens (`http://www.ist-ubisecsens.org`). UbiSec&Sens receives research funding from the European Communitys Sixth Framework Programme. Apart from this, the European Commission has no responsibility for the content of this paper. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The work presented in this paper has also been partially supported by the Hungarian Scientific Research Fund (contract number OTKA T046664). The first author has been further supported by the HSN Lab.

References

[1] J. N. Al-Karaki and A. E. Kamal. Routing techniques in wireless sensor networks: a survey. In *IEEE Wireless Communications*, Volume 11, pp. 6-28, 2004.

[2] W. Heinzelman, A. Chandrakasan and H. Balakrishnan. Energy-Effcient Communication Protocol for Wireless Microsensor Networks. In *Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00)*, January 2000.

[3] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed Diffusion: a scalable and robust communication paradigm for sensor networks. In *Proceedings of ACM MobiCom'00*, Boston, MA, 2000, pp. 56-67.

[4] A. Manjeshwar and D. P. Agarwal. TEEN: a routing protocol for enhanced effeciency in wireless sensor networks. In *1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing*, April 2001.

[5] A. Manjeshwar and D. P. Agarwal. APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. *Parallel and Distributed Processing Symposium*, Proceedings International, IPDPS 2002, pp. 195-202.

[6] D. Braginsky and D. Estrin. Rumor Routing Algorithm for Sensor Networks. In *Proceedings of the First Workshop on Sensor Networks and Applications (WSNA)*, Atlanta, GA, October 2002.

[7] M. Chu, H. Haussecker, and F. Zhao. Scalable Information-Driven Sensor Querying and Routing for ad hoc Heterogeneous Sensor Networks. In the *International Journal of High Performance Computing Applications*, Vol. 16, No. 3, August 2002.

[8] S. Lindsey, C. Raghavendra. PEGASIS: Power-Effcient Gathering in Sensor Information Systems. In *IEEE Aerospace Conference Proceedings*, 2002, Vol. 3, 9-16 pp. 1125-1130.

[9] F. Ye, A. Chen, S. Liu, L. Zhang. A scalable solution to minimum cost forwarding in large sensor networks. In *Proceedings of the tenth International Conference on Computer Communications and Networks (ICCCN)*, pp. 304-309, 2001.

[10] F. Ye, H. Luo, J. Cheng, S. Lu, L. Zhang. A Two-tier data dissemination model for large-scale wireless sensor networks. In *Proceedings of ACM/IEEE MOBICOM*, 2002.

[11] V. Rodoplu and T. H. Meng. Minimum Energy Mobile Wireless Networks. In *IEEE Journal Selected Areas in Communications*, vol. 17, no. 8, Aug. 1999.

[12] Q. Li and J. Aslam and D. Rus. Hierarchical Power-aware Routing in Sensor Networks. In *Proceedings of the DIMACS Workshop on Pervasive Networking*, May, 2001.

[13] Q. Fang, F. Zhao, and L. Guibas. Lightweight Sensing and Communication Protocols for Target Enumeration and Aggregation. In *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking and computing (MOBIHOC)*, 2003, pp. 165-176.

[14] R. C. Shah and J. Rabaey. Energy Aware Routing for Low Energy Ad Hoc Sensor Networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*, March 17-21, 2002, Orlando, FL.

[15] N. Sadagopan et al. The ACQUIRE mechanism for efficient querying in sensor networks. In *Proceedings of the First International Workshop on Sensor Network Protocol and Applications*, Anchorage, Alaska, May 2003.

[16] Y. Yu, D. Estrin, and R. Govindan. 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.

[17] B. Karp and H. T. Kung. GPSR: Greedy perimeter stateless routing for wireless sensor networks. In *Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom '00)*, Boston, MA, August 2000.

[18] T. He et al. SPEED: A stateless protocol for real-time communication in sensor networks. In *Proceedings of International Conference on Distributed Computing Systems*, Providence, RI, May 2003.

[19] I. Stojmenovic and X. Lin. GEDIR: Loop-Free Location Based Routing in Wireless Networks. In *International Conference on Parallel and Distributed Computing and Systems*, Boston, MA, USA, Nov. 3-6, 1999.

[20] F. Kuhn, R. Wattenhofer, A. Zollinger. Worst-Case optimal and average-case efficient geometric ad-hoc routing. In *Proceedings of the 4th ACM International Conference on Mobile Computing and Networking*, Pages: 267-278, 2003.

[21] C. Schurgers and M.B. Srivastava. Energy efficient routing in wireless sensor networks. In *MILCOM Proceedings on Communications for Network-Centric Operations: Creating the Information Force*, McLean, VA, 2001.

[22] B. Leong, S. Mitra, and B. Liskov. Path vector face routing: Geographic routing with local face information. In *Proceedings of ICNP 2005*, pages 147-158, November 2005.

[23] B. Leong, B. Liskov, and R. Morris. Geographic Routing Without Planarization. In *Proceedings of NSDI 2006*, pages 339-352, May 2006.