Research Article

Reliable, Real-Time Routing in Wireless Sensor and Actuator Networks

Martin Krogmann,1 Mike Heidrich,1 Daniel Bichler,2 Daniel Barisic,2 and Guido Stromberg2

1 Fraunhofer Institute for Communication Systems ESK, 80686 Munich, Germany
2 Infineon Technologies, Munich, Germany

Correspondence should be addressed to Martin Krogmann, martin.krogmann@esk.fraunhofer.de

Received 26 November 2010; Accepted 3 January 2011


Copyright © 2011 Martin Krogmann et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a novel Reliable, Real-time Routing protocol (3R) based on multipath routing for highly time-constrained Wireless Sensor and Actuator Networks (WSANs). The protocol consists of a newly designed routing metric and a routing algorithm utilizing this metric. Our routing metric enables strong Quality-of-Service (QoS) support based on parallel transmissions which significantly reduces transmission delays in WSANs. A routing algorithm utilizing this metric is presented based on Dijkstra’s shortest path. A novel Medium Access Control (MAC) layer that supports dynamical adjustments of retransmission limits, reduces traffic overhead in multipath routing protocols. Thorough simulations have been performed to evaluate the routing protocol, and the results show that real-time performance of WSANs can be vastly improved.

1. Introduction

Wireless Sensor Networks (WSNs) have moved into real-world applications, and their extension to wireless sensor and actuator networks (WSANs) is in progress. Traditional application areas for WSNs include building automation, environmental monitoring, and habitat monitoring [1], and one major challenge is to cope with the energy limitations of the battery-powered sensor nodes. Real-time aspects are not exhaustively considered so far, since such WSN applications hardly need real-time communication. With the introduction of WSANs, applications are emerging that would significantly benefit from the possibility of real-time communication [2].

In the industrial sector many applications for WSANs can be found that have hard real-time requirements such as open- or closed-loop controlled systems [3]. These applications use the WSAN for measuring and processing data and for controlling the system if necessary. In most closed-loop controlled systems, the control functionality requires hard real-time communication, while in open-loop controlled systems a human is in the loop and the timing requirements are less stringent. The lack of reliable, real-time protocols for WSANs state a big problem for enabling WSANs in such control systems. One reason for this lack of suitable protocols might be that energy awareness and real-time performance are often conflicting objectives [4].

Advances in energy-harvesting technologies have the potential to mitigate the energy limitation of wireless sensor nodes. Energy-harvesting techniques steadily improve, and a variety of self-sustaining products already exist that do not require battery changes anymore [5, 6]. If this trend continues, the energy limitations of WSANs will become less demanding, especially in some special purpose deployments where powerful energy-harvesting solutions exist such as thermoelectric energy harvesting in industrial environments with significant temperature differences. Overcoming the energy limitations is one key to enable more powerful real-time performance in WSANs, since novel ideas can come up that might require more energy but increase the reliability and real-time behavior of communication protocols.

In our paper we present a novel routing protocol increasing the reliability and real-time performance at the cost of a higher energy consumption. The idea is to send copies of a packet in parallel via different routes to its destination and thus reduce the probability of sequential packet losses.
on a single route, which has been identified as a serious problem [7]. The calculation of optimal routes utilizes information about the deadline of a packet, its requested reaching probability that defines the ratio of packets reaching their destination in time, and the link quality. At the same time, the overhead for parallel transmissions is kept low by introducing a new MAC layer with support for dynamic limitation of retransmissions.

2. Related Work

RAP, a real-time communication architecture for WSNs [8], was one of the first real-time protocols for WSNs. It consists of a bundle of different communication layers. The central layer, concerning the real-time capability of this architecture, is the Carrier Sense Multiple-Access (CSMA) Media Access Control (MAC) layer, which prioritizes traffic by adjusting inter-frame spaces according to a packet's priority in the wait times and back-off times. The priority of a packet is calculated by a velocity monotonic scheduling algorithm that determines the priority according to the distance between the destination and the current node as well as the packet's deadline.

The SPEED [9] protocol also uses velocity as a criterion for choosing a route. SPEED measures velocities of links and considers only fast links that have the required velocity as forwarding paths. No adjustments of the underlying MAC layer have been done. Therefore, velocity depends on link quality and network load.

A routing metric reflecting the velocity in nongeographical routing protocols is the number of expected transmissions. Using the number of expected transmissions as a routing metric has been evaluated in several papers, for example, as the Expected Transmission count (ETX) metric in [10] or as the Minimum Transmission (MT) metric in [11]. The core of the related routing algorithm is to choose the route that causes the minimal number of transmissions including MAC layer retransmissions. This approach has been proven to have remarkably low energy consumption and a very high throughput. In [12] it was shown that the energy consumption can be further reduced with a more accurate estimation of the number of transmissions.

The Multipath Multi-SPEED protocol (MMSPEED) [13] is an extended version of SPEED. Besides several improvements to the velocity-based real-time capability of SPEED, MMSPEED adds support for reliable data transmission. Similar to the approach in ReInForM [14], the Packet Reception Rate (PRR) is used to estimate the reaching probability of a packet and, if necessary, to start parallel transmissions to increase the reaching probability. Using parallel forwarding paths usually results in systematic congestion and high energy consumption, both of which are major problems with these kinds of reliability enhancements.

In [15, 16], a scheduling algorithm similar to the well-known Earliest Deadline First (EDF) algorithm is introduced to schedule time slots for sending packets on a Time Division Multiple Access (TDMA) MAC layer. The key idea is that if nodes are placed in a cell structure topology, all nodes inside one cell can synchronize their transmit schedule implicitly because all nodes are able to receive the same messages and thus have complete information of the cell's transmit schedule. Enhanced router nodes are responsible for forwarding data.

Emerging standards such as WirelessHART [17, 18] specify MAC and routing layer protocols and are designed especially for harsh industrial environments. They contain a framework that can be used for implementing routing algorithms. However, these standards do not provide a complete routing protocol for WSNs because one of the most important parts in mesh networks is missing, namely, the routing metric. The routing metric is especially important if reliable, real-time communication is demanded. In our work, we propose a routing metric that enables reliable, soft real-time communication in wireless mesh networks.

We partially reuse the idea of expected transmissions, but extend this metric with more elaborate calculations regarding the reaching probability and deadline of a packet. Multipath transmissions enhance the reliability of a transmission if necessary. Similar to MMSPEED, our routing metric also uses PRR estimations for calculating the necessary number of forwarding paths to ensure a certain reaching probability. The difference is that we do not consider the reaching probability and transmission latencies apart from each other. Instead, we correlate these factors to yield more accurate estimations. Further, we also introduce a novel MAC layer as an approach for effectively reducing the energy consumption and network load of the routing metric, based on parallel forwarding of packets by adjusting dynamically the maximum number of retransmissions on the MAC layer.

3. Routing Protocol Architecture

The architecture of our 3R routing protocol partially integrates functionality of the transport layer in the ISO/OSI model. The routing protocol consists of the proposed routing algorithm that is responsible for calculating optimal paths according to the proposed routing metric. The routing metric ranks alternative paths and partially integrates transport layer functionality since it considers possible packet retransmissions and the reaching probability of a packet. There is no need for any additional reliability mechanism on the transport layer. The proposed MAC layer is tightly coupled with the routing metric to reduce the energy consumption and network traffic. In the following we present details about the protocol architecture. First, we formalize our network model, and then we give detailed information about our routing metric, an important part of the protocol. After that we present an algorithm utilizing this metric and a traffic-reducing MAC layer exploiting the characteristics of our routing protocol.

3.1. Network Model. In order to ensure a common understanding of our network model, we start with a formal description. We represent the network by a bidirectional, weighted graph $G(V,E)$ with the weighting function $\lambda$. Here, node $v_i \in V$ represents a node in the network, edge $e_i \in E$
represents a wireless link between two adjacent nodes, and \( \lambda(e_i) \) denotes the link quality of an edge \( e_i \). The link quality is equivalent to the complement of the Packet Error Rate (PER), namely, \((1 - \text{PER})\). A route is a simple, connected path in the graph. The length of a route \( r \) is the number of edges on the path denoted by \( \kappa_r \), and the route contains the edges from \( e_0 \) to \( e_{r-1} \), which describes a path from the transmitting node to the destination. As in [10, 12], we assume that packet losses are independent and identically distributed. Each packet that is sent inside the network has a fixed deadline \( d \) and a requested reaching probability \( r_p \).

3.2. Routing Metric. Our routing metric is based on parallel multi-path transmissions as a technique to reduce transmission latencies, based on the awareness that sequential retransmissions cost precious time. Our approach to reduce transmission latencies below those latencies achieved with state-of-the-art routing metrics is to send packets at the same time via several disjoint routes and thus have immediate retransmissions on parallel routes, which takes no additional time.

3.2.1. Reaching Probability and Transmission Time. In our network model, each packet has an assigned reaching probability \( r_p \) that must be fulfilled. We assume that the transmission of a packet via a route is a Bernoulli process in which the success of each transmission is independent of the previous one. The estimated maximum reaching probability of a packet that is sent via route \( r \) is calculated with the following equation:

\[
\sigma_r(m) = \prod_{e_i \in r} (1 - (1 - \lambda(e_i))^m).
\]  

Here, \( m \) is the maximum number of retransmissions per hop on the MAC layer. Assuming low bandwidth radios in WSNs and no complex packet processing, the total transmission time of a packet is usually dominated by the number of transmissions. Therefore, latencies resulting from packet processing inside the network stack can be neglected and we assume that the transmission time is \( T_r = T_0 b_r \), where \( T_0 \) is the average transmission time of the packet and \( b_r \) the number of transmissions including retransmissions on the current route \( r \). The worst-case transmission time is therefore \( T_r = T_0 \kappa_r m \) for the case in which the maximum retransmission limit is necessary for each hop.

Considering that each packet has reliability requirements as well as timing constraints, we introduce a maximum allowed number of transmissions which is defined as \( b_{\text{max}} = \lfloor d/T_0 \rfloor \). If not more than this \( b_{\text{max}} \) transmissions are used on a single route, then the packet’s deadline \( d \) will be satisfied with the probability stated in the packet’s reliability requirements \( r_p \).

So, we have two constraints for each transmission, that is, \( T_r < d \) for the time domain and \( \sigma_r \geq r_p \) for the reliability domain. The reliability domain will be handled in Section 3.2.2. For the time domain, we will estimate the expected worst-case transmission regarding the reaching probability \( r_p \) since \( T_r \) is dependent on the number of transmission. In our routing metric, \( m \) is not considered as a general limit of retransmissions in the MAC layer but as vector \((m_0, \ldots, m_{\kappa_r - 1})\) of expected worst-case number of transmissions per hop on a certain route concerning a packet’s requested reliability \( r_p \). Each \( m_i \) is calculated according to the next hop’s link quality and the packet’s requested reaching probability by

\[
m_i(r) = \log_{(1 - \lambda(e_i))}(1 - r_p).
\]  

If we find a vector with \( |m_1| \leq b_{\text{max}} \), then the packet arrives with the requested reaching probability \( r_p \) before its deadline \( d \) at the destination. If we are not able to find a route that fulfills this requirement, we relax the reliability constraints to get more flexibility for searching a suitable route. To compensate for this possible loss in reliability, we use parallel multi-path transmissions.

3.2.2. Parallel Multipath Transmission. Parallel multi-path transmissions increase the reliability of a transmission [13, 14]. If we assume that \( R' \) contains all used routes, then the total reaching probability \( \text{trp} \) over multiple paths is calculated as

\[
\text{trp} = 1 - \prod_{r \in R'} (1 - \sigma_r(m)).
\]  

As long as \( \text{trp} < r_p \) and additional routes are available, new routes are added. This mechanism compensates for the reliability relaxation in our routing metric which is needed for achieving shorter latencies.

3.2.3. Disjoint Routes. An inherent problem of parallel multi-path transmissions is self-created congestion that results from sending several copies of a packet at the same time to the same destination. Consequently, the formula in (3) for the total reaching probability is only correct if we ensure that transmissions via different routes are random events that are independent from each other.

Our approach is to use disjoint, noninterfering routes. Disjoint routes have no common nodes except the destination node. Alternatively, edge-disjoint routes could be used, but this would not be sufficient since copies of a packet might be sent over the same node. If this happens, the later copies might just be queued in the transmitting buffer and thus delayed. This delay would be equal to a delay caused by a normal sequential retransmission. A major reason for introducing parallel multi-path transmission is to reduce the probability of sequential retransmissions; thus disjoint routes are an important requirement.

3.3. Routing Algorithm. The routing metric requires choosing noninterfering disjoint routes that satisfy the timing and reliability requirements of a packet. This problem is very similar to the k-shortest paths problem, which already has been extensively researched [19]. Our solution in order to find disjoint routes is to apply Dijkstra’s shortest path [20] repetitively and remove used edges. Although this algorithm will not be able to find
maximal disjoint paths, its results are sufficient for our purposes since it finds nearly optimal solutions [21]. The channel scheduling problem is solved with a sequential vertex coloring algorithm [22].

We use a centralized design and implement the routing algorithm inside a central network coordinator that has complete knowledge about the network topology and manages the routing tables of each single node. The proposed routing metric requires complete knowledge about the network, so no distributed routing algorithms such as AOMDV [23] or MDSDV [24] can be used here without modifications. The route discovery uses counter-based flooding and the network coordinator considers the network as a graph as described in Section 3.1.

3.3.1. Channel Scheduling. We obtain noninterfering traffic by assigning different channels to nodes within signal range. This channel assignment is a casual coloring problem. We use a sequential vertex coloring algorithm [22] to solve this problem. During the initialization phase, each node will be assigned a channel that it will listen on during idle mode. If a neighboring node sends a packet, it will use the destination node’s receive channel. No other node in its neighborhood is allowed to use the same channel. Since routes are disjoint and no node inside the network will be used twice for forwarding a copy of the same packet, different routes will not affect each other.

3.3.2. Route Creation. For the creation of disjoint routes according to our routing metric, we reuse Dijkstra’s shortest-path algorithm. In Algorithm 1 we see the pseudo code of the algorithm. The following steps are executed as long as the total reaching probability of a packet trp is smaller than the packet’s requested reaching probability rp. At the beginning, we initialize our graph with the help of Dijkstra’s shortest-path algorithm. That means that we define a distance function that the algorithm will use for calculating the weight of a link, and as a result each node will be assigned its minimum weight. As a distance function we use (2) whose input is the requested path reliability, here denoted as tmpProb. The result of the distance function is the worst-case number of estimated transmissions needed for achieving the desired reliability, which will be considered as the weight of a link. Thus, after running Dijkstra’s algorithm, each node inside the network is weighted with the number of estimated worst-case transmissions regarding the requested reliability tmpProb. According to Section 3.2.1, timing constraints can be expressed as the number of transmissions, so here the weight of a node is equivalent to the worst-case transmission time regarding the packet’s reaching probability rp. If a route can meet a packet’s timing constraints, it will be chosen. After adding a new route, the total reaching probability must be adjusted using (3). Since we have to ensure that all routes are disjoint, all used nodes of the new route are removed from the network graph and the shortest-path algorithm is run again. If no routes can be found that can meet the timing constraints, the reliability constraints are relaxed and we continue from the start. If rp cannot be reached with the available routes in the network graph given the restriction of the timing, no route will be suggested and packets will not be sent.

Algorithm 1: 3R routing algorithm.

```plaintext
SET trp to zero
SET tmpProb to rp
WHILE trp < rp
    CALL Dijkstra with tmpProb on current graph
    WHILE routes satisfying timing constraints exist
        add new route to the packet
        increase trp by reaching probability of new route
    remove route’s nodes from graph
    IF trp ≥ rp
        RETURN route has been found
    ENDIF
    CALL Dijkstra with tmpProb on current graph
ENDWHILE
IF tmpProb < 0
    RETURN no route has been found
ENDIF
ENDWHILE
```

3.4. Energy Saving MAC Layer for Time-Constrained Multipath Routing Protocols. Multi-path routing metrics based on parallel transmissions of packets such as ours also have the inherent problem of creating considerable transmission overhead. In this section, we present our novel MAC layer mitigating this overhead.

The key feature is a mechanism exploiting the reaching probability requirements of each packet. In Section 3.2, we have introduced the vector m with the number of expected worst-case transmissions on a hop for ensuring a certain reaching probability. If the requested reaching probability for a certain route is quite low, then too many retransmissions might take place on a single hop. In this case, the resulting reaching probability would be unnecessarily higher than the requested one. If we artificially restrict this number of retransmission by adjusting the retransmission limit on the MAC layer, we will save some transmissions caused by an unwanted increase of the reaching probability. The effect of this mechanism will be evaluated in the next section. Furthermore, we have implemented well-known features in this MAC layer, such as discarding duplicate copies of a packet and dropping of packets that have already missed their deadlines.

4. Evaluation

First, we briefly introduce our simulation environment. After that we start evaluating the timing behavior of the different protocols by examining average- and worst-case transmission latencies. Since multi-path protocols suffer from a large overhead, the energy consumption will also be considered. The last important benchmark is a reliability analysis.
4.1. Simulation Environment. For the evaluation, we have reused a simplified 802.11-like CSMA MAC layer from the Mobility Framework [25], which postulates multichannel transceiver hardware. The channel model assumes that channels are strictly orthogonal to each other so that interference among channels can be excluded. Our network consists of 36 nodes, arranged in a grid structure, thus avoiding unintended, random topology-related effects on performance. We send short packets from the lower right-side node to the upper left-side node. Traffic is Poisson distributed with $\lambda = 1$ s. We compare two versions of our 3R protocol to a shortest-path routing algorithm using the well-known ETX metric. One version, namely, 3R, uses a casual CSMA MAC layer similar to the one which ETX uses. The other version, 3R with dynamic Limitation of Transmissions (3R LT), uses our own proposed MAC layer (see Section 3.4).

The link quality depends exponentially on the distance, because of the path-loss model implemented in Mobility Framework’s physical layer. For our grid-structured network, that means that we have two different link quality classes, one for diagonal links and one for vertical or horizontal links. We vary the link quality by changing the transmit power, while all other parameters remain unchanged. Thus, diagonal links are influenced the most, while horizontal and vertical links remain very high, virtually 100%. The requested reaching probability of packets, $rp$, is always 99%.

4.2. Transmission Latencies Analysis. Figures 1 and 2 show the results of our three tested protocols concerning their average transmission delay and the maximum transmission delay, respectively. For this evaluation, the link quality was varied between 90% and 70%, which represents quite common wireless sensor network characteristics. During these tests, the 3R protocol variants were configured for best real-time behavior. We made 5 simulation runs for each link quality level. A total of 2000 packets were sent in each run.

Figure 1 shows the arithmetic mean of the test run results. The 3R protocols generally perform better than ETX. There is virtually no difference between the two 3R derivatives. Figure 2 shows the results for the worst-case transmission time. Here, the arithmetic means of the results are represented as error bars with their 95% confidence intervals since our measurements were scattered for this benchmark. It is clear that the 3R protocols also perform much better than ETX. For low link qualities, 3R LT is slightly better than 3R. The reason for this result is the MAC layer of 3R LT, which automatically discards packets that already have missed their deadlines. This mechanism leads to a slightly higher packet loss rate, but as we will see in the reliability analysis in Section 4.4, the PRR is in its predefined range of 99%.

4.3. Energy Consumption Analysis. For multi-path protocols such as 3R, the energy consumption is large since packets are sent over several routes at the same time. Thus, we have introduced our own MAC layer. The energy consumption is measured by calculating the average number of transmitted packets during a complete transmission of a packet from the transmitting node to the sink. Figure 3 clearly shows that ETX has always a far better energy consumption. The reason is that ETX always chooses the path with the lowest average number of transmissions, which yields very low energy consumption. By comparison, the 3R protocols choose the path with the lowest worst-case number of transmission concerning the requested reaching probability $rp$ of the packet. The different paths explain the high energy consumption for a rather high deadline of 0.012 s; the 3R protocols choose a more reliable route twice as long as the ETX route. Thus, the energy consumption is also nearly twice as large. As the timing constraints tighten, shorter paths are also used in 3R although they are not able to guarantee the reliability constraints. Therefore, additional forwarding paths are chosen and the energy consumption jumps up to
22 packets for a deadline of 0.0118 s. Note that because of the utilization of shorter paths, the overall energy consumption increases up to 22 packets and not to 40 which one might expect. With a deadline of 0.0102 s, the energy consumption jumps to approximately 34 packets, which is caused by needing one additional route. In this scenario, the limitation of retransmission in 3R L T is able to save about 10% energy by reducing the network load if multiple paths are used.

4.4. Reliability Analysis. For the reliability analysis, we examine how many transmissions are successful, that is, packets which have reached their destinations within their deadlines, for different timing constraints and network loads. Figure 4 shows the percentage of successful transmissions. The 3R protocols perform significantly better than ETX if packet deadlines are shortened. At 0.0085 s, ETX is only able to deliver 90% of the packets within their deadline while the 3R protocols still achieve 99%. For deadline below this border, 3R protocols will refuse to deliver the packet since the requested reaching probability rp of 99% cannot be met. Figure 5 shows the relationship between the rate of successful transmissions and the network load. Therefore, we adjust our traffic model to send Poisson distributed packets with $\lambda = 0.02$ s. As the deadline for the packets, we have chosen 0.009 s. We call a node that continuously sends data to the sink a “flow”. All protocols work fine for just one flow, but if the number of flows and thus the network traffic increase, the performance of all protocols decreases dramatically. For many flows, 3R LT performs slightly better than 3R because the traffic is reduced by its MAC layer.

5. Discussion

The results show that the routing metric can decrease latencies and increase the reliability at the cost of a higher energy consumption. Since energy consumption states an essential problem in WSNs, these results have to be discussed.

In general WSNs, devices are battery driven and the main objectives are energy-efficiency and prolonging the system lifetime. However, in these general WSNs there is often no need for real-time communication with constrained deadlines. In such applications, routing metrics such as ETX perform very well. If the applications require real-time communication, such as closed-loop controlled systems, the proposed routing metric in this paper offers a better real-time performance and reliability at the cost of a higher energy consumption. So, for specific applications with actual real-time requirements, the higher energy consumption might be a price that has to be paid. If powerful energy-harvesting solutions are available, this option must be considered.

The performance of energy harvesters differs significantly depending on several factors, for example, the
environment and the used technology. Unfortunately, not all energy-harvesting solutions are powerful enough to cope with the higher energy demands. However, energy harvesting is an active research area that improves steadily. Currently, there already exist energy harvesters that might be able to provide enough energy if they are deployed and sized properly and if the environment provides sufficient energy [6, 26]. Of course, having these requirements limits the application scenarios, since the deployment efforts and costs increase. However, if these deployment requirements can be fulfilled, for example, in some static deployments in a machinery hall, then the energy constraints can be relaxed.

Alternatives in the literature, such as MMSPEED, also increase the reliability of the network by increasing the energy consumption. In contrast to the proposed routing metric, MMSPEED lacks the possibility to trade off latency, reliability, and energy consumption and it bases on geographical information which is often not available in industrial (indoor) environments. The high energy consumption for the improved reliability remains a disadvantage of the protocol.

In the future, energy harvesting might be a key feature of WSANs, since it has the potential to offer more energy than batteries in some cases. As a result, gaining energy by using energy harvesters might not only prolong the lifetime of wireless networks, it also has the potential to increase the performance of WSANs in terms of reliability and latency.

6. Conclusion

In this paper we have proposed a routing protocol that enables reliable, real-time routing in industrial WSNs and WSANs. A key feature of our study is a routing metric that balances timing constraints against reliability requirements, that is, reliability constraints are relaxed in order to find a sufficiently fast route while increasing reliability by using parallel multi-path transmissions. The drawback of this routing metric, the increased energy consumption and network load, is mitigated by the proposed MAC layer. Simulations have shown that transmission latencies have been significantly reduced and the routing protocol assures a reliable packet transmission. However, the proposed routing protocol cannot be seen as a reliable, real-time routing protocol for all kinds of WSN applications, but it contributes to WSAN deployments in which real-time communication is demanded and powerful energy-harvesting solutions exist. In our future work we intend to examine the impact of link quality estimation errors [27] and to consider varying energy levels of nodes caused by nonuniform power supplies utilizing energy-harvesting solutions.

References


