Proactive Highly Ambulatory Sensor Routing (PHASeR) protocol for mobile wireless sensor networks

Article  (Published Version)


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/59774/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher’s version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

http://sro.sussex.ac.uk
Proactive Highly Ambulatory Sensor Routing (PHASeR) protocol for mobile wireless sensor networks

T. Hayes*, F.H. Ali

Communications Research Group, Department of Engineering and Design, University of Sussex, Brighton, East Sussex, BN1 9QT, UK

A R T I C L E   I N F O

Article history:
Received 20 August 2014
Received in revised form 9 February 2015
Accepted 2 April 2015
Available online 20 April 2015

Keywords:
Mobile environments
Routing protocols
Sensor networks
Wireless

A B S T R A C T

This paper presents a novel multihop routing protocol for mobile wireless sensor networks called PHASeR (Proactive Highly Ambulatory Sensor Routing). The proposed protocol uses a simple hop-count metric to enable the dynamic and robust routing of data towards the sink in mobile environments. It is motivated by the application of radiation mapping by unmanned vehicles, which requires the reliable and timely delivery of regular measurements to the sink. PHASeR maintains a gradient metric in mobile environments by using a global TDMA MAC layer. It also uses the technique of blind forwarding to pass messages through the network in a multipath manner. PHASeR is analysed mathematically based on packet delivery ratio, average packet delay, throughput and overhead. It is then simulated with varying mobility, scalability and traffic loads. The protocol gives good results over all measures, which suggests that it may also be suitable for a wider array of emerging applications.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

WIRELESS sensor networks (WSNs) are traditionally made up of a number of small nodes with the ability to communicate wirelessly [1]. The aim of the system is to gather sensory data, which is usually accumulated at the sink. In static WSNs, once the nodes have been deployed, they rarely move, which simplifies the task of routing data to the sink. However in some situations the sensor nodes maybe required to be dynamic, creating a mobile wireless sensor network (MWSN). One advantage of this is in the increased coverage that the network can provide [2]. MWSNs have no fixed topology, which makes routing a more difficult task due to the frequent link breaks that could cause a route to become unusable. Link breaks could occur from varying channel conditions or nodes travelling away from each other until they are out-of-range. MWSNs also suffer from the same problems as static WSNs, such as energy constraints, cost, bandwidth limitations and required self-configuration.

The number of applications requiring the use of a MWSN is growing with emerging technologies. Mobile nodes maybe required for moveable objects in smart environments, people who require constant health monitoring [3], research on wildlife [4] or surveillance performed by drones [5]. The target application for this work is in the area of environment monitoring [6–8], for example in an area that has been flooded with radiation [9]. An irradiated zone would be hazardous for people to enter, so it may be useful to map the levels of radiation to assess the extent of the contamination, locate safe areas and determine the point of origin. This may be done with a swarm of unmanned aerial vehicles (UAVs) equipped with radiation sensors, which could map the areas radiation levels and an image could be built up at the sink. The sink would be a
manned ground station with a power supply, where aggregation and analysis of the data will take place. Protocols designed for these types of emerging application will require high packet delivery reliability and low overhead to decrease latency even in highly mobile environments. The large quantity of applications, with varying characteristics, will often mean that one single protocol will not be suitable for every scenario. As such, the objective of Proactive Highly Ambulatory Sensor Routing (PHASer) is to provide robust and timely data routing in the challenging mobile environment of radiation mapping using UAVs. Although the protocol can be used in a variety of scenarios, the application of radiation mapping will be used to motivate the design choices of the protocol.

The next section will outline some of the existing literature designed for MWSNs in order to position the work. Section 3 will motivate the design choices and give an in depth description of how the proposed protocol works, Section 4 then presents mathematical analyses on various metrics. Section 5 gives details of the modelling and the simulation results are given in Section 6 before the paper is concluded in Section 7.

2. Literature review

In comparison to static sensor networks, there is limited work on MWSNs and even less so for routing in radiation mapping applications. For this reason, in this section, the existing literature will be reviewed from a more generic MWSN point of view and conclusions will be drawn as such.

Essentially, in terms of existing protocols, MWSNs are the overlap between WSNs and mobile ad-hoc networks (MANETs). Generally, WSNs only require data to flow one-way; from many sources to a single sink. Whereas, MANETs require that data must be able to flow both ways between any two nodes. This additional functionality often adds overhead, which is unnecessary in a MWSN. On the other hand WSN protocols are intended for static nodes and therefore cannot cope with the fast changing topology of MWSNs. So, in order for the demands of emerging applications in this area, new MWSN protocols are needed.

Routing protocols for MWSNs can be broadly split into two categories: hierarchical and flat. The hierarchical protocols assign roles to different nodes, whereas in flat protocols all nodes perform the same tasks.

Protocols designed for MANETs are normally flat; they are intended to be capable in mobile scenarios and are often split into two categories; proactive and on-demand. The rapidly changing topology of MWSNs can cause proactive protocols to flood the network with topology information so frequently that the amount of data delivered is severely reduced. Alternatively, the topology information may not be distributed often enough and a large number of packets may be lost. So, proactive MANET routing protocols, like OLSR (Optimised Link State Routing) [10], are often deemed unsuitable for MWSNs [11].

On-demand protocols suffer from similar effect to proactive protocols, in that the mobility of the network may warrant the discovery of a new route so frequently that the network becomes clogged up with control traffic, or the data will be lost by attempting to use an out dated route. However in low traffic scenarios this approach may still be feasible making MANET routing protocols such as DSR (Dynamic Source Routing) [12] possible candidates. However, it is the popular AODV (Ad-hoc On-demand Distance Vector) routing protocol [13], which is most commonly used in MWSN scenarios [14–16]. Even though in highly dynamic scenarios it is unable to react fast enough to the frequent topology changes [17]. An improvement on this is Ad-hoc On-demand Multipath Distance Vector (AOMDV) [18], which introduces a multipath element to the protocol.

For MWSNs Data Centric Braided Multipath (DCBM) [19] has been proposed. The protocol is query-driven so when the sink requires a certain piece of data a query is broadcast. The queries are flooded through the network in the same way as a route reply in AODV, such that nodes record the ID of the node from which they received the query. This allows each node to forward the data from the response of the query onwards. As it is likely that each node will receive multiple copies of the query the multipath element comes from the nodes storing the IDs of more than one node, similarly to AOMDV. DCBM reduces the overhead associated with route discovery by performing it in conjunction with data queries from the sink.

Another protocol, with similarities to AODV is AODV++ [20], which is based on the same route request-reply framework, but the choice of route is made based on link reliability, node energy levels and traffic rates. In this way the protocol attempts to prolong the network lifetime whilst trying to find the quickest, most reliable path to the sink.

Alternatively, Geographically Opportunistic Routing (GOR) [21] is a protocol designed for MWSNs and splits the network area into grids and nodes use GPS to determine which grid they are in. GOR eliminates the need to distribute topology information as the nodes forward data to a grid that is closer to the sink rather than a node. The sink must remain static at a known location, so that each grid can be given a priority based on its distance from the sink. GOR is opportunistic in the fact that a node will transmit to a specific grid based on the nodes transmission range, if no nodes in the intended grid hear the transmission then the data is forwarded through other nodes in a closer grid. Essentially this is a proactive protocol in which a node may determine a path for its data based on its GPS coordinates and a grid system rather than the distribution of topology information.

RRP (Robust cooperative Routing Protocol) [22] assumes a path has already been found and then cooperatively aids the transmission. It does this by enabling nodes that are not on the intended path, to relay overheard transmissions. This means that if a link is broken the packet can still be passed forward, making it able to handle frequent topology changes.

OR–RSSI (Opportunistic Routing–Received Signal Strength Indicator) [23] uses an opportunity probability and is based on Extremely Opportunistic Routing (Ex-OR) [24]. Opportunistic protocols are particularly applicable in MWSNs as they can exploit the transient connections that are created and destroyed by the mobile nodes.
Angle-based Dynamic Source Routing (ADSR) [25] uses the angle between potential forwarding neighbours and the sink to determine its next hop neighbour. This information is constructed from the sharing of location data and enables the protocol to ensure that packets are always being forwarded towards the sink.

The proposed mechanism in [26] allows nodes with data to transmit, to request location information from their neighbours. The paper suggests that disseminating location information through the use of beacons may cause nodes to forward data based on out-of-date information. So, by allowing location information to be retrieved on-demand, nodes will make better choices.

Also, the recently proposed protocol MACRO (Mobility Adaptive Cross-layer Routing) [27], is designed specifically for MWSNs. It utilises information such as average speed and RSSI data across multiple layers. Its route discovery method is similar to that of AODV, although it reduces the amount of flooding by restricting the subset of nodes able to forward the requests. Additionally, the most reliable routes are chosen for packets, based on link quality and the mobility of nodes.

Overall, in terms of reactive protocols, there is generally an initial delay caused by the discovery of routes. Though in comparison to proactive protocols, this is minimal compared to the much larger delays caused by the flooding of routing tables. The delay caused by flooding topology information is significantly higher in large mobile networks due to the higher number of nodes and the frequency of topology change, which often makes reactive protocols the preferred choice in MWSNs.

With regard to hierarchical protocols, generally nodes are split into groups and one node receives all the data from that group and passes it on to the sink. This type of routing is mainly influenced by WSN protocols such as LEACH (Low-Energy Adaptive Clustering Hierarchy) [28]. Network mobility causes hierarchical protocols to either generate large amounts of overhead to regroup nodes each time the topology significantly changes or suffer from packet losses from nodes that are out of range of their cluster head. So although cluster based protocols are generally energy efficient and very scalable, “they are not capable of handling node mobility” [11].

However Zone Based Routing (ZBR) [29] tries to rectify this by determining clusters geographically. So, each node always knows which cluster it is associated with by using GPS to determine which zone it is in. The zone heads are quickly appointed by broadcasting a mobility factor, then the least mobile node will become the zone head. The zone heads then aggregate the received data and route it to the sink.

Additionally, the WSN protocol LEACH has been adapted to MWSNs in [30], which presents LEACH-M (LEACH-Mobile). LEACH-M makes it easier for nodes to switch between clusters, which may occur frequently in mobile scenarios. This is done by providing a mechanism in which both the node and its cluster head can determine when they have become disconnected. When a node realises it has become disconnected from its cluster head, it then tries to associate with a new cluster. When a cluster head realises it has lost a node it then reallocates its timeslots accordingly.

LEACH-M was improved upon by LEACH-ME (LEACH-M Enhanced) [31], which determined cluster heads based on mobility. This meant the least mobile nodes would become cluster heads, making the network more stable.

Enhanced Cluster Based Routing Protocol for MWSN (ECBR-MWSN) [32] is based on the same principles of its predecessor CBR Mobile-WSN (Cluster Based Routing for Mobile WSNs) [33] and LEACH-M. It attempts to prolong the life time of the network by selecting the cluster head with the most residual energy, lowest mobility and closest to the sink. Further to this MBC (Mobility Based Clustering) [34] performs cluster head selection based on estimated connection time, residual energy, the cluster heads node degree and physical distance.

Location Aware Fault-tolerant Clustering Protocol for Mobile WSNs (LFCP-MWSN) [35] utilises GPS information at each node to define clusters. The cluster heads are selected using both mobility and energy metrics. It also reduces the number of control messages to reduce energy consumption and increase efficiency.

The literature suggests that the hierarchical structure in WSNs can reduce energy consumption and delay in very large networks. However, mobile nodes will constantly need to become associated with different clusters, which may cause large overhead. In comparison, the nodes in flat protocols can overcome this as they do not rely on any infrastructure, making flat protocols more suited to high mobility situations.

Flat protocols designed for static WSNs, such as Directed Diffusion [36], simply cannot handle the mobility of a MWSN. Though, Directed Diffusion introduces the idea of gradient based routing, in a query based format, which requires no infrastructure and a reduced amount of overhead. Directed Diffusion also allows more than one path to be used simultaneously, which improves the protocols reliability.

Similarly GBR (Gradient Based Routing) [37] initially floods the network to establish the gradient field so that nodes can forward data down the gradient to the sink. GBR uses a hop-count metric as the gradient and once the setup phase has been completed nodes can broadcast their gradient to their neighbours so that a forwarding node can be determined. Since the network is expected to be static, the gradient field does not need to be maintained.

Taken further, [38] suggests a protocol for a MANET which uses a similar concept but is designed for use in a mobile environment, using landmark nodes to define gradients. After an initial setup phase, overhead is kept low by only sharing local topology information to maintain the gradient field in a changing topology.

Alternatively, another method of keeping the gradient field up-to-date in a changing topology is using geographical information as the metric. GPSR (Greedy Perimeter Stateless Routing) [39] uses this technique by allowing each node to use location awareness to determine their distance from the sink. In this way, when a node want to forward a packet it will determine the next hop by listening to the gradient broadcasts of its neighbours. Additionally, [40] attempted to adapt GPSR for use with mobile sensors by using a gradient metric derived from the nodes speed, direction and the forwarding nodes distance to the sink.
DBO (Directed Broadcast with Overhearing) [41] is a gradient based protocol designed for static WSNs. Similarly to Directed Diffusion, it establishes the gradient field through an initial discovery phase and then proceeds to forward data through the network. However, contrary to Directed Diffusion, DBO uses a blind forwarding method of propagating data through the network. Blind forwarding is a technique of sending data down an established gradient through the broadcasting of data packets. Instead of a node choosing a single forwarding neighbour to transmit to, the nodes blindly broadcasts its packet, with its own gradient value, to all of its one hop neighbours. The neighbours receive the packet and each individually decide whether to forward the data on by comparing the gradient value from the packet with their own. This technique has low overhead and has been used in a variety of static WSN protocols [41–44], dating back to 1991 [45].

Although there are currently no protocols designed purposely for the application of radiation mapping with UAVs, the literature review suggests that a flat approach with low overhead would be preferable. One popular, readily available and commonly used protocol is AODV, which is also the basis of the ZigBee standard [46]. The ZigBee standard is one of the most commonly used protocols in sensor networks. AODV is also preferable to a wireless RS232 link, which was used in a testbed implementation of radiation mapping [47]. For these reasons the simulation results in Section 6 include AODV as a representative of current network deployments. Since AODV is a reactive protocol, results for the popular proactive OLSR routing protocol have also been included for completeness. In addition to this, as a comparison of the current state-of-the-art, the MWSN protocol MACRO [27] will also be included in the simulation results.

The proposed multihop routing protocol, PHASEr, applies the technique of blind forwarding in a MWSN, which increases the reliability of data delivery through its inherent use of multiple routes. This approach requires a gradient metric to be continuously maintained, which is problematic in a dynamic topology. The literature commonly uses either flooding or location awareness, however flooding creates large amounts of overhead and location determination schemes can often be inaccurate, power hungry and create the issue of the dead end problem. PHASEr uses a novel method of gradient maintenance in a mobile network, which requires the proactive sharing of only local topology information. This is facilitated by a global TDMA (time division multiple access) MAC (medium access control) layer and further reduces the amount of overhead, which in turn will decrease packet latency. PHASEr is also set apart by its use of encapsulation, which allows data from multiple nodes to be transmitted in the same packet in order to hande high volumes of traffic. It utilises node cooperation to create a robust multipath routing solution. As such, the contribution of this paper is a cross-layer routing protocol for MWSNs that can handle the constant flow of data from sensors in highly mobile situations.

3. PHASEr protocol

To motivate this work the target application of UAV radiation mapping was selected, which will require a small number of nodes to map the radiation levels in a contaminated area. Since the mission will use a fixed number of nodes and the network will be deployed for a limited amount of time, it can be assumed that the number of nodes will not change as long as each node has enough power to complete the mission. The authors in [9] give the operational time of a UAV carrying communications equipment and sensors, to be more than one hour, which is more than enough for a swarm of drones to map a medium sized area. The authors also state that the drones are autonomous, as such very few commands will need to be sent from the sink to the nodes. For this reason, this work will focus on the routing of data from the sensors to the ground station.

It is assumed that all of the nodes will be generating data periodically with fixed time intervals, which can be taken advantage of by letting each node periodically transmit in predefined slots. Given that there are a relatively small number of nodes, each one may be assigned a unique identification number, which will indicate the time slot in which the node should transmit. In this way, each node will transmit in turn and the time slots will loop cyclically, with a single cycle consisting of a single transmission from each node.

This method of using a fixed time slot assignment creates a collision free global TDMA MAC layer, which does not require any dynamic scheduling. Whilst this method will reduce overhead and computational efficiency it will require global synchronisation, for which a selection of available methods are surveyed in [48]. Additionally, since the sink is also allocated a time slot, this may be used to transmit a high powered network wide beacon to all sensor nodes. This beacon can provide synchronisation as well as network wide control commands. In the target application, the additional power cost for this is not a large concern due to the sink being a fixed ground station with a large power supply.

The length of each time slot will be predefined since the length of a cycle will be dictated by the sampling frequency of the sensor. Though a minimum bound for the time slot length will be dictated by the transmission time of a full packet, which will also need to be fixed.

In order to allow a node to forward data received from other nodes, a packet will have a variable capacity. The capacity of a packet will be expressed in frames, where the number of frames is the number of available data fields in a packet. The packet structure in Fig. 1 shows a two frame packet. The first frame, type 1, is always for the transmitting nodes data and protocol overhead and consists of four fields; the transmitting nodes ID, current hop-count and generated data, as well as an indication of which other nodes data is present.

The second frame shown in Fig. 1, type 2, is made up of only three fields, the data, priority and time stamp. Depending on the number of frames required, a node may repeat the structure of the type 2 frame as needed. For example, if a node has the data from four nodes to forward it will fill the type 1 frame with its own data and then repeat the type 2 frame structure four times; once for each additional piece of data to be transmitted. The forwarding node IDs field indicates which nodes
data is present in the packet, this is done using a single bit to represent each node. So, if the third and sixth bits were set high then it would indicate that there are two additional frames in the packet, containing data from nodes three and six respectively.

Since nodes are able to transmit data from multiple node in a single packet, future work may also look at using aggregation techniques to combine the received data and reduce the size of transmissions.

In order to calculate the minimum allowable time slot length it is necessary to put an upper limit on the number of frames a node may transmit. This is non-trivial since nodes closer to the sink will require more space, whereas nodes further from it will need less. However, the varying topology means that a node’s requirement will change over time. As such, if the maximum number of frames is too low, the bottleneck effect will cause data to be lost. In contrast, if the maximum packet size is too large then there may be wasted bandwidth. Although, a large slot length will reduce the frequency of transmissions and save energy.

In order to determine an appropriate maximum for the number of frames, a metric, \( \alpha \) is derived in Appendix A and given as:

\[
\alpha = 2^{1-n} \sum_{c=0}^{n-1} \left( \frac{n-1}{c} \right)
\]

where \( n \) is the total number of nodes in the network and \( F \) is the maximum frame capacity of a packet. The value, \( \alpha \), is the fraction of possible topologies that will not suffer from the bottleneck effect for a given \( n \) and \( F \).

For example, in a network of 5 nodes there are 1024 possible topologies. Allowing each node to forward its own data and two other nodes’ data, would give a frame capacity of 3. Out of the 1024 possible topologies there are 704 that would not suffer from the bottleneck effect. This means that there are 320 possible topologies in which the frame capacity of 3 would not be sufficient. This example gives \( \alpha \) as 0.6875, which is equal to \( \frac{704}{1024} \). Subsequently we could say that in a network of 5 nodes with a frame capacity of 3, 68.75% of topologies would suffer no losses due to the bottleneck effect. In this way (1) can be used to select a suitable value of \( F \) based on the number of nodes in the network and a given acceptable \( \alpha \) threshold.

In PHASER, each transmission is received by all of the transmitting nodes neighbours, this allows nodes to gather local topology information; namely the hop-counts of the nodes neighbours. Using this information a node will determine its own hop-count as one more than the lowest hop-count of its neighbours. So, if a node had three neighbours, with the hop-counts two, four and five, the node would set its own hop-count to three. Since a deterministic global TDMA scheme is in use, a node will hear a single transmission from each of its neighbours in every cycle. This means that the node can update its hop-count every cycle, which will enable the gradient metric to be maintained across the whole network without flooding. This greatly reduces the protocol overhead and is one of its main advantages.

PHASER uses the hop-count gradient to implement blind forwarding; so a transmission from a node is heard by all of its neighbours and it is the receiving nodes who independently decide whether they should forward any of the received data. In other words, when a node overhears a transmission it compares the hop-count of the transmitting nodes with its own. If the received hop-count is lower than its own, then the transmitting node is closer to the sink and the packet can be ignored. If the received hop-count is higher than its own, then the transmitting node is further from the sink, so the data is extracted from the packet and stored. If the received hop-count is equal to its own, then the transmitting node is the same distance from the sink, so the data is extracted from the packet, evaluated and either stored or dropped.

Each frame has a status, frames designated a priority status is known as priority frames, and frames that do not have priority and have not been dropped is known as diversity data. Frames generated by the transmitting node are automatically treated as priority, whereas forwarded frames have their priority indicated by a frame priority bit.

When a packet is received from a node further away from the sink, the received priorities do not change. However, priority frames received from a node the same distance from the sink get demoted to diversity status and diversity frames are dropped.

When a node is compiling a packet for transmission it will first fill available frames with priority data, then any empty frames will be filled with diversity data. This allows nodes to utilise more of the available packet capacity.

By using blind forwarding, the decision to forward a frame is made by the receiver and since packets are broadcast multiple nodes will receive the data. This inherently creates a multipath scenario, where the same data is forwarded along multiple routes to the sink. The use of route diversity improves reliability because if one path fails to deliver a piece of data, there is
likely to be another path that will succeed. In other words, multiple copies of data will be created in the network, which can create congestion. However the trade-off for this is in the increased probability of the data successfully being delivered.

The algorithm used to determine which pieces of data in the queue are to be encapsulated needs to be run before a transmission can be made. As it is a simple sort algorithm, the worst case time complexity is given as $O(n^2)$, this simplicity makes the protocol suitable for cheap, low energy processors.

In order to allow the protocol to keep memory requirements to a minimum and cope with high levels of traffic, superseded frames are dropped. In other words, if a node receives two pieces of data which originated from the same node, the older piece will be disregarded in favour of the newer data. Alternatively, if a node receives a piece of data that is older than a piece of data already in the queue, the received data will be considered out-of-date and be dropped. Although, it should be noted that, depending on the application, using this method of dropping old data is not always preferable. In these cases an appropriate queuing algorithm may be implemented instead.

As an overview, Fig. 2 shows the flow chart for a single node's operation at each timeslot and illustrates the simplicity of the protocol. The node begins by determining if the current time slot is its own dedicated time slot, if it is then it should compile a packet from the queued frames and its own data. Priority data is stored first and then any remaining room is filled with diversity data. The packet is then broadcast to any nodes within the transmission radius.

An example of PHASer operation is given in Fig. 3, which uses the five node example given before, where the frame capacity is limited to three. The five nodes, A, B, C, D and E have the hop counts 2, 2, 2, 1 and 1 respectively. The transmitted packets are described in frames; so, $F_1(A)$ is the 1st frame, which contains data from node A. The first part of the figure shows the first transmission from node A, which consists of only one frame with node A's own data inside. This is received by nodes B, C and D. The second part of the figure shows a transmission from node B, which contains both its own data and node A's data. This transmission is received by both nodes A and D, however node A has a greater hop count then node B, so the packet is not forwarded. The second part of the figure also shows node C's transmission, which contains its own data as well as node A's. This is received by nodes A, D and E. Since node A's hop count is greater than node C's, the data is not forwarded. The third part of the figure shows node D's transmission. At this point node D has received data from nodes A, B and C, however the packet only has enough space for its own data and two other frames. So, in this case, node C's data is dropped and node D forwards node A's data, node B's data and its own data. The third part of the figure also shows node E transmitting a full packet containing node A's data, node C's data and its own. This example shows how blind forwarding is used to pass data along multiple paths to create redundancy and how the frame capacity limits the number of duplicate packets so that the network does not become saturated.
Generally, if it is not the nodes dedicated time slot then it should listen for a packet. On the reception of a packet the node will extract the transmitted hop-count and update its own if necessary. Then the relevant data should be retrieved and stored for forwarding.

In PHASER each frame is treated individually, which means that different frames can take multiple different paths through the network. This route diversity provides robustness and has been shown to provide improved performance [49]. Also, by only sharing local topology information means that overhead is kept to a minimum and a gradient metric can be maintained in highly mobile environments.

4. Mathematical analyses

This section will present a mathematical analysis of PHASER, designed to characterise the protocols performance for varying parameters. The metrics analysed are average end-to-end delay, packet delivery ratio, throughput, overhead and energy consumption.

4.1. Average end-to-end delay

The first metric is average end-to-end delay, $D_{av}$, and is the average between a packet being created and being delivered to the sink. It is given by:

$$D_{av} = h \cdot T_q $$

where $h$ is the average number of hops and $T_q$ is the delay of a single hop. To determine the average number of hops, the analysis in [50] is used, which gives the average number of hops as:

$$h = \frac{d_{av}}{d_{hop}} = \frac{2 \cdot L}{3 \cdot r \cdot \cos \left( \frac{\pi}{2N_n} \right)}$$

where $d_{av}$ is the average distance between source and destination and $d_{hop}$ is the average distance of a single hop. $L$ is the length of one side of the area of the network, $r$ is the transmission radius of the nodes and $N_n$ is the expected number of neighbours to each node, as given by:

$$N_n = \left( \frac{\pi \cdot r^2}{L^2} \right) (n - 1)$$

where $n$ is the number of nodes in the network.

The average delay in a TDMA multihop based protocol depends greatly on the order of the allocated time slots of the forwarding nodes. The best case scenario is that the forwarding nodes are in sequence, which is shown in Fig. 4.

Fig. 4 shows the source of the data, S, transmit the data to the destination, D. The nodes are shown at each timeslot, where the transmitting node is highlighted, and the progress of the data can be seen by the arrows. The timeslot in which a node should transmit is given underneath the respective nodes. The expression to calculate the delay in this scenario is:

$$T_{q_{\text{min}}} = h \cdot \Delta$$

where $\Delta$ is the length of a single timeslot.

However if the nodes allocated time slots are in the reverse order then the delay will be considerably longer, as shown in Fig. 5.
The expression for this worst case scenario is given as:

$$T_{q_{\text{max}}} = ((h - 1) (n - 1) + 1) \cdot \Delta.$$  \hspace{1cm} (6)

Assuming that the time slots are uniformly distributed, then the average end-to-end delay is given by:

$$T_{q_{\text{av}}} = T_q = \left( \frac{n \cdot (h - 1)}{2} + 1 \right) \cdot \Delta.$$  \hspace{1cm} (7)

This equation shows that as the number of nodes in the network increases, so will the delay, which is unsurprising considering the use of a TDMA MAC layer. The increase in nodes will increase the $n$ term as well as the packet length, which will in turn increase the time slot length causing the delay times to increase quickly.

4.2. Packet delivery ratio

The second metric is PDR, which is the number of packets successfully received, $P_{\text{rx}}$, per number of packets transmitted, $P_{\text{tx}}$. Given as:

$$\text{PDR} = \frac{P_{\text{rx}}}{P_{\text{tx}}}.$$ \hspace{1cm} (8)

As the TDMA MAC scheme is contention free there will be no loss from collisions. However, nodes being disconnected from the network and path breaks may cause packet losses. The use of multipath routing creates redundancy in the network, however the dropping of out dated frames may cause data to be lost, the characteristics of which are also captured by this metric.
The average link lifetime, $t_{av}$, adapted from [51], is

$$t_{av} = \frac{d_{\text{link}}}{\bar{v}} = \frac{4 \cdot r}{\pi \cdot v_{\text{max}}}$$  \hspace{1cm} (9)

where $d_{\text{link}}$ is the link distance, $\bar{v}$ is the relative velocity between the transmitter and receiver and $v_{\text{max}}$ is the maximum speed that a node is capable of.

From this, the probability of a link breaking in the time between a packet being transmitted and its reception at the sink, $P_{\text{break}}$, can be determined as:

$$P_{\text{break}} = \frac{D_{\text{av}} \cdot \pi \cdot v_{\text{max}}}{4 \cdot r}$$ \hspace{1cm} (10)

and therefore the expected number of broken links, $L_{\text{broken}}$, can be determined by multiplying $P_{\text{break}}$ by the expected number of active links:

$$L_{\text{broken}} = P_{\text{break}} \cdot \left( \frac{n}{2} \right) \cdot \left( \frac{\pi \cdot r^2}{L^2} \right).$$ \hspace{1cm} (11)

Since the average number of hops is given as $h$, this also indicates the number of links that a packet must traverse before it reaches the sink. Assuming that a link break on the packets path will cause packet loss then the packet loss ratio (PLR) is given by:

$$PLR = L_{\text{broken}} \cdot \left( \frac{n}{2} \right) \cdot \left( \frac{\pi \cdot r^2}{L^2} \right).$$ \hspace{1cm} (12)

PDR can now simply be calculated as:

$$PDR = 1 - PLR.$$ \hspace{1cm} (13)

This expression can be simplified as follows:

$$PDR = 1 - \left[ \frac{D_{\text{av}} \cdot \pi \cdot v_{\text{max}}}{4 \cdot r} \cdot \left( \frac{n}{2} \right) \cdot \left( \frac{\pi \cdot r^2}{L^2} \right) \cdot \frac{h}{\left( \frac{n}{2} \right) \cdot \left( \frac{\pi \cdot r^2}{L^2} \right)} \right]$$

$$= 1 - \left[ \frac{D_{\text{av}} \cdot \pi \cdot v_{\text{max}} \cdot h}{4 \cdot r} \right].$$ \hspace{1cm} (14)

From this expression the effect of various parameters can be seen; increasing the number of nodes will increase the end-to-end delay causing lower PDR. Also, the faster the maximum speed of the nodes and the smaller the transmission radius, the lower the PDR. Increasing the size of the network will increase the average hop-count and also diminish the PDR.

4.3. Throughput

In this paper, throughput, $TP$, is defined as the number of data bits successfully delivered to the sink, per second, over the entire simulation time. So the expression is given as

$$TP = \frac{L_{\text{data}} \cdot N_p \cdot PDR}{T_t}$$ \hspace{1cm} (15)

where $N_p$ is the number of packets produced and $T_t$ is the total deployment time of the network. The expression describes how the number of lost packets decreases the throughput, as does the number of packets produced.

4.4. Overhead

Overhead, $OH$, is a major factor in designing routing protocols for mobile networks since too much can cause congestion, which will limit the throughput of data. There are generally two types of overhead; packet overhead and control overhead. Packet overhead is the ratio of non-data bits to data bits in a data packet. Control overhead is the ratio of bits in control packets to bits in data packets. Control packets are often used to negotiate channel access, discover routes or share topology information.

The overall overhead is characterised by the total number of bits transmitted per successfully delivered data bit:

$$OH = \frac{B_{\text{tx}}}{L_{\text{data}} \cdot N_p \cdot PDR}$$ \hspace{1cm} (16)

where $B_{\text{tx}}$ is the total number of bits transmitted.
To determine $B_{tx}$ the following expression is used,

$$B_{tx} = \left( (N_f - 1) \cdot N_p \cdot L_{F_2} \right) + \left( (n - 1) \cdot \frac{T_i}{\Delta n} \cdot L_{F_1} \right) + \left( \frac{T_i}{\Delta n} \cdot L_{f_1} \right)$$  \tag{17}

where $N_f$ is the number of forwarding neighbours, $L_{F_2}$ is the size of the type 2 frame, $L_{F_1}$ is the length of a type 1 frame and $L_{f_1}$ is the size of the sink frame. The sink frame is a reduced size type 1 frame. The first set of square brackets calculates the number of additional type 2 frames that contain data forwarded from other nodes. The second set of square brackets determines the number of type 1 frames transmitted, which is one per cycle for every sensor node. The last set of square brackets is used to evaluate the contribution of transmissions from the sink, which also transmits once every cycle.

### 4.5. Energy consumption

When analysing the energy consumption of the protocol, only the power used to transmit and receive messages is considered. This is because the transceiver has the largest energy cost compared with that of the processor. The other factors such as the compiler used, which may make code more or less efficient, will affect the processors energy consumption, as well as other tasks that need to be run. There are also other energy costs attributed to things like the sensors and other peripherals, the mobility platform and the battery type, which are hardware specific and difficult to account for. There is also the time in which the node is listening for a message, which, in PHASER, occurs for one symbol at the beginning of each timeslot. However, for comparison purposes this has been omitted, such that protocols which do not consider sleep cycles are not given an unfair disadvantage.

The energy consumption, $EC$, is characterised in terms of joules per second per node:

$$EC = \left( \frac{V_{batt}}{R_b} \right) \cdot \left( \frac{(I_{tx} \cdot B_{tx}) + (I_{rx} \cdot B_{rx})}{n \cdot T_i} \right)$$  \tag{18}

where $R_b$ is the bit rate of the transceiver, $V_{batt}$ is the voltage of the batteries, $I_{tx}$ and $I_{rx}$ are the current consumptions of the transceiver when transmitting and receiving respectively and $B_{tx}$ is the total number of bits received.

Since PHASER broadcasts packets to all neighbours, $B_{tx}$ is given as

$$B_{tx} = B_{tx} \cdot N_f.$$  \tag{19}

This expression requires knowledge of the hardware but, $V_{batt}, I_{tx}$ and $I_{rx}$ can be substituted for temporary values based on potential hardware for comparison purposes.

### 5. Modelling and simulation

The simulation was done with the popular OPNET Modeler \[52\], which provides a discrete event simulator used to evaluate various parameters of network performance. In order to focus on the merits of the routing protocol the nodes were designed to be able to communicate without error whilst within a certain transmission range. Also, any collisions that occur will cause all packets involved to be considered as corrupted and subsequently dropped. Essentially, the physical layer was assumed to be perfect and channel effects, such as fading and interference are not taken into account. This was done in order to isolate the performance of the proposed protocol. It should also be noted that the transmission rates given in this work, are the rates seen by the MAC layer and so any additional preambles and checksums are absorbed by the physical layer.

The simulations were performed with varying parameters of speed, data generation rate, number of nodes and size of network. However, the transmission radius was kept constant at 250 m, as was the transceiver transmission rate at 250 kbps. This is to emulate a low cost, low power transceiver such as the Memsic IRIS mote \[53\]. These motes were also used for the $V_{batt}, I_{tx}$ and $I_{rx}$ values, to model the energy consumption. Node mobility was controlled using the random waypoint mobility model \[12\], the pause time was set to zero and the speed is set by a uniform distribution between 0 m/s and a maximum. All nodes, including the sink, are mobile.

The base parameter values chosen were intended to model the application of radiation mapping, as described in Section 3. Whereby a group of UAVs are kept at a constant altitude and deployed in an irradiated area in order to report radiation levels to a sink, such that these levels can be plotted on to a two dimensional map. These parameters will be varied around the base values in order to test the protocols robustness and adaptability. The base value for maximum speed will be 25 m/s, which is slightly higher than the flight speed given in \[9\], but is close to both the cruising speed of a fixed wing UAV and the top speed of a rotary wing UAV. The base values in terms of scalability are set to 25 nodes in a 600 m by 600 m network area. This represents a medium size swarm of drones surveying an area such as a nuclear power station. The base data generation rate of 1 pk/s (packet per second) was taken from \[9\] and will give a reasonable resolution for building up a radiation map of the contaminated area. Every sensor node is considered a source and able to produce data at this rate.

The data length is a fixed parameter for these simulations at 32 bits, within which is contained the nodes spatial coordinates and the sampled sensor data. The longitudinal and latitudinal coordinates are allocated 12 bits each, leaving 8 bits for the radiation measurement. Since the swarm will be deployed in a defined area, a single geographic coordinate can be selected as the origin beforehand. As such, nodes can report their position relative to this point. Therefore the
maximum value of 4096 can be directly translated into 4.096 min. Longitudinally, this translates to a maximum distance of approximately 7.6 km with a resolution of 1.9 m. In terms of latitude, this is a maximum distance of approximately 7.5 km and a resolution of 1.8 m.

Five metrics are used to compare the protocols; PDR, average end-to-end delay, overhead, throughput and energy consumption. All of the metrics are as defined in Section 4.

6. Analytical and simulation results

In the first set of results, shown in Fig. 6, the maximum speed of the nodes is varied: [0, 5, 15, 25, 50, 75, 100] m/s. The number of nodes is kept constant at 25 and the data generation rate of each node is set to 1 pk/s. Also, the network size is 600 m by 600 m. At low speeds this scenario could simulate people walking or rotary wing drones cruising. Whereas, the higher speeds could represent the top speed of fast fixed wing drones. Overall the PDR and throughput are high, with a slight decline towards higher speeds. The analytical results predict this trend but tend to overestimate the packet loss as speeds increase. MACRO shows a slightly improved PDR, however it yields a slightly diminished throughput. In both PDR and throughput AODV and OLSR show relatively poor performance, this is also true of overhead. This large overhead from AODV and OLSR is caused by route discovery and the sharing of topology information. This causes congestion in the network, which in turn creates high delays and packet loss. Additionally, the increasing speed in the simulation means that topology changes become more frequent and consequently more overhead is required to maintain up-to-date routes. The analytic results for overhead are very close to PHASERs simulated results, showing minimal increase in the two metrics as the speed is increased. The delay results show a minor increase at high speeds, with the analytical results anticipating slightly higher delay times than were simulated. AODV has the longest delay times, whereas the average delay in OLSR is low and comparable to PHASER. The low PDR of OLSR means that many packets are dropped; as such there are fewer packets in the network, which means the packets that are delivered can traverse the network much faster. MACRO shows a slightly better delay than PHASER, but PHASER gives an improved level of overhead and subsequently better energy performance as well. In general AODV and OLSR perform badly in these scenarios as they are not intended for these kinds of high speeds. Whereas, PHASER and MACRO show much better performance because they were designed for this type of scenario.

The set of results in Fig. 7 shows how PHASER responds to changes in the number of nodes: [15, 25, 50, 75, 100] nodes. In order to keep a roughly similar node density the square network area was also varied by changing the side lengths: [400, 600, 1000, 1200, 1500] m. The maximum speed was kept constant at 25 m/s and the data generation rate of the sensor nodes was also maintained at 1 pk/s. These simulations illustrate the scalability of the protocol and, in terms of radiation mapping, how it may be used for various sizes of irradiated areas. The PDR results show that PHASER, MACRO, AODV and OLSR all follow the same trend as the network size increases, and all give better results in smaller network sizes. PHASERs increased packet loss is generally due to the increase in cycle time from the larger number of nodes requiring access to the medium. The analytical results tend to overestimate the PDR. Again MACRO shows the best PDR of the four protocols. Contrastingly, the throughput gets higher as the number of nodes increases, which is simply due to the fact that every node is producing data, so with more nodes more data is being generated. Again, this is overestimated by the analytical results. The overhead and energy consumption results for AODV and OLSR are high and show a steep increase as more nodes are added. PHASERs analytical results are close to the simulated, both of which show only a slight increase as the number of nodes...
get larger. Whereas MACROs overhead and energy consumption is much larger than that of PHASER and show a significant increase as more nodes are added. The delay results highlight AODV as having the worst performance and OLSR having comparable results to PHASER, however this is due to the significantly lower PDR, as discussed previously. MACROs delay is generally worse than PHASERs and shows an increase with high numbers of nodes. The overall trend indicates lengthening delay times and generally worse performance overall for all four protocols as the number of nodes is increased.

Fig. 8 gives results for the protocols over varying traffic loads: [0.1, 0.5, 1, 2.5, 5, 10, 24.43] pk/s. The number of nodes and maximum speed are kept constant at 25 nodes and 25 m/s respectively. Every sensor node is generating data, so the network-wide traffic load varies from 2.4 pk/s to 586.32 pk/s. The highest generation rate of 24.43 pk/s represents the maximum allowable packet transmit rate. This is dictated by the length of a TDMA cycle; since a node is only able to transmit once in a cycle its maximum data generation rate is $1 / \Delta$, which results in a network-wide data generation rate of $1 / \Delta$. With relation to the radiation mapping application, changing the data generation rate can control the resolution of the mapping. The PDR is generally high and shows a distinct increase towards higher data generation rates, which is characterised well by the analytic results. The reason for this increase is that the TDMA cycle time is equal to the inverse of the data generation rate. In practise this is done by evenly lengthening each time slot. This means that with low packet rates the cycle time is quite long, which will make the gradient metric slow to update. So, with a data generation rate of 0.1 pk/s the cycle time is...
10 s and an average link lifetime, \( t_a \), of 12.73 s, the gradient field is not updated regularly enough. However, this is only a problem with very low data generation rates since 0.5 pk/s has a cycle time of 2 s, which is enough to maintain the gradient field as the topology changes. The choice of lengthening the slots to accommodate the required data generation rate was made to save energy, which is shown by the energy consumption results, which gradually increases as more packets are introduced into the network.

Contrastingly, AODV and OLSR use a large amount of energy, even at low packet rates, which then only slightly increases as the data generation rate increases. MACRO also shows an increase in energy consumption with higher packet rates, however after 10 pk/s it plateaus due to saturation. PHASER's analytical results slightly overestimate the increase in energy. The overhead results show how AODV and OLSR produce a lot of overhead even though there is not much data to send, but as more data is introduced the ratio between data bits and overhead evens out. PHASER shows a constantly low amount of overhead, which is closely predicted by the analytical results. MACRO shows a slight increase in overhead but its throughput seems to reach a saturation limit at packet generation rates above 5 pk/s. PHASER's throughput rises steadily as the data generation rate is increased, which is modelled well by the analytical results. The throughput of AODV and OLSR begins to plateau as more and more packets are lost. The delay results show PHASER consistently low delay, which actually decreases as more data is generated. This is again due to the decrease in cycle time, which means each node is allowed access to the medium more frequently so packets may be passed through the network faster. The analytical results capture this behaviour well. Contrastingly MACRO shows a steep rise in delay times after 5 pk/s. In general, PHASER shows a consistently high level of performance above 1 pk/s, whereas the other three protocols show severe degradations as the level of traffic is increased.

Overall, these results have shown PHASER to be suitable for a wide range of scenarios, including those specifically designed to model the application of radiation mapping with UAVs. The AODV and OLSR results, as expected, show an inferior level of performance in these scenarios. In comparison to MACRO, PHASER shows consistently better energy consumption, which is key in UAV orientated applications where power resources are limited. However, MACRO shows an improved level of PDR in some scenarios. In terms of delay, PHASER's end-to-end delay is either comparable or better than that of MACRO. PHASER also shows a superior ability to handle high levels of packet generation, which would be necessary if the resolution of the mapping was to be increased.

7. Conclusion

This paper has presented an original routing protocol designed for MWSNs. It specifically targets the application of radiation mapping using UAVs. However, it should be noted that the protocol could be used in other similar scenarios and for different applications. PHASER uses a novel, low overhead, method of maintaining a gradient metric, even in high speed scenarios, through the use of a global TDMA MAC layer. PHASER is also particularly robust from its use of the blind forwarding technique, which allows data to simultaneously take multiple paths through the network.

Extensive analysis and simulations, modelled on the radiation mapping application, have shown the protocol to be effective in a wide array of situations, over varying mobility, scalability and traffic levels. MACRO, AODV and OLSR results have also been given to illustrate the performance level of the routing protocols currently used in sensor network implementations.

Presently, testbed implementations of the protocol are being developed to further validate its use in radiation mapping applications and also explore other applications for which PHASER is suited. Other future work will investigate the effects of channel fading on the performance of the protocol.

Acknowledgement

This work is partly supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant number EP/K503198/1.

Appendix

In PHASER each packet has a fixed maximum number of frames, which it is able to forward. The number of frames held by a packet is called the frame capacity. If the maximum frame capacity is too small a node will not be able to forward all the data required and so some frames will be dropped. If the frame capacity is too big there could be a large amount of bandwidth wasted. In this appendix, we solely consider losses from the dropping of frames.

The ideal required frame capacity, \( F \), for a fixed topology is directly related to the number nodes in the network, \( n \), and the number of existing connections to the sink, \( c \).

So the maximum required frame capacity for a given \( n \) and \( c \) is given as:

\[
F_{\text{max}} = n - c. \tag{20}
\]

Conversely, the minimum required frame capacity is given by:

\[
F_{\text{min}} = \left\lceil \frac{n - 1}{c} \right\rceil. \tag{21}
\]
The total number of possible topologies for a network with \( n \) nodes is:

\[
T = 2^{\frac{n(n-1)}{2}}. \tag{22}
\]

Further to this, the number of possible topologies for a given \( n \) and \( c \) is:

\[
T(n, c) = \left(\frac{n-1}{c}\right) \cdot 2^{\frac{(n-1)(n-2)}{2}}. \tag{23}
\]

If a network had a maximum number of connections for which no losses could occur, \( x \), then no losses would occur as long as \( c \geq x \). So the total number of topologies that are possible without loss, for a given \( x \), is:

\[
T(n, x) = 2^{\frac{(n-1)(n-2)}{2}} \cdot \sum_{c=x}^{n-1} \left(\frac{n-1}{c}\right). \tag{24}
\]

Eq. (20) can be rearranged to give \( x \) for a fixed \( F \):

\[
x = n - F. \tag{25}
\]

Substituting (25) into (24) will give the total number of topologies that are possible without loss for a fixed \( F \):

\[
T(n, F) = 2^{\frac{(n-1)(n-2)}{2}} \cdot \sum_{c=n-F}^{n-1} \left(\frac{n-1}{c}\right). \tag{26}
\]

This can subsequently be described as the fraction of possible topologies without loss, \( a \), by simply dividing (26) by (22), which gives:

\[
\alpha = 2^{1-n} \cdot \sum_{c=n-F}^{n-1} \left(\frac{n-1}{c}\right). \tag{27}
\]

Now various values for \( F \) may be evaluated using (27), in order to achieve a satisfactory proportion of topologies that can occur without loss. This will help to minimise \( F \), which will limit wasted bandwidth and reduce delay by making packet sizes smaller. Also, \( a \) can be maximised, which will decrease the chances of losses occurring from insufficient frame capacity.

References


