## Contents

### REGULAR PAPERS

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCA: New Cooperative Algorithm for Reducing Topology Control Packets in OLSR</td>
<td>125</td>
</tr>
<tr>
<td>Adil Benabbou, Abdelali Boushaba, Azeddine Zahi, Rachid Benabbou, Mohammed Oumsis, and Said El Alaoui Ouatik</td>
<td></td>
</tr>
<tr>
<td>A Distributed UWB-based Localization System in Underground Mines</td>
<td>134</td>
</tr>
<tr>
<td>Yuanqing Qin, Fang Wang, and Chunjie Zhou</td>
<td></td>
</tr>
<tr>
<td>Wi-Fi-based Positioning in a Complex Underground Environment</td>
<td>141</td>
</tr>
<tr>
<td>Matteo Cypriani, Gilles Delisle, and Nadir Hakem</td>
<td></td>
</tr>
<tr>
<td>Learning Algorithm of Neural Networks on Spherical Cap</td>
<td>152</td>
</tr>
<tr>
<td>Zhixiang Chen, Jinjie Hu, and Feilong Cao</td>
<td></td>
</tr>
<tr>
<td>Research on Channel Selection Algorithms in Cognitive Radio Networks</td>
<td>159</td>
</tr>
<tr>
<td>Lin Li, Yi-na Deng, Yang Yuan, and Wen-jiang Feng</td>
<td></td>
</tr>
<tr>
<td>Gigabit Wireless Networking with IEEE 802.11ac:Technical Overview and Challenges</td>
<td>164</td>
</tr>
<tr>
<td>Farhan Siddiqui, Sherali Zeadally, and Khaled Salah</td>
<td></td>
</tr>
</tbody>
</table>
NCA: New Cooperative Algorithm for Reducing Topology Control Packets in OLSR

Adil Benabbou 1, Abdelali Boushaba 1, Azeddine Zahi 2, Rachid Benabbou 2, Mohammed Oumsis 3,4, and Said El Alaou Ouatik 1

1. Laboratory LIM, Dhar Mehraz Sciences Faculty, Sidi Mohamed Ben Abdellah University, Fez, Morocco
2. Laboratory LSIA, Faculty of Sciences and Technology, Sidi Mohamed Ben Abdellah University, Fez, Morocco
3. LRIT laboratory Associate Unit to CNRST (URAC 29), Mohammed V University of Rabat, Rabat - Morocco
4. High School of Technology -Salé, Mohammed V University of Rabat, Rabat - Morocco
Email: adil.benabbou@usmba.ac.ma, abdelali.boushaba@usmba.ac.ma, dinzahi@yahoo.fr, rachid.benabbou@usmba.ac.ma, oumsis@gmail.com, said.ouatik.elalaoui@gmail.com

Abstract—OLSR (Optimized Link State Routing Protocol) is a routing protocol designed especially for MANETs (Mobile Ad hoc Networks) and is currently the most widely deployed for this type of networks. OLSR uses multipoint relay (MPR: MultiPoint Relay) flooding mechanism to optimize topology control messages broadcasting throughout the network. Each node of the network selects a subset of nodes (MPR) among its neighbors to disseminate control messages. Reducing topology control packets (TC) is the key functionality of OLSR. In the Ad hoc network, the number of TC depends on the choice of MPR. In this paper, we propose a new cooperative MPR selection algorithm with the aim to reduce TC packets. We implemented this algorithm in NS2 and three other algorithms are known to be performant in the domain. Simulation results show that our algorithm reduces TC packets relative to other algorithms. This reduction is reflected positively on other performance parameters.

Index Terms—Ad hoc Network; MANET; OLSR; UM-OLSR; NFA; MPR; TC; NS-2

I. INTRODUCTION

Nowadays, the need for more mobility and power to share or to exchange information, at any time and by using mobile devices only (mobile phones, laptops, etc.), has made the notion of Mobile Ad hoc Networks (MANETs) very widespread. MANET is a kind of distributed multi-hop wireless network, which does not rely on a pre-existing infrastructure, such as routers or access points. Such network is composed of a number of autonomous, wireless and mobile nodes. Unfortunately, the classical routing protocols, designed for the networks with strong infrastructure, are inappropriate with the inherent characteristics of MANETs. Therefore, a new routing technique is used; the nodes agree to relay communications of their neighbors so that messages can spread beyond the coverage area of their original transmitters. This routing technique has been widely studied. As a result, several protocols have been proposed namely AODV [1], DSR [2], DSDV [3] and OLSR[4].

OLSR (Optimize Link State Routing Protocol) [4] is the well-known and often implemented protocol in MANETs. It is a proactive routing protocol on which each node regularly exchanges topology information with other nodes. The main idea of OLSR is the optimization of the broadcast discovery messages and routes updating throughout the network. The optimization is performed by the use of the Multipoint Relay (MPR) as a flooding mechanism. Indeed, each node selects, independently, its MPR nodes as a subset of one-hop neighbors that allows it to reach all two-hop neighbors. So, only nodes that are selected as MPR are allowed to broadcast control messages containing topology information. In conclusion, we can say that the use of the multipoint relay technique reduces, significantly, the impact of the dissemination of control messages and contributes to the overall performance of the protocol. This makes the MPR selection problem a keystone of the OLSR protocol studies.

The MPR selection problem, which aims to find an optimal set of MPR nodes covering the whole of two-hop neighbors, has been proven to be NP-complete [5] [6]. Therefore, several heuristic algorithms have been proposed to address this problem in practice [4] [7] [8] [9]. They mostly used only the knowledge of one-hop neighbors and two-hops, and are aimed to optimize the MPR selection with a specific purpose such as minimizing the number of MPR nodes, reducing the TC parquets, improving QoS, etc. In this respect, we have proposed, in an earlier work [10], an algorithm with the purpose of reducing the number of topology control packets (TC) by minimizing the number of MPR nodes, locally in each node of the network. While analyzing its simulation results, we found that reducing the number of MPR locally does not necessarily lead to the reduction of the whole TC packets in the network. We have also observed an increase in the total number of MPR nodes. Therefore, we have concluded that the reduction of the total number of MPR in the network is relaying to the degree of cooperation between nodes when selecting their MPR. In this case, each node needs to know, among its one-hop neighbors those which have been selected as MPR by other nodes.

© 2015 ACADEMY PUBLISHER
doi:10.4304/jnw.10.3.125-133
In this paper, we propose a new algorithm, called New Cooperative Algorithm (NCA), to select MPR in OLSR. It is a cooperative algorithm which is based on a heuristic that enhances the one presented in [10]. The purpose of this heuristic is to minimize the number of MPR locally in each node, and secondly, to favor the one-hop neighbor nodes which have already been selected as MPR by the maximum of nodes. We have implemented our algorithm, the NFA algorithm [7], and the cooperative algorithm [9] on NS2 [11]. The simulation results show that our algorithm reduces the number of topology control packets (TC) up to 15% for fixed networks and 19% for mobile networks.

The rest of the paper is organized as follows: In Section II, we discuss the related work on the MPR selection; In Section III, we present the functionalities of OLSR and the basic MPR selection algorithm; Section IV is devoted to our MPR selection algorithm. Parameters and simulation results of our algorithm are presented in Section V.

II. RELATED WORK

The reduction of routes flooding is the key to all routing protocol in MANETs. OLSR performs this reduction by using a set of nodes, called multipoint relays (MPRs), which are responsible for generating and broadcasting the topology control messages (TC) through the network. Indeed, each node (selector node) independently selects a set of MPR nodes, from its one-hop neighbors with the aim of covering all two-hop neighbors. These MPR nodes act as relay nodes for messages sent by the selector node, and thus organize the broadcast of the communications in the network. This broadcasting is very consumer of resources such as bandwidth and throughput. It also generates collisions and can disrupt the data traffic routing. Hence, selecting an optimal MPR-set is primordial for more performance and efficiency in OLSR protocol.

Finding an optimal MPR set has been proven to be an NP-complete problem [5] [6]. So, heuristic algorithms are used to select MPR sets in practice. In the literature, the existing heuristics are mostly designed to find the adequate MPR sets with a specific objective such as minimizing the number of MPR nodes, as was the goal in the original specification [4] and in other works [5] [12]; improving QoS [13]; reducing the number of collisions; minimizing the overlap between MPRs or maximizing the global bandwidth [14]. In this section, we present some important works that focus on reducing topology control packets (TC).

In [12] the authors proposed a heuristic to select the MPR-set. This heuristic has been used in the original specification of OLSR [4]. The basic idea is to select, as MPRs, the one-hop neighbors having the strongest coverability or covering two-hop neighbors that cannot be covered by other neighboring nodes. This heuristic has been deeply analyzed [15] [16], the results show that the heuristic allows to provide a set MPR near optimal that greatly reduces the retransmission of messages topology control (TC).

In [17] the authors have found incoherence between the RFC 3626 specification of OLSR and the UM-OLSR implementation. Indeed, two bugs are identified, in the MPR selection procedure, as a result of a deep inspection of the code of the UM-OLSR implementation. Thus, the authors have proposed the solutions for correcting this incoherence with the aim to be conforming to the RFC specification of OLSR. The corrected implementation was then validated by simulations, which show that these solutions provide more credible results.

Li et al have presented in [7] a new algorithm, called Necessity First Algorithm (NFA), to select MPR-sets in OLSR. In contrast to the original heuristic, NFA is based on the notion of “necessity of selecting” that may be associated with a one-hop neighbor node, which is the only to cover some two-hop neighbor nodes. This notion describes the ability of a node to be, imperatively, selected as MPR. Thus, the heuristic of NFA selects, as MPRs, the one-hop neighbors with “necessity of selecting”. Then, the neighbor with poorest cover ability is deleted to purposely create nodes with “necessity of selecting”. The results of simulation show that NFA improves the original heuristic; the algorithm reduces the number of MPR nodes and topology control messages (TC).

A cooperative algorithm for selecting MPR sets is described in [9]. It consists to give priority to the nodes that are already selected as MPR by other nodes. Thus, before calculating its MPR, each node i determines the set, MN(i), of its one-hop neighbors that have been selected as MPR for other nodes. If the MN(i) set covers all two-hop nodes of i, the MPR nodes are selected from this set, if not the basic algorithm is applied. This algorithm reduces the number of TC packets compared with the UM-OLSR implementation [4].

In [8] the authors claimed that the local optimization of the MPR sets does not necessarily lead to a global optimization MPRs through the network. They proposed two strategies: The first one attempts to minimize the total number of MPR with the aim of reducing the number of topology control messages. This strategy chooses among the candidates to become MPR, those having the maximum MPR selectors. The second strategy aims to increase the stability of routes; it focuses on reducing changes in the sets of MPR selectors by favoring the candidates which were already MPR in the previous run of the algorithm.

The authors proposed in [18] two enhancements of the original MPR selection algorithm described in the UM-OLSR implementation [4]. These enhancements are meant to extend the visibility of a node to three hops when two nodes are candidate to be MPR having the same reachability and the same degree. The first enhancement favors the candidate which has more isolated nodes from the two hop neighbors. A two hop node is isolated if it has no one-hop neighbors, except one which is a candidate to become MPR. While the second enhancement gives priority to improving the node that has more not-MPR nodes from the two hop neighbors. The results of simulation show that both variants can
reduce the number of TC packets, the routing cost and the efficiency compared to UM-OLSR implementation [4].

In [19] a multipoint relay selection method for robust broadcast in a wireless ad hoc network is proposed. In this method, a node selects additional MPR nodes so that it can cover two-hop nodes m-times. The proposed method can improve network throughput and delivery ratio compared to the original MPR selection algorithm of OLSR [4].

III. OLSR PROTOCOL

OLSR (Optimized Link State Routing Protocol) is a proactive routing protocol for ad hoc networks, which establishes routes based on local knowledge of the topology. A node discovers its one-hop neighbors and two-hop neighbors. It uses an optimized technique based on nodes called multipoint relays (MPR) to ensure optimized broadcast messages. The choice of MPRs is done after the discovery phase of the neighbors of all nodes in the network using HELLO messages (default broadcast every 2 seconds). These messages allow each node u to know its neighbors in one-hop noted N(u) and 2-hop neighbors denoted N2(u) and then calculate its MPR. Each node stores a database containing information network topology which is constructed from topology control message (TC: Topology Control). TC message contains a list of nodes which selected the original node of TC as MPR. These messages are broadcast by default to all nodes in the network every 5 seconds. Neighbor nodes of a node u that are not MPR of u, receive and process messages TC, but do not broadcast them. Each node maintains all its selectors and multipoint relays and retransmits only the packets received for the first time from these multipoint relays selectors.

The principle of MPR is to use the knowledge of two-hop neighbors to locally optimize the broadcast. Thus, each node u is a set of nodes MPRs among its neighbors in the first level that can reach all two-hop nodes. The example in Fig. 1 shows the difference between the classical flooding (Fig. 1(a)) and the optimized multipoint relays (Fig. 1(b)) [12].

![Figure 1. The classical diffusion (a) and the optimized diffusion through MPR nodes (b)](image)

In the classical case it takes 24 retransmissions to achieve the 3-hop nodes; while only 12 retransmissions are needed with MPRs.

To transmit a message to all nodes in N2(u) of a node u using optimized diffusion by MPR, [MPR(u)+1] emissions are necessary. So, to increase the network performance, we must find a minimum number of multipoint relays for each network node. The choice of MPR nodes is a NP-complete problem [5] and several algorithms have been proposed to solve it. The basic MPR selection algorithm is described in RFC 3626 [4] specification as shown in Fig. 2. Before introducing this algorithm, some notations should be described first:

- u a node in Ad hoc Network
- N(u) the subset of one-hop neighbors of the node u
- N2(u) a set of two-hop neighbors of u by excluding:
  1. The nodes only reachable by members of N(u) with willingness WILL_NEVER. The willingness is a parameter stored and maintained by each node in MANET. Willingness may be set to any integer value from 0 to 7, and specifies how willing a node is to be forwarding traffic on behalf of other nodes. Nodes will, by default, have willingness WILL_DEFAULT (3). WILL_NEVER (0) indicates a node which does not wish to carry traffic for other nodes, for example due to resource constraints (like being low on battery). WILL_ALWAYS (7) indicates that a node always should be selected to carry traffic on behalf of other nodes, for example due to resource abundance (like permanent power supply, high capacity interfaces to other nodes).
  2. All the symmetric neighbors: the nodes for which there exists a symmetric link to this node on some interface.
- D(y) the degree of a one-hop neighbor node y (where y is a member of N(u)), is defined as the number of symmetric neighbors of node y, excluding all the members of N(u) and excluding the node u performing the computation.

**MPR(u)** the MPR set of u.

After the explanation of these notations and concepts, the heuristic used by OLSR is described as follows:

<table>
<thead>
<tr>
<th>Algorithm 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Start with an MPR set made of all members of N(u) with N williness equal to WILL ALWAYS.</td>
</tr>
<tr>
<td>2. Calculate D(y), where y is a member of N(u), for all nodes in N(u).</td>
</tr>
<tr>
<td>3. Add to the MPR set those nodes in N(u), which are the only nodes to provide reachability to a node in N2(u). For example, if node b in N2(u) can be reached only through a symmetric link to node a in N(u), then add node a to the MPR set. Remove the nodes from N2(u) which are now covered by a node in the MPR set.</td>
</tr>
<tr>
<td>4. While there exist nodes in N2(u) which are not covered by at least one node in the MPR set:</td>
</tr>
<tr>
<td>4.1. For each node in N(u), calculate the reachability, i.e., the number of nodes in N2(u) which are not yet covered by at least one node in the MPR set, and which are reachable through this one-hop neighbor.</td>
</tr>
<tr>
<td>4.2. Select as a MPR the node with highest N_williness among the nodes in N (u) with nonzero reachability.</td>
</tr>
<tr>
<td>4.2.1. In case of multiple choices select the node which provides reachability to the maximum number of nodes in N2(u).</td>
</tr>
<tr>
<td>4.2.2. In case of multiple nodes providing the same amount of reachability, select the node as MPR who’s D(y) is greater.</td>
</tr>
<tr>
<td>4.3. Remove the nodes from N2(u) which are now covered by a node in the MPR set.</td>
</tr>
</tbody>
</table>

![Figure 2. The basic algorithm for selecting MPR](image)
This heuristic does not always provide an optimum solution. Fig. 3 shows that this heuristic gives three MPR nodes for node u, whereas the optimal solution is two MPRs nodes.

![Figure 3. Comparison between the solution of the basic algorithm and the optimal solution](image)

**IV. PROPOSED ALGORITHM FOR SELECTING MULTIPoint RELAYS**

In this section we present our new heuristic for selecting MPR nodes in OLSR. The undelaying algorithm enhances the previous one presented in [10], which attempts to reduce the number of MPR locally in each node by introducing cooperation between nodes. So, our heuristic selects MPR for a source node by giving priority to one-hop neighbor nodes that have already been selected as MPR by other nodes. The heuristic maximizes the intersection of MPR sets, and therefore, minimizes the total number of MPR in the network.

Before stating our algorithm, some notations should be presented first. For each node u of the network and its one-hop neighbor a, we propose the notation shown in Fig. 4:

![Figure 4. Algorithm parameters illustration](image)

- **u** is a node in the MANET.
- **N(u)** is a subset of one-hop neighbors of the node u.
- **N2(u)** is the set of two-hop neighbors of u.
- **NNC(a)** is the number of nodes in N2(u) covered by node a.
- **NNA(a)** is the number of nodes N(u) that share the coverage with node a (named associates of a).
- **CA(a)** is the number of nodes in N2(u) covered by the associates of node a.
- **MPRSEL(a)** is the neighbors nodes that have selected the node a as MPR.
- **MPRSEL_ONEHOP** is the nodes that have selected a one-hop node of u as MPR.
- **Weight(a)** is the weight of the node a defined as follows:

\[
Weight(a) = \frac{NNC(a)}{NNC(a) + CA(a)} \quad (1)
\]

**Weight_C(a):** is the cooperative weight of the node a defined as follows:

\[
Weight_C(a) = Weight(a) \ast cf \quad (2)
\]

where \( cf = \left( 1 + \frac{MPRSEL(a)}{MPRSEL_ONEHOP} \right) \)

The weight parameter, weight(a), defined by the formula (1) represents the coverage rate of the two-hop nodes of u by the node a according to its associates. The choice of this parameter gives more chance to associate nodes, which cover more nodes of N2(u), to become MPR. Using this parameter reduces the number of MPR locally for each network node. But the number of TC packets did not follows the same trend; instead it has increased in the majority of cases [10]. Additionally, to take into account the cooperation between the nodes, we introduce, in the formula (2), the multiplicative factor, cf. This factor allows the MPR selection procedure, to favor the nodes that have more than MPR selectors, and increases the chance of the node, a, to become MPR. Accordingly, the cooperative weight parameter, weight_C, defined in the formula (2), makes to combine two purposes at the same time: favoring the nodes that have more than MPR selectors and locally reducing, the number of MPR.

Our algorithm is based on the weight_C parameter to select the MPR set for each source node. It iteratively eliminates one-hop node with the minimal weight_C parameter to bring up the nodes of N (u) that only reach one or some nodes in N2(u). These nodes will be added to MPR set. The sketch of our algorithm is presented in Fig.5.

**Algorithm 2**

1. Start with an MPR set made of all members of N(u) with N willingness equal to WILL AL WAYS.
2. While there exist nodes in N2(u) which are not covered by at least one node in the MPR set:
   2.1. Add to the MPR set those nodes in N(u), which are the only nodes to provide reachability to a node in N2(u).
   For example, if node b in N2(u) can be reached only through a symmetric link to node a in N(u), then add node a to the MPR set. Remove the nodes from N2(u) which are now covered by a node in the MPR set.
   2.2. For each node v of N(u) Calculate the weight:
   \[
   Weight_C(v) = \frac{Weight_C(a)}{N(u) - N2(u)} \quad (3)
   \]
   2.3. Delete the node v0 with a minimal weight_C of N(u).

![Figure 5. Our MPR selection algorithm](image)

In our algorithm, in order to select its MPR, a network node needs both a number of MPR selectors for each one-hop neighbor and the total number of MPR selectors of all the one-hop neighbors (without repetition). These numbers are not available for each node. HELLO packet can be used to exchange the numbers of MPR selectors between the neighbor nodes in the network. For example, if a and b are two neighbor nodes, and the number of MPR selectors of a is 4 and the number of MPR selectors of b is 3. Then node a informs the node b indicating in a Hello message that the number of its MPR selector is 4. Conversely, b informs the node a specifying in the
HELLO message that the number of MPR selector of node b is 3. Fig.6(b) shows the format of the HELLO message described in RFC3626 [4]. This message contains the reserved fields and other fields to exchange important information between nodes. We have extended the structure of the HELLO message as shown in Fig.6(c) to use a reserved field in the HELLO messages to exchange the number of MPR selector (MPRSEL). In addition, we have extended the structure of a one-hop neighbor node (neighbor tuples) by adding a field (N_Nb_MPRSEL) to keep updating the number of MPR selectors of a one-hop neighbor nodes as shown in Fig. 6(a).

<table>
<thead>
<tr>
<th>N_neighbor_main_addr</th>
<th>N_Status</th>
<th>N_Willingness</th>
<th>N_Nb_MPRSEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link Code</td>
<td>Reserved</td>
<td>Link Message Size</td>
<td>Neighbor Interface Address</td>
</tr>
<tr>
<td></td>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) The extended structure of neighbor tuples</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) The HELLO message format of OLSR

<table>
<thead>
<tr>
<th>MPRSEL</th>
<th>Htime</th>
<th>Willingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link Code</td>
<td>Reserved</td>
<td>Link Message Size</td>
</tr>
<tr>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Interface Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) The HELLO message format of OLSR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) The extended HELLO message format of OLSR

Figure 6. OLSR HELLO message formats and structure of neighbor tuples

So, there is no additional signaling overhead introduced by our algorithm; we just used a reserved field of HELLO messages to disseminate the number of MPR selectors.

To calculate the number of MPR selectors of all one-hop nodes (MPRSELONEHOP), the topology table is used, since this table contains the fields: T_dest_addr, T_last_addr, T_seq and T_time. T_dest_addr is the main address of node, which may be reached in one-hop from the node with the main address T_last_addr [4]. Thus, T_last_addr is a MPR of T_dest_addr [4].

Let’s take an example to better illustrate the operating principle of our algorithm. As shown in Fig. 7(a), the network consists of 12 fixed nodes placed in an area of 1000m x 1000m. Numbers in brackets next to each node represent the MPR selectors of this node, for example nodes 10, 11, 5, 9, 3 and 0 chose node 4 as a MPR node.

In this example we will calculate only the MPR of node 3. Where N(3) = [0, 4, 10, 2, 7, 8, 9] and N2(3) = [1, 5, 11, 6]. We have, also, shown only the wireless links between the node 3 and the nodes of N(3) (represented by continuous lines) and those between the nodes of N(3) and nodes of N2(3) (represented by dashed lines).

As there is no node in N(3) that only covers nodes in N2(3), we proceed to calculate the weight_C parameter for each node in N(3). We have | MPRSEL_ONEHOP | = 12 because the network nodes that have chosen one or nodes in N(3) as MPR are: 2, 5, 1, 4, 0, 3, 10, 11, 9, 8, 7 and 6.

We remove the node in N(u) which has the minimum weight_C. We can remove either node 7 or node 10 as they have the same weight_C. If we choose to remove the node 7 of N(u), Node 2 is the only in N(u), which covers 6, then it is chosen as MPR and nodes 1 and 6 are removed as shown in Fig. 7(b).

We recalculate the weight parameter for the remaining nodes because there is no node in N(3) that covers only nodes in N2(3). Table II presents the results of parameters calculation.

<table>
<thead>
<tr>
<th>TABLE I.</th>
<th>PARAMETER SELECTION OF MULTIPROET RELAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>N(3)</td>
<td>NNC Associates CA Weight</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
</tr>
</tbody>
</table>
Now we remove the node 10 with the minimum weight_C of N(u) and we get the network shown in Fig. 7 (c).

Node 4 is the only in N (3), which covers 11, then it is chosen as MPR and 5, 11 are removed. The algorithm stops because N2(3) becomes empty, finally, the nodes 2 and 4 are selected as MPR of node u, while the basic MPR selection algorithm selects three nodes 0, 2 and 4 as MPR for the same example.

We have also calculated the MPR of each node using a three other algorithms: UM-OLSR [20], NFA [7] and cooperative algorithm [9]. Table III summarizes the obtained results.

<table>
<thead>
<tr>
<th>TABLE III.</th>
<th>MPR nodes, total number of MPR and total number of TC for the four algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>NCA</td>
</tr>
<tr>
<td>0</td>
<td>2.4</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>2.4</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
</tr>
<tr>
<td>5</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>2.3</td>
</tr>
<tr>
<td>9</td>
<td>3.4</td>
</tr>
<tr>
<td>10</td>
<td>3.4</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Total MPR number</td>
<td>4</td>
</tr>
<tr>
<td>Total TC number</td>
<td>670</td>
</tr>
</tbody>
</table>

We note that our algorithm reduces, locally and globally the number of MPR. The global reduction in MPR versus NFA implies a decrease in TC packets generated by MPR every 5 seconds. And the local reduction of MPR compared to UM-OLSR implies a reduction of TC packets broadcasted. In both cases (locally and globally) this leads to a reduction of the total number of TC packets in the network as shown in Table III.

V. SIMULATION AND RESULTS

A. Evaluation Criteria

The objective of the experiments with the simulator NS-2 is to validate the influence of our MPR algorithm selection on the performance parameters by analyzing the following metrics:

- **Number of TC packets**: The number of topology control packets (TC) in the network.
- **PDR**: Is the ratio between the number of received data packets and the number of data packets sent.
- **Routing Cost**: Is the ratio between the number of routing packets sent and the number of data packets received by destinations.
- **Efficiency**: Is the ratio between the number of delivered packets which are transmitted (Data) and (Data + RoutingOverhead).

B. Simulation Parameters

The simulation parameters are summarized in Table IV.

<table>
<thead>
<tr>
<th>TABLE IV.</th>
<th>SIMULATION PARAMETERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters used for traffic model</td>
<td>S 300 s</td>
</tr>
<tr>
<td>Type of traffic</td>
<td>CBR</td>
</tr>
<tr>
<td>Number of connections</td>
<td>30% of network nodes</td>
</tr>
<tr>
<td>Packet size</td>
<td>512 Octet</td>
</tr>
<tr>
<td>Parameters used for mobility model</td>
<td>RWP</td>
</tr>
<tr>
<td>Adhoc network area</td>
<td>1000 m x 1000 m</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>30/50/75/100/125/150</td>
</tr>
<tr>
<td>Pause time</td>
<td>20 s</td>
</tr>
<tr>
<td>Maximum speed of nodes</td>
<td>0/2/5/10/15/20/25/30 (m/s)</td>
</tr>
<tr>
<td>Mobility model</td>
<td>RWP</td>
</tr>
<tr>
<td>Parameters used for physical and link layers</td>
<td>IEEE 802.11</td>
</tr>
<tr>
<td>MAC protocol</td>
<td>IEEE 802.11</td>
</tr>
<tr>
<td>Propagation model</td>
<td>Two-ray ground</td>
</tr>
<tr>
<td>Transmission range</td>
<td>250 m</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>2 Mbps</td>
</tr>
<tr>
<td>Maximum queue size</td>
<td>50</td>
</tr>
</tbody>
</table>

C. Simulation Results

The objective of these simulations is to evaluate our MPR algorithm selection and to compare the obtained results with those of UM-OLSR implementation [20], the NFA algorithm [7] and cooperative algorithm [9]. To study the influence of MPR nodes selection on the number of topology control packets (TC), the RDP, the routing cost and the Efficiency, we conducted two experiments: the first, concerns the simulation of a static Ad hoc network; while the second deals with a simulation of a mobile ad hoc network.

1) Experiment 1

In this experiment, the static network nodes are randomly distributed in an area of 1000m x 1000m. To evaluate the impact of network density, we varied the number of nodes between: 30, 50, 75, 100, 125 and 150. The number of nodes that will generate data traffic represents 30% of the network nodes. For each number of nodes, four different scenarios are simulated, and the average of the four results is presented in the following figures.

Fig. 8 shows the number of topology control packets (TC) depending on the number of network nodes. We note that, regardless of the density of the ad hoc network (the number of nodes is varied between 30 and 150). Our MPR selection algorithm reduces the number of TC compared with the three other algorithms. This reduction can respectively reach 6% for original UM-OLSR, 15% for NFA and 5% for cooperative algorithm. The reduction of TC appears clearly in the case of dense networks (between 100 and 150 nodes). These improvements can be explained by the fact that, in the dense networks, the probability of selecting nodes that are already MPR for other nodes increases. So, our algorithm, which favors...
those nodes, reduces the global number of MPR nodes in the network. Subsequently the number of topology control packets (TC) is also reduced because the MPR nodes are the only ones that generate the TC packets, and nodes that are not MPR, receive and process messages TC, but do not broadcast them.

Fig. 9 represents the PDR, the routing cost and the efficiency depending on the number of nodes. We note, in Fig. 9 (b), that the cost of routing followed the same trend as the number of TC. Indeed, whatever the density of the network, the routing cost given by our algorithm is less compared to other algorithms. It can be reduced up to 14% compared to the NFA algorithm. The reduction rate can also reach 5% for the other two algorithms (UM-OLSR and cooperative algorithm).

In Fig. 9 (c), we also note that our proposal is more efficient. It can increase efficiency up to 6% compared to NFA and up 2% compared to the other two algorithms. Despite the reduction in the number of TC packets and routing cost, we see in Fig. 9 (a) the PDR remains almost the same for all algorithms.

2) Experiment 2

Like the first experiment, we chose the area of 1000m x 1000m mobility. However, the number of node is fixed at 50. These nodes use the Random Waypoint mobility model [21]. The traffic between nodes is generated by a traffic generator that can create connections CBR at times uniformly distributed between 5 and 295 seconds. The size of packet data is 512 bytes. To study the impact of mobility on the four implementations, we varied the speed of mobile nodes between 2.5m/s and 30m/s.

Fig. 10 shows the number of TC in the network depending on the speed of nodes. We can confirm that regardless of the maximum speed of nodes in the network, our MPR selection algorithm reduces the number of TC packets in the network. Reducing TC is varied between 1% and 19% for the UM-OLSR implementation; between 12% and 19% for the algorithm NFA and between 1% and 4% for cooperative algorithm. We note that the reduction rate increases in the low and medium speeds. Whereas, moving with a relatively high speed, the reduction of TC packets is not significant. Indeed, the number of MPR selectors of a node sent to a neighbor node at a given time can change rapidly due to the high speed movement of the nodes. This may induce a node to select its MPR based on outdated information.
routing cost compared to the other three algorithms (until: 15% for NFA, 4% for the cooperative algorithm and 7% for UM-OLSR implementation). Fig. 11 (a) shows a small improvement in PDR at low speeds, indeed, our algorithm increased the PDR up to 2% to less than 10 m/s speeds. In addition as shown in Fig. 11 (c), we can note that our proposal is more efficient in mobile networks; it can increase efficiency up to 5% compared to NFA and can improve it by 1% compared to the other two algorithms.

![Figure 11. The performance metrics (mobile network)](image)

D. Analysis of Simulation Results

The principle of MPR is to optimize the broadcast of topology control packet (TC). Reduction of the topology control packets number is the key of routing protocol in Ad hoc networks. Indeed, the principle of MPR is to optimize the broadcast of topology control packets (TC). Each MPR node in the network generates every 5 seconds a TC message announcing the nodes that have elected the TC original node as MPR. These TC packets are broadcast by MPR nodes to all nodes in the network. The choice of different MPR nodes increases the number of TC generators packets; while cooperation between these nodes in the MPR selection gives them more opportunity to choose the most common MPR nodes. This, necessarily leads to a reduction of global MPR number in the network, and allows the reduction of the total number of TC packets.

VI. CONCLUSION AND PROSPECT

In this article, we proposed a new cooperative algorithm for selecting MPR nodes in OLSR protocol. The algorithm is based on heuristic favoring neighbors that have already been selected as MPR by the maximum of nodes. This reduces the total number of MPR nodes in the networks. This reduction leads to the reduction of the total number of TC packets and to the improvement of other performance parameters such as PDR, routing cost and efficiency. We have implemented our algorithm, the NFA algorithm and the cooperative algorithm on NS2 simulator. The simulation results show that our algorithm reduces the number of topology control packets compared to the other three algorithms for networks, both fixed and mobile. The reduction of TC affects positively other performance parameters.

As a future extension of this work, it would be interesting to consider other parameters of quality service during the cooperation between network nodes.

REFERENCES


Optimisation and Topology Control Traffic in Ad hoc Networks.


Adil Ben Abbou received a BSc degree in Computing from Faculty of Science Dhar Mehraz in 2002, a MSc degree in Operations Research and computer Sciences from Faculty of Science and Technology (FST, Fez, Morocco) in 2006. He is now a PhD degree in Networks, Telecommunications and Multimedia from National Higher School of Computer Science and Systems Analysis (ENSIAS), Rabat, Morocco in 2007. His research interests include Performance Evaluation, Quality of Service, Quality of Experiment, Multimedia Communications, and Topology Control Traffic Optimisation in Wireless Ad hoc Networks.


Abdelali Boushaba received his PhD degree in 2014 in Computer Sciences from Dhar Mahraz Sciences Faculty, Fez, Morocco. Received the MSc degree in Networks, Telecommunications and Multimedia from National Higher School of Computer Science and Systems Analysis (ENSIAS), Rabat, Morocco in 2007. Received the BSc degree in Computer Sciences from Dhar Mahraz Sciences Faculty in 2002. He is a researcher at the LIM laboratory of Dhar Mahrz Sciences Faculty. His research interests include Performance Evaluation, Quality of Service, Quality of Experiment, Multimedia Communications, and Topology Control Traffic Optimisation in Wireless Ad hoc Networks.

Azeddine Zahi received his PhD degree in 1997 in Computer Sciences from Mohammed V University in Rabat. He is a research professor at Sidi Mohamed Ben Abdellah University of Fez since 1995. His research interests include fuzzy data mining, heuristics and automatic learning.

Rachid Benabbou received his PhD degree in Computer Sciences in 1999 from the University Mohammed V of Rabat. He is a research professor at Sidi Mohamed Ben Abdellah University of Fez since 1997. His research interests focus on Ad hoc Network, Social Web and Communities of Practice.

Mohammed Oumsis received in 1999 his PhD degree in Computer Sciences from the Sciences Faculty of Rabat. Graduated in 2010 with a HDR in Computer Sciences. He is currently Research Professor in Computer Sciences at the High School of Technology, Mohammed V Agdal University. Currently, his research activity focuses on the problems at the Network and Application Layer of various Wireless Ad hoc Networks.

Said El Alaoui Ouatik received in 1999 his PhD degree in Computer Sciences from the Sciences Faculty of Rabat. Graduated in 2008 with a HDR in Computer Sciences. He is currently Research Professor in Computer Sciences at the Sciences Faculty of Fez.
A Distributed UWB-based Localization System in Underground Mines

Yuanqing Qin 1, Fang Wang 2, and Chunjie Zhou 1
1. Key Laboratory of Ministry of Education for Image Processing and Intelligent Control, School of Automation, Huazhong University of Science and Technology, Wuhan, China
2. School of Physics and Information Engineering Jianghan University, Wuhan, China
Emails: qinyuanqing@mail.hust.edu.cn, wangfang1979@jhun.edu.cn, cjiezhou@126.com

Abstract—The location of people, mobile terminals and equipment is highly desirable for operational enhancements in the mining industry. In an indoor environment such as a mine, the multipath caused by reflection, diffraction and diffusion on the rough sidewall surfaces are the main sources of range measurement errors. In this paper, a UWB time of flight based localization system is proposed to address the multipath effect in underground mines. To reduce the communication cost and time delay of localization in such a chain type wireless network, a distributed localization algorithm based on particle swarm optimization (PSO) is proposed and implemented on the blind node (the node to be localized). Without extra hardware needed, an accurate but low cost localization system has been achieved. Experimental results verify the proposed scheme.

Index Terms—Ultra Wideband; Wireless Sensor Networks; Localization; Particle Swarm Optimization; Underground Mines

I. INTRODUCTION

Underground mining operations are considered as hazardous industrial activity because of the poor ventilation/visibility, the danger of collapse/gas explosion, and the presence of toxic gas. Accidents often happen and cause severe casualty and capital lost. It is of great significance to establish advanced monitoring system, which can obtain the real-time information about the miners and the environment, especially the real-time location of the miners, to safeguard them in case of emergency.

In emergencies wireless communication may become vital for survival, for example, during a disaster (such as a fire, rock falls), the conventional wired communication system may become unreliable, necessitating a wireless radio system. In fact, the idea of using wireless underground sensor networks, can be traced back to [1]. The utilization of WSNs to implement the monitoring system often leads to a rapid and flexible deployment. Additionally, the multi-hop transmitting method conforms to the mines structure and provides more scalability for system construction. Another important reason for choosing WSNs to monitor mines is that it can also be used to localize the miners or equipments, where infrastructures needed for other localization method such as Global Position System (GPS) are not available.

The Ultra Wideband (UWB) technology has been the subject of extensive research in last two decades. It has emerged as a promising candidate for many wireless personal area network (WPAN) applications, sensor networks and ubiquitous computing. It has been also selected as a viable candidate for precise ranging and geolocation, due to the high time resolution (large bandwidth) of UWB signals, which enables accurate Time of Flight (TOF) measurement between nodes [2].

In this paper, we focus on designing localization system based on UWB wireless sensor networks in underground mines, however, it can be also generalized to different underground environments such as, underground city, tunnels, subway, etc.

The remainder of the paper is organized as follows: Section 2 reviews the localization techniques in mining industry. A UWB time of flight based localization algorithm is proposed in Section 3. Section 4 details the localization scheme and the implementation of the system. Section 5 demonstrates the experimental results and conclusions are presented in Section 6.

II. LOCALIZATION TECHNIQUES IN MINING INDUSTRY

A. Traditional Localization Method

The traditional localization techniques are based on a procedure of manual reporting of miner’s location (i.e., using talky–walky). However, the precision are limited to the knowledge of the level, gallery name, segment or section number where the miner is located. Infrastructure based wireless techniques, such as RFID, 802.11, are reported for localization in mine [3-5]. However, the precision depends on the deployment of RFID readers or access points, and the cable part makes the localization system less scalable.

B. Wireless Sensor Network Based Localization Schemes

Another kind of wireless network, ad-hoc wireless sensor network, has attracted extensive interests for monitoring in underground mines, due to its low cost, flexibility, and the capability of localization service. In this paper, we focus on the localization function of WSN.
Localization in WSNs refers to estimating the location of a target node (a node with unknown location, i.e., a blind node) according to the relationship between itself and several anchor nodes (nodes with known location). It has received considerable attention, as data collected from sensors makes sense only if the location of those sensors is known, and sometimes the location itself is the information of interest, such as the asset localization, target tracking etc [6].

Underground mines are quite special indoor environments, with lengths of tens of kilometers and widths of several meters. WSNs deployed there usually line or chain type networks with low density. Data transmission from sensor nodes to central server costs more energy because of the multi-hop manner. The network topology is dynamic with the advance of production. The surface of the tunnels is usually rough and the multi-path effect of radio propagation is severe. These factors make underground mines quite challenging environments for WSNs localization application.

With regard to the mechanisms used for estimating location, localization based on WSN can be divided into two categories: range-based and range-free. Reference [7] provided an extensive review of range measurement techniques and algorithms for WSN localization.

1) Range-based Localization Schemes

Range-based localization schemes consist two steps: 1) measurement of the relative distances or angles between the blind node and the anchor nodes; 2) estimation the position of the blind nodes with the information collected in the first stage. Several techniques have been researched for physical distances/angles measurement in WSN, such as the Angle of Arrival (AOA), the Received Signal Strength (RSS) and the Time Of Arrival (TOA) etc. In the second stage, trilateration or triangulation algorithms are often used to calculate the position of the target node. We will discuss it in section 3.

The necessary of antenna arrays for AOA mechanisms makes it impractical to implement while maintaining the low-cost demands of node. RSS is considered as useful information to estimate the distance, because it is a mandatory standard in IEEE 802.15.4, which means every node can get RSS when receiving a package, without additional hardware required. It has been used in underground mines [8,9]. The main challenge is that the RSS ranging method relies on a path loss model, which is very sensitive to the channel parameters, especially to multipath channel in indoor environment. TOA method measures the time of the flight of acoustics or radio signals, and then the distance between the two nodes can be calculated because the speeds of acoustics and radio in the air are known as constants. Acoustic TOA method can get sub-meter accuracy at the cost of additional acoustic hardware requirement on the node [10]. With the same transceiver used for communication, Radio Frequency Time of Flight (RF-TOF) is preferred for ranging in WSN. In narrowband systems, measuring RF-TOF requires accurately resolving the phase offset of a signal. Ref. [11] proposed a method for pair-wise ranging called code modulus synchronizaition that did not require either mote to determine the absolute phase offset of system clocks, the correlation function or the TOF in real time. Meter level accuracy could be obtained even in coal mine environment. Jennic embedded a 2.4GHz TOF engine on a SoC JN5148 [12], which could enable a low cost localization system be included on wireless sensor nodes [13].

UWB has been regarded as an ideal candidate for accurate indoor localization application due to its excellent time domain resolution, multipath immunity, and simultaneous ranging and communication capability. Commercial indoor UWB positioning system already exists. A precision-location UWB system has been developed which achieves sub-meter accuracy with Time Difference Of Arrival (TDOA) range measurement [14]. Combining TDOA with AOA measurement, Ubisense’s precise location system [15] can achieve 15-cm accuracy. However, the high accuracy was achieved in a much smaller coverage because of the synchronization requirement of the anchor nodes. Ref. [16] carried out a feasibility study of using UWB based-WSN as a future solution for localization in underground mines. In 2013, DecaWave company published a low cost UWB chip DW1000 [17], which could provide 10cm ranging accuracy. In this paper, we propose an accurate, low cost, large scale UWB localization system in underground mine based on this chip.

2) Range-free Localization Schemes

In order to address the multipath channel character and the None Line of Sight (NLOS), range-free localization scheme are also proposed in underground mines, such as DV-hop [18], weighted centroid localization [19] and fingerprinting scheme (or pattern recognition technique) [20, 21] etc. The former two connection-based might suffer from the localization precision due to the sparse character of WSN in underground mine. And the main disadvantage of fingerprinting method is the requirement that the training database should be large enough and representative of the current environment for accurate position estimation. It could be laborious or even impossible in practice in harsh dynamic underground mine environment.

III. UWB TOF-BASED LOCALIZATION ALGORITHM IN UNDERGROUND MINES

In this paper, we proposed a UWB TOF-based localization algorithm, which could be implemented in a distributed manner to improve the timeliness of localization and save energy. It consists of three stages: 1) the blind node finds out the anchor nodes in its communication range, and chooses three or four (corresponding to 2-dimensional or 3-dimensinal localization scenarios) of them that are not on a line as reference nodes; 2) the blind node measures the distances between the reference nodes and itself, and 3) the blind node estimates its location according to the coordinates of the reference nodes obtained in the first stage and the distances measured in the second stage.
A. Discovery of the Localization Reference Nodes

After joining the network, the blind node broadcasts a localization request in its communication range and starts a waiting timer. The anchor nodes that received the request reply with their coordinates and network IDs. When the waiting timer expires, the blind node checks if there are enough anchor nodes that can serve as localization reference nodes (3 or 4, not on a line). If true, the blind node will start the ToF measurement; otherwise, the blind node will broadcast the localization request and start the waiting timer again.

B. Distance Measurement based on UWB ToF

In this stage, the blind node performs TOF measurement with the chosen reference nodes one by one, as shown in Fig. 1.

1) The blind node sends a Poll message addressed to the current reference node and notes the send time \( T_{SP} \). The blind node listens for the Response message. If no response arrives after some period, the blind node will time out and send the poll again.

2) The reference node listens for a Poll message addressed to it. When the reference node receives a poll, it notes the receive time \( T_{SR} \), and sends a Response message back to the blind node, noting its send time \( T_{SR} \).

3) When the blind node receives the Response message it notes the receive time \( T_{SR} \) and set the future send time of the Final response message \( T_{SP} \), it embeds this time in the message before initiating the delayed sending of the Final message to the reference node.

4) The reference node receives the Final message and notes the receive time \( T_{RF} \). Now the reference node has enough information to work out the TOF between the blind node and itself according to (1).

\[
\text{TOF} = ((T_{SR} - T_{SP}) - (T_{SR} - T_{SP}) + (T_{SR} - T_{SP}) - (T_{SR} - T_{SR})) / 4
\]

5) The reference node reports the result to the blind node with ToF Report message. Multiplying the TOF by \( c \), the speed of the light (and the radio waves), the blind node gets the distance between the current reference node and itself. This ranging algorithm does not require clock synchronization between the two nodes and the average of the four trips time removes the effects of each end’s clock frequency differences.

C. Trilateration Based on Particle Swarm Optimization

With the distances and the coordinates information of the reference nodes, the blind node now can calculate its own position (in 2 dimensional scenarios):

\[
(x - x_i)^2 + (y - y_i)^2 = d_i^2, \quad i = 1, 2, \ldots, N \tag{2}
\]

where \((x, y)\) is the coordinates of the blind node, \((x_i, y_i)\) is the coordinates of the \(i^{th}\) reference node, \(d_i\) is the measured distance between the blind node and the \(i^{th}\) reference node, and \(N\) is the number of the reference nodes. In 2 dimensional scenarios, \(N\) needs to be greater than 2, while in 3D scenarios, \(N\) needs to be greater that 3. And the reference nodes should not be on a line, or the solution of (2) is not unique.

This geometric technique, called trilateration, yields ambiguous solutions in the presence of range error in the system, since the circles defined by (2) may intersect at multiple points due to the erroneous distance measurement, as shown in Fig. 2. A popular statistical localization algorithm is the Nonlinear Least Squares (NLS) techniques, by which the location of the blind node is calculated as follows:

\[
[x, y] = \arg \min_{x,y} s(x, y) = \arg \min_{x,y} \sum_{i=1}^{N} \beta_i (\sqrt{(x-x_i)^2 + (y-y_i)^2} - d_i)^2 \tag{3}
\]

where \(s(x, y)\) is the cost function, \(\beta_i\) represents a weighted coefficient for the \(i^{th}\) measurement, which commonly reflects the reliability of this measurement. The solution of (3) usually requires numerical search methods such as the steepest descent or the Gauss-Newton techniques, which can have high computational complexity and typically requires good initial value in order to avoid converging to the local minima of the cost function [7].

![Figure 1. UWB ToF ranging algorithm](image1)

![Figure 2. Trilateration yields multiple intersections due to the distance estimation error](image2)

To minimize the cost function, heuristic algorithms such as simulated annealing [22], Particle Swarm Optimization (PSO) [13], Bacterial Foraging Algorithm (BFA) etc. [23], were used in trilateration.
To reduce the computational complexity and enable the algorithm to be implemented in a distributed manner, a global version PSO algorithm is designed to estimate the position of the blind node. The algorithm is described as follows:

1) After obtaining coordinates of the reference nodes and the distances between the blind node and those reference nodes, the blind node can define a searching space for the current localization, shown as the dotted rectangle in Fig. 2, where

\[
\begin{align*}
  x_{\text{min}} &= \min_{i=1,2,\ldots,N} \{x_i - d_i\} \\
  x_{\text{max}} &= \max_{i=1,2,\ldots,N} \{x_i + d_i\} \\
  y_{\text{min}} &= \min_{i=1,2,\ldots,N} \{y_i - d_i\} \\
  y_{\text{max}} &= \max_{i=1,2,\ldots,N} \{y_i + d_i\}
\end{align*}
\]  

(4)

2) The blind node defines the fitness function as (3) and initializes M particles’ position according to (5),

\[
\begin{align*}
  x_j &= \text{rand}(1) \times (x_{\text{max}} - x_{\text{min}}) + x_{\text{min}} \quad j = 1, 2, \ldots, M. \\
  y_j &= \text{rand}(1) \times (y_{\text{max}} - y_{\text{min}}) + y_{\text{min}} \quad (5)
\end{align*}
\]

where \((x_j, y_j)\) is the position of the \(j^{th}\) particle, \(\text{rand}(1)\) generates a random number with uniform distribution in the range of \([0,1]\) and \(M\) is the number of the particles.

3) Each particle updates its position based on its own best exploration, the best swarm overall experience and its previous velocity according to the following model:

\[
\begin{align*}
  v_{jx}(k+1) &= w \cdot v_{jx}(k) + c_1 \cdot \text{rand}(1) \cdot \left[ pBest_{jx} - x_j(k) \right] \\
  &\quad + c_2 \cdot \text{rand}(1) \cdot \left[ gBest_{jx} - x_j(k) \right] \\
  v_{jy}(k+1) &= w \cdot v_{jy}(k) + c_1 \cdot \text{rand}(1) \cdot \left[ pBest_{jy} - y_j(k) \right] \\
  &\quad + c_2 \cdot \text{rand}(1) \cdot \left[ gBest_{jy} - y_j(k) \right]
\end{align*}
\]

(6)

\[
\begin{align*}
  x_j(k+1) &= x_j(k) + v_{jx}(k+1) \\
  y_j(k+1) &= y_j(k) + v_{jy}(k+1)
\end{align*}
\]

(7)

where \((v_{jx}(k),v_{jy}(k))\) is the current velocity vector of particle \(j\); \((v_{jx}(k+1),v_{jy}(k+1))\) is the velocity vector of particle \(j\) for the next iteration; \((x_j(k),y_j(k))\) is the current position of particle \(j\); \((x_j(k+1),y_j(k+1))\) is the position of particle \(j\) of the next iteration; \((pBest_{jx},pBest_{jy})\) is the best position particle \(j\) achieved based on its own experience during previous \(k\) iterations; \((gBest_{jx},gBest_{jy})\) is the best particle position based on overall swarm’s experience during previous \(k\) iterations; \(w\) is the inertia weight; \(c_1\) and \(c_2\) are two positive constants; \(\text{rand}(1)\) is a randomly generated number with uniform distribution in the range of \([0,1]\); and \(k\) is the iteration index. \(pBest\) and \(gBest\) are selected in terms of the fitness value calculated. The location estimation algorithm based on the global-best version of PSO is depicted in Fig. 3.

IV. DISTRIBUTED LOCALIZATION SYSTEM IN UNDERGROUND MINES

A. Structure of the Proposed Localization System

WSN deployed in underground mines has a chain type topology because of the special geographic restriction. The WSN deployed in underground mines has a chain type topology because of the special geographic restriction. The communication cost and time delay are relatively high due to the multi-hop transmission. A distributed localization scheme which can be implemented on the blind node is preferred with only the localization result reporting to the sink node.

The coordinator is responsible for establishing the network. It also acts as a gateway to the surveillance PC through a serial port. The surveillance PC is responsible for the configuration of the anchor nodes and localization data management.

Anchor nodes are routers of the ZigBee network. They collect data from the tunnel environment and participate in localization. When receiving a localization request from a blind node in one hop range, the anchor nodes respond it with their ID numbers and coordinates.

The blind node must be a ZigBee router, because it needs to communicate with multiple anchor nodes directly within its communication range. The blind node
carries out the localization algorithm. The proposed localization system structure can be directly applied to multiple blind nodes which are moving in the tunnels simultaneously, without any modification. It is because of the mesh topology of the ZigBee network.

B. The Deployment of the Anchor Node

To ensure the network communication is reliable and the blind node can find enough reference nodes that are not on a line, the anchor nodes should be deployed along both sides of the tunnel. The distance between any two adjacent nodes on the same side keeps equal, and is shorter than the valid communication range between two nodes. The anchor nodes on the opposite side should be placed alternately, in other words, one anchor node on one side is to be placed in the middle of two anchor nodes on the opposite side, as shown in Fig. 4.

Each anchor node has a unique ID number and coordinates which can be configured. The ID numbers of the anchor nodes on one side are odd numbers and those on the other side are even numbers. This rule helps the blind node to choose proper reference nodes on both sides, because if all the chosen reference nodes are on the same side, which means they are in a line, it will lead to a failure of our localization algorithm.

V. EXPERIMENTAL RESULTS

To test the performance of the proposed localization system, experiments were carried out in an abandoned air-raid shelter, with the similar environment characteristics to underground mines. The length in X axis is 120 meters and the width in Y axis is 4 meters respectively, as shown in Fig. 4.

A. TOF Ranging Accuracy Contrast Between UWB and Narrowband

To prove the advantage of UWB ranging method in mine environments, a contrast measuring experiment was carried out in a point to point way, using a pair of UWB evaluation boards EVB100 [17] and a pair of 2.4GHz narrowband JN5148 [12] nodes respectively. One was placed at the entry of the air-raid shelter and the other was moving along the air-raid shelter. At different test points, 20 times of UWB ranging and 20 times of narrowband ranging were performed. The average and the standard deviation of the measuring results are shown in Table 1.

<table>
<thead>
<tr>
<th>Real distance(m)</th>
<th>UWB ranging</th>
<th>Narrowband ranging</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average(m)</td>
<td>Standard deviation(m)</td>
</tr>
<tr>
<td>10</td>
<td>3.86±9.74</td>
<td>1.15±0.13</td>
</tr>
<tr>
<td>20</td>
<td>29.85±9.74</td>
<td>2.16±0.16</td>
</tr>
<tr>
<td>30</td>
<td>49.78±9.69</td>
<td>3.21±0.21</td>
</tr>
<tr>
<td>40</td>
<td>59.7±9.68</td>
<td>4.19±0.23</td>
</tr>
<tr>
<td>50</td>
<td>69.7±9.68</td>
<td>5.19±0.23</td>
</tr>
</tbody>
</table>

As can be seen from the result, the Line of Sight (UWB) ranging errors based on UWB were much smaller than those based on narrowband. It is because the UWB signal has the character of excellent time domain resolution and the immunity to multipath effect in indoor environment. This provides a good foundation for establishing an accurate localization system in underground mines.

B. Localization Experiment

The deployment of WSN in the shelter is shown as Fig. 5. The distance between the adjacent anchor nodes on the same side was 20 meters, and the deployment of the two sides was alternate. The ZigBee localization network consisted of one coordinator and 13 routers, which were all based on DW1000 UWB chip. One of the routers acted as the blind node, which was localized in real time. A localization software was designed on the surveillance PC. Fig. 6 showed the user interface of the software, through which users could configure the parameters of the anchor nodes, including the ID and the coordinates, and inquire the real-time position of the blind node.

The PSO parameters used in the experiment were as follows:

- Population of particles $M = 10$ and the target fitness value $f_j = 0.3$
- The coefficients $c_1 = c_2 = 1.494$
- The inertia weight $\omega$ is decreased linearly from 0.9 in the first to 0.4 in the last iteration, i.e.
  \[ \omega(k) = 0.9 - k \times 0.5 \]
- The weighted coefficient for the $i^{th}$ measurement $\beta_i$ was defined as follows:
  \[ \beta_i = \frac{r_{SS_i}}{\sum_j r_{SS_j}} \]

where $r_{SS_i}$ is the RSS value obtained when the blind node measures RF-TOF between the $i^{th}$ reference node and itself.

At each test point, the localization error (the distance between the estimated position and the real position of the blind node) was less than 0.3 meter.

DW1000 is a fully integrated single chip UWB low-power low-cost transceiver IC compliant to IEEE802.15.4-2011, with transmitting current 70mA, receiving current 30mA and supply current in SLEEP mode.
mode $2 \mu$A[17]. The blind node can work in low duty cycle mode so that it can be battery powered. But the other ZigBee routers (i.e. the anchor nodes) need to be supplied with main power in long-term run because they cannot be put into sleep mode when idle. Schedule-based network protocols taking localization task into consideration may address this problem.

Several research challenges remain to be addressed. None line of sight propagation is not taken into account in our scheme, which can be a main cause of range error. Time delay of communication is another issue that should be addressed when the scale of the system becomes bigger.

ACKNOWLEDGMENT

This work was supported in part by NSFC grants 61272204 and “the Fundamental Research Funds for the Central Universities”, HUST: 2013QN149.

REFERENCES


Yuanqing Qin received the B.Sc. degree in applied electronic techniques from Shandong Institute of Science and Technology in 2000, and M.Sc. and Ph.D. degree in control science and control engineering from Huazhong University of Science and Technology (HUST) in 2003 and 2007 respectively. Currently, he is a lecturer of control science and control engineering at HUST, Hubei, China. His current research interests include wireless sensor network, industrial field bus, machine vision and networked control system.

Fang Wang received B.Sc., M.Sc., Ph.D. degrees from Shandong Institute of Science and Technology, Northwest University of Polytechnique, Harbin Engineering University in 2000, 2003 and 2011. Currently, she is a lecturer of communication engineering at Jianghan University, Hubei, China. Her research interests include wireless communications and wireless sensor networks.

Chunjie Zhou received the MS and PhD degrees in control theory and engineering from Huazhong University of Science and Technology, Wuhan, China, in 1991 and 2001, respectively. He is currently a professor at School of Automation, HUST. His research interests include industrial communication, artificial intelligent, theory and application of networked control system.
Wi-Fi-based Positioning in a Complex Underground Environment

Matteo Cypriani, Gilles Delisle, Life Fellow, IEEE, Nadir Hakem, Member, IEEE
Laboratoire de Recherche Télébec en Communications Souterraines (LRTCS)
Université du Québec en Abitibi-Témiscamingue (UQAT), Val-d’Or, QC, Canada
E-mail: surname.name@uqat.ca

Abstract—Underground mining tunnels constitute a very particular environment for radio wave propagation – with characteristics of both indoor and outdoor “regular” environments. This paper shows how a general-purpose Wi-Fi-based indoor positioning system, OwlPS (Owl Positioning System) was adapted to work in that particular environment. A series of experiments, conducted in a formerly exploited gold mine at 70 metres below the surface, across about 400 metres of drifts, is then introduced; it primarily aims at determining the positioning accuracy that can be reached in such a context with a Wi-Fi-based positioning system using the signal strength at 2.4 GHz. The results obtained are improved with a filter, and the mean Euclidean distance error is under 10 metres in most cases when the terminal is carried by an operator; this makes OwlPS usable as is for asset management and emergency positioning of workers underground.

Index Terms—positioning, localization, underground, mines, IEEE 802.11, Wi-Fi, self-calibration, autocalibration

I. INTRODUCTION

Mining tunnels constitute a unique environment regarding the propagation of radio signals. Propagation through rock depends on the frequency and on the material’s conductivity, and is very limited when it comes to microwaves. At 2.4 GHz, for two devices to be in coverage, they must most of the time be in the same tunnel or even in the same section of a tunnel. On the other hand, tunnels sometimes act as waveguides [1]–[3] that make the signal propagate on longer distances; this is mainly dependent on the radio frequency, the characteristics of the walls and the width and height of the tunnels, along with the signal’s polarisation direction.

This paper presents a series of experiments that aim at studying the feasibility of a positioning system based on 2.4 GHz IEEE 802.11 (Wi-Fi) operated underground. This work is justified by the fact that it is extremely difficult to localise people and equipment in the complex tunnel network of a real underground mine. Such a positioning system is to be used for normal operations such as asset management, but could become vital in case of emergency situations.

Experiments were conducted at the CANMET laboratory mine (Canada Centre for Mineral and Energy Technology), which is a formerly exploited gold mine near Val-d’Or, in north-west Quebec. The deployment uses about 400 metres of galleries at 70 metres under the surface. The experimental results presented in this paper extend some of the preliminary results published in [4].

The Owl Positioning System [5] was used for all the experiments; it is a free open-source software that was originally released by the University of Franche-Comté (France). It is a Wi-Fi-based positioning system that targets generic indoor environments, tested mainly in office and residential contexts. It implements several positioning techniques and algorithms based on propagation models, on location fingerprinting, or on both. It also features a self-calibration (or autocalibration) mechanism that generates reference points to be used by fingerprinting-based algorithms the same way real measurements would be. This mechanism was used in previous experiments [6], [7] to generate a regular, rectangle-shaped meshing of reference points. Obviously this operating mode, whereas well suited for use in buildings, does not work as is in a network of long and narrow tunnels. Therefore, the self-calibration mechanism has been modified to enable generating the reference points along a path following the tunnels.

With the change of environment, the question of the localisation space is raised. If we define a two-dimension coordinate system suitable for a tunnel, the first dimension is longitudinal to the tunnel’s orientation, and the second dimension is perpendicular, i.e. determines the mobile’s position in relation to the walls. In classical drift mines, the tunnels are narrow enough that it does not make sense to take this second dimension into consideration outside the scope of high-precision positioning systems used for use cases such as machine guidance. In this work, it has therefore been elected to localise in only one dimension, but to keep providing the positions in a regular 2-D coordinate system – switching to a 1-D system would be burdensome and bring few advantages, if any.

This paper is organised as follows. Section II extends the introduction by presenting some work from the literature, then section III briefly introduces the Owl Positioning System and explains the changes that have been made to the autocalibration mechanism. The characteristics of the deployment area, the hardware, as well as the network and software configuration are detailed in section IV. The measurement scenarios are presented in section V and the results obtained are discussed in section VI. Finally, section VII explains how the results can be improved by...
implementing a filter.

II. RELATED WORK

Indoor localisation has been the focus of many research teams over the past decade. Although fewer teams have underground localisation in mind when developing a positioning system, several technologies have been experimented as well.

In [8], the authors studied the geomagnetic field in a copper and zinc mine and noticed that it presents enough anomalies that it can be used as the support of a positioning system. The proposed system uses the Monte-Carlo localisation method. Its main advantage is that no infrastructure is required, so positions can be obtained even in case of power failure. On the other hand, it requires an extensive fingerprinting of the deployment area, which is even more problematic in a daily-changing environment such as a mine than it is in a regular indoor environment. Moreover, it may not be usable in a zone that does not feature “a complex distribution of metallic minerals in the upper lithosphere.”

Dead reckoning has also been explored; in [9], a positioning system composed of accelerometers, a gyroscope, a speed sensor and distance sensors mounted on a vehicle is considered. The distance sensors provide the distance of the vehicle to the walls of the tunnel, while the other sensors allow to accurately estimate the distance travelled. The initial fix is provided by short-range radio beacons. The authors announce a very good accuracy, the distance error being below 1 metre in most cases, but the results were obtained only by simulation.

For specialised, high-precision tasks, total stations are frequently used; it has been shown in [10] that the iGPS system commercialised by Nikon Metrology [11] could also be suitable for such use.

Of course, several teams also concentrate their efforts on radio-based systems. In [12], a solution using a leaky cable that transmits a signal at 2.4 GHz and receives a signal at 1.2 GHz from the mobile terminal is proposed. The position is computed by measuring the time difference between both signals. Although this system does not seem to target specifically a mining tunnel environment, it seems to offer a good accuracy when tested on a length of 100 metres in an underground facility, where the distance error ranges from 0.2 to about 3 metres in most cases and does not exceed 8.1 metres. One of the drawbacks of this system is that it requires proprietary hardware both on the infrastructure side and on the mobile terminals.

In [13], the authors present an extensive study of the feasibility of a positioning system based on the time of flight of the radio waves in a ZigBee network deployed in an underground gallery. They conduct tests in two kinds of tunnels and conclude that the accuracy of such a system would be within 20 metres. In [14], the authors announce a precision of 5 to 15 metres with a Wi-Fi-based positioning system, but the experiments did not take place in an actual mine; this system uses a different positioning technique, since it uses a manually-calibrated, mobile-centred fingerprint of the signal strength. A similar technique is used in [15] for location-based services in subway stations.

Compared to these solutions, the system presented in this paper has the advantage of being low-cost thanks to the use of off the shelf Wi-Fi devices and low-maintenance thanks to its self-calibration mechanism. Moreover, the infrastructure devices can be used to provide data access to mobile devices in the mine – or a mine’s existing wireless access points could be used as an infrastructure devices for the positioning system (assuming the hardware is compatible).

There is a significant number of other related papers, but most of these are either theoretical or are set up in environments that are much easier to deal with than an underground mine. The studies available for practical applications of location-based services in mining environments are practically inexistent, hence reinforcing the need for our study.

III. POSITIONING SYSTEM

The details of the Owl Positioning System have been given in previous publications and the next section solely recalls its major concepts and describes its main architecture; the details can be found in [16], [17] and [7]. Section III-B presents the principle of OwIPS’ autocalibration mechanism as it was implemented until now and explains the changes made to adapt it to the new environment of underground tunnels.

A. General presentation

As stated in introduction, OwIPS is based on the Wi-Fi (IEEE 802.11) wireless data network; it uses the signal strength as the main information to compute positions. It is infrastructure-centred, which means the hardware and software components of the infrastructure deployed by the system administrator are in charge of measuring the signal strength from the mobile terminals and to compute their coordinates. More precisely, this infrastructure is composed of three main kinds of elements:

- the capture points (CPs), that receive (capture) the positioning requests transmitted by the mobile terminals and measure the signal strength;
- the aggregation server, that collects the captured packets from the capture points and consolidates those corresponding to the same positioning request into a single data structure;
- the positioning server, that computes the positions of the mobile terminals from the aggregated requests.

The role of the mobile terminals is simply to transmit positioning requests, each request being made up of a bunch of UDP packets.

B. Autocalibration

In the context of location fingerprinting-based positioning, when using a manual calibration, a reference point (or fingerprint) is a set of measurements from several
capture points, associated to the coordinates of the mobile terminal at the moment it transmitted the calibration request. When autocalibration is used, OwlPS generates reference points that contain pseudo-measurements extrapolated from real measurements. The autocalibration mechanism works by measuring the signal between the capture points in coverage to deduce the quality of the signal in the corresponding areas, given the distances between the capture points. These measurements are done by having the capture points transmit so-called autocalibration requests, that are similar to the positioning (and calibration) requests sent by the mobile terminals. Once a measurement is done between two capture points, the Friis transmission equation is used to calculate the path loss exponent (or Friis index) for this link; then, using the same formula, the signal strength at various locations between the two capture points can be extrapolated.

The generation process of reference points in a wide area such as an office environment or an open space has been described in detail in [17], and in deeper detail in [7]. To be adapted to an environment where the mobile terminals evolve most of the time between two capture points, the autocalibration mechanism had to be redesigned to generate reference points only along a path that includes the coordinates of the capture points. The new process is much simpler than the older one and obeys the following rules to generate a reference point $RP$:

1) If $RP$ is between two capture points, it is assumed that the two capture points are in coverage, and $RP$ will contain measurements for both. The signal quality between the two capture points is used to extrapolate the signal strengths in $RP$.

2) If $RP$ is at one of the extremities of the path and is not between two capture points, it will only contain measurements for the nearest capture point. The signal quality between this capture point and the second nearest capture point in coverage is used to extrapolate the signal strengths in $RP$.

For example, on the deployment map (Fig. 1), a reference point between $CP_2$ and $CP_3$ includes signal strengths from both $CP_2$ and $CP_3$, and the signal level is extrapolated from measurements corresponding to the autocalibration requests transmitted by these two capture points. Although $CP_1$ is also in good coverage in this section of the tunnel, it is not taken into account. A reference point between the emergency exit and $CP_6$ includes signal strengths only for $CP_6$, and the signal level is extrapolated from autocalibration requests sent by $CP_3$ and received by $CP_6$.

Despite the existence of more accurate propagation models for underground tunnels, it has been elected to rely only on the free-space Friis transmission equation. In [18], Y. P. Zhang shows that tunnel propagation is similar to free-space propagation up to a certain distance, and then follows a lossy dielectric waveguide model. The location of this breakpoint depends on several factors, but the examples provided by Y. P. Zhang show that it is located several dozens of meters away from the transmitter. This model, although interesting, is of impractical use for the purpose of autocalibration, because it requires to characterise accurately the environment – tilt and cross-sectional dimensions of the tunnel (which, in mines, can vary even in a single tunnel) – as well as the propagation losses due to wall roughness and refraction.

IV. EXPERIMENTAL SET UP

A. Deployment area

The deployment area is a part of the level -70 metres of the CANMET laboratory mine, near Val-d’Or (Quebec), pictured in Fig. 1. The average width of the tunnels is between 2.50 and 3 metres, with some exceptions such as larger rooms or recesses. The walls are very irregular, and the actual width can vary by several dozens of centimetres depending on where it is measured.

The drifts are fairly long, each of the three main sections used in this experiment being around 130 metres long. They are relatively straight, but turn enough that it is easy to loose the line of sight between two devices that are a few dozens of metres apart.

In some tunnels, “doors” made of flexible plastic blades hanging from the ceiling are installed; they are represented by short gray lines on the map. The short black line at the south of $CP_3$ is a wooden wall with a door and a plastic tarp window, that closes a small heated room used as an office when experimenting at this level of the mine.

For the sake of clarity, some important coordinates (measurement points) were defined; they appear in Fig. 1 under the form $MP_i$, where $i$ is an integer number.

B. Hardware

Five capture points are used, all being small routers of the MikroTik brand but with slightly different models: $CP_1$, $CP_2$ and $CP_3$ are RouterBoard 433 (which have 3 Ethernet connectors), and $CP_5$ and $CP_6$ are RouterBoard 411. The mobile terminal is a RouterBoard 411AH powered by a portable 12 V battery (actually a car battery booster) that provides 110 V AC power plugs. All these devices are equipped with Mini PCI MikroTik RouterBoard R52Hn IEEE 802.11a/b/g/n cards (Atheros AR922X chipset) and the transmission power is set to 25 dBm. Since only 802.11b/g is used, a single 3 dBi omnidirectional antenna is connected to one of the two antenna connectors of each card.

The aggregation server is a laptop running Debian GNU/Linux, whereas all the other devices run the OpenWrt embedded Linux distribution [19] in its version 12.09.

The coordinates of the capture points are displayed on the map (Fig. 1); they are all attached near the ceiling between 2.20 and 2.40 m above the floor (Fig. 2), except for $CP_3$ which sits on a block of concrete at 0.70 m above the floor. As can be seen on the map, they are not always located in the lateral centre of the tunnel, as their installation was dependent on the hardware available on the ceiling at each location. The aggregation server is in the small heated room at the south of $CP_3$. 

© 2015 ACADEMY PUBLISHER
Given the type of environment and the distances between the capture points, it is obvious that they cannot all be in coverage of the aggregation server; therefore, it was elected to install a wired network to allow the capture points to communicate with the server. However, the capture points still need to be able to communicate over the radio network to transmit autocalibration requests; for this purpose, an ad hoc network was set up and each capture point is configured to transmit its autocalibration requests to another capture point in coverage.

The mobile terminal also needs a destination host for its positioning requests. While it would have been feasible to use the same ad hoc network as for the capture points, it would have required to set up routing between the capture points. In order to avoid a complex configuration, a simpler solution was opted for: a smartphone running CyanogenMod 10.1.0 was configured to be a Wi-Fi access point serving a BSS (IEEE 802.11 Basic Service Set, i.e. a "Wi-Fi network") on the same channel as the capture points (channel 6, 2437 MHz). The mobile terminal is connected to this BSS and the destination of its positioning requests is set to the smartphone's IP address. The smartphone is configured to communicate as well to the aggregation server, and the aggregation server requests to send position beaconing messages. The smartphone powers CP1, CP2, CP3, and CP5, and works as a communication hub for CP1 and CP2, which are powered by PoE adapters plugged next to CP3. Since CP1 and CP2 have multiple Ethernet sockets, no switch is needed to connect them to each other and to CP3. The wiring and the direction of the wireless communication is shown in Fig. 3. A Power over Ethernet (PoE) switch allows them to communicate as well to CP3. Since CP3, CP4, and CP5, do not exchange requests, even though they are in coverage, it is not necessary to connect them to each other and to CP3.
D. OwlPS configuration

The mobile terminal is set to transmit a positioning request of 20 packets approximately every second (10 ms between two packets of the same request, 800 ms between two requests).

The autocalibration requests transmitted by the capture points also contain 20 packets, but they are separated by 25 ms. Only one request is transmitted every second for all the capture points: since there are five capture points, a given capture point transmits a request every 5 seconds.

The positioning server is configured to use the so-called nearest neighbour (NN) algorithm [20], also called “nearest in signal strength space” (NSS), which simply searches the database of reference points to find the one that is the most similar to the measurements corresponding to the last positioning request sent by the mobile terminal. The similarity function, which actually compares two sets of measurements, is based on a Euclidean distance of the mean signal strength of the two sets.

The autocalibration mechanism generates reference points along a path constituted by the following devices’ coordinates and measurement points (cf. Fig. 1): $MP_{61}$, $CP_6$, $CP_5$, $CP_2$, $CP_1$, $CP_5$, $MP_{52}$, and $MP_{51}$. Unless stated otherwise, the distance between two generated reference points is set to 1 metre for all the results presented in the next sections.

V. Scenarios

The results presented in this paper come from two scenarios in which the terminal is moved along the tunnels.

In the first scenario, a human operator carrying the terminal walks in the lateral centre of the drifts, starting from the south-west end of the deployment area ($MP_{61}$), goes all the way along the covered sections of tunnel and stops under $CP_5$.

The second scenario is similar to the first one, but the mobile terminal is mounted on a small mine transporter (2 m long, 1.40 m large and 1.90 m high). The terminal is attached at 1.70 m above floor level, and is 0.30 m away from the left side of the vehicle (Fig. 4). Since it is too big to go at the very end of the tunnel near the emergency exit ($MP_{61}$), the vehicle starts from the western ore loading point ($MP_{62}$). It then follows the same path as in the first scenario, but it continues past $CP_5$ until $MP_{51}$.

In the sequel, these scenarios are referred as “human” and “vehicle” scenarios.

VI. Results

A. Survey of the propagation in the mine

The “human” scenario covers the majority of the deployment area, and therefore the data generated can provide an interesting overview of the signal propagation quality in the tunnels. Fig. 5 and Fig. 6 display respectively the percentage of packets received from the mobile and the average signal strength per positioning request for each of the five capture points, in function of the mobile terminal’s position along its path. Please note that each step represents a transmitted positioning requests and, despite the fact that each step is indeed an indicator of the distance covered, it cannot be translated to an actual distance in metres because the mobile terminal’s speed is not rigorously constant.

The number of packets received is extremely variable but remains quite high as long as the mobile and the capture point are in line of sight. It drops quickly as soon as the line of sight is lost. The signal strength seems to be a far better indicator of the distance between two devices.
TABLE I

<table>
<thead>
<tr>
<th>Distance (m)</th>
<th>0.5</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>13.37</td>
<td>13.21</td>
<td>13.20</td>
<td>14.11</td>
<td>15.61</td>
<td>17.14</td>
<td>22.04</td>
<td>30.39</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>12.89</td>
<td>12.90</td>
<td>13.03</td>
<td>12.67</td>
<td>12.49</td>
<td>14.17</td>
<td>18.46</td>
<td>25.10</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>63.25</td>
<td>63.74</td>
<td>63.74</td>
<td>69.29</td>
<td>69.29</td>
<td>69.29</td>
<td>80.46</td>
<td>128.18</td>
</tr>
<tr>
<td>75th percentile</td>
<td>18.62</td>
<td>18.24</td>
<td>17.48</td>
<td>17.92</td>
<td>20.33</td>
<td>23.51</td>
<td>31.26</td>
<td>40.39</td>
</tr>
<tr>
<td>90th percentile</td>
<td>28.84</td>
<td>28.80</td>
<td>28.10</td>
<td>28.78</td>
<td>32.03</td>
<td>32.59</td>
<td>44.41</td>
<td>56.16</td>
</tr>
</tbody>
</table>

Indeed, the curve for each capture point has a Gaussian-like shape and the peak corresponds approximately to the moment when the mobile is the closest to the capture point. It can be observed that the signal drops more quickly once the mobile passes a capture point than it raises when the mobile approaches it; this indicates that the operator’s body has a significant impact on the signal transmitted by the mobile, probably increased by the fact that the antenna is very close to the operator’s abdomen. This observation is confirmed by another test in which the operator walks in the opposite direction, i.e. from $CP_2$ to the emergency exit (results not shown here): there again the signal is weaker in the direction opposite to the operator’s movement (i.e. towards his back).

Note that $CP_2$ seems to receive less packets and with a weaker signal than the other capture points. This is not only true for the signal from the mobile: an analysis of the signal between the capture points show that the signal is weaker between $CP_2$ and $CP_3$ than between $CP_1$ and $CP_3$. This may be due to its position, since it is closer to the ceiling than the other capture points: the ceiling’s texture can impact the signal and cause partial loss of line of sight. This may also be due to the topology of the tunnel around this location, or could simply come from a hardware defect. Whatever the reason, the weaker signal of $CP_2$ combined with the attenuation due to the operator’s body causes inconsistent coverage between the mobile and $CP_2$, especially when the mobile approaches $CP_3$. This is disruptive to the quality of the positioning, as will be shown in section VI-C.

It is also notable that the signal strength is sometimes higher than 0 dBm when the mobile terminal is close to a given capture point. This is probably due to the sensitivity of the IEEE 802.11 chips used, but leads to somewhat unrealistic values given the transmission power and the antenna gain of the devices.

B. Influence of the distance between the generated reference points

One of the most important parameters of the new auto-calibration function is the distance between two generated reference points. In order to know the real impact of this parameter, positioning results for the “human” scenario were generated with different distances; these results are presented in Table I. Similar results, not shown here, were obtained with a scenario in which the terminal stays at the same position at the end of a tunnel.

It appears that, as long as it is between 0.5 and 5 metres, the distance between two generated reference points has very little impact on the quality of the positioning. It begins having a visible effect starting from 10 metres, although in some cases increasing the distance has the positive side effect that reference points that are closest to the real position are chosen more often.

These results also show that the similarity function (cf. section IV-D) is stable enough and provides consistent results.

C. Measurements in mobility

Table II shows the detailed results for each stretch of the “human” scenario, as well as the overall scores. Table III does the same for the “vehicle” scenario; in this scenario, the vehicle stops next to each capture point for some time, which allows to show the difference of error when the mobile terminal is still and when it is in movement. Fig. 7 and Fig. 9 show (with different scales) the Euclidean distance error for each positioning request sent by the mobile along the path of both scenarios. In

Fig. 4. Photograph of the mobile terminal mounted on the mine vehicle

© 2015 ACADEMY PUBLISHER
parallel, Fig. 6 and Fig. 8 display the mean signal strength for the same requests.

In the “human” scenario, the results are generally acceptable, except for two stretches: between the emergency exit and $CP_6$, and between $CP_2$ and $CP_1$. In the former one, it can be observed on Fig. 6 that the signal received by $CP_6$ increases very slowly as the mobile terminal progresses towards the capture point, then abruptly reaches high values a few metres away from it. When comparing the curves of Fig. 6 and Fig. 7, it is clear that there is a correlation between the signal strength and the error in this part of the scenario. Regarding the section between $CP_2$ and $CP_1$, the results are very poor when the mobile terminal is close to $CP_2$, and improve as it moves towards $CP_1$. A plausible explanation is the weak signal received by $CP_2$ from the terminal and the intermittent coverage, which was already noted in section VI-A; since the coverage is so bad between $CP_2$ and $CP_1$, the autocalibration could not fully compensate for this difference of signal strength. However, as the mobile approaches $CP_1$, the signal received by $CP_2$ is less and less significant for the similarity function, hence the observed improvement in accuracy.

TABLE II

<table>
<thead>
<tr>
<th>Start position</th>
<th>exit</th>
<th>$CP_6$</th>
<th>$CP_3$</th>
<th>$CP_2$</th>
<th>$CP_1$</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>20.42</td>
<td>8.10</td>
<td>5.22</td>
<td>21.62</td>
<td>11.19</td>
<td>13.21</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>14.74</td>
<td>5.48</td>
<td>4.11</td>
<td>16.93</td>
<td>8.87</td>
<td>12.90</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.01</td>
<td>0.01</td>
<td>0.34</td>
<td>1.01</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>Maximum</td>
<td>63.74</td>
<td>20.18</td>
<td>19.59</td>
<td>61.56</td>
<td>33.34</td>
<td>63.74</td>
</tr>
<tr>
<td>50th percentile</td>
<td>19.12</td>
<td>7.53</td>
<td>4.15</td>
<td>17.69</td>
<td>10.38</td>
<td>9.60</td>
</tr>
<tr>
<td>75th percentile</td>
<td>25.24</td>
<td>12.11</td>
<td>7.14</td>
<td>28.71</td>
<td>17.06</td>
<td>18.24</td>
</tr>
<tr>
<td>90th percentile</td>
<td>37.12</td>
<td>16.47</td>
<td>11.08</td>
<td>49.43</td>
<td>22.91</td>
<td>28.80</td>
</tr>
</tbody>
</table>

© 2015 ACADEMY PUBLISHER
TABLE III

<table>
<thead>
<tr>
<th></th>
<th>CP1</th>
<th>CP2</th>
<th>CP3</th>
<th>CP4</th>
<th>CP5</th>
<th>CP6</th>
<th>CP7</th>
<th>CP8</th>
<th>CP9</th>
<th>CP10</th>
<th>CP11</th>
<th>CP12</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1</td>
<td>6.3</td>
<td>8.0</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP2</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP3</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP4</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP5</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP6</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP7</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP8</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP9</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP10</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP11</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP12</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Stop position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP2</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP3</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP4</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP5</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP6</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP7</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP8</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP9</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP10</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP11</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CP12</td>
<td>6.2</td>
<td>7.8</td>
<td>6.2</td>
<td>3.0</td>
<td>1.9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
</tbody>
</table>

In the “vehicle” scenario, the results are globally worse than when the mobile is carried by a human operator. Except for when the vehicle is next to a capture point, in which case the results are very good and very consistent, the error reaches very high values. Once again, the worst section is between CP2 and CP1, but unlike in the “human” scenario, the best one is between the start position and CP6 – although we must be careful in this comparison, since the terminal starts from MP61 in the “human” scenario and from MP62 with the vehicle, the latter being closer to CP6. A comparison of the signal strength in both scenarios (Fig. 6 and Fig. 8) can help understand why the results are so poor in the “vehicle” scenario. Indeed, the shapes of the curves are quite different: in the latter scenario, the signal strength for a given capture point stays low for a long time, with important drops of coverage, and then increases abruptly when the terminal is close to the capture point, in a way similar to the signal strength for CP2 in the “human” scenario. This can be explained by the higher height of the terminal, since as noted earlier the irregularities of the ceiling can impact the signal. Also, the vehicle is made of metal and the terminal’s antenna almost touches it (cf. Fig. 4); attaching the terminal so that the antenna is above the vehicle’s roof could help, but it could not be done because of height constraints.

As can be observed on the graphs, the results are quite variable in both scenarios, with peaks of error that are sometimes very high even in areas where the accuracy is otherwise good.

VII. IMPROVING THE RESULTS WITH A FILTER

The observations of the previous section, especially those on the high variability of the results, tend to indicate that the results could be improved by some kind of filtering.

To verify this assumption, a basic filter was implemented, that uses the maximal movement speed of the mobile terminal as the main parameter; this speed is provided by the user as an option of the positioning algorithm.
server. For each new position computed, the movement speed since the last position – which can itself be the result of filtering – is calculated, taking into account the timestamps of the two corresponding positioning requests. If the observed speed exceeds the maximum speed, the filtered position is the interpolation of the previous position in the direction of the new unfiltered position, respecting the maximum speed. In other words, the mobile is assumed to have moved in the same direction, but at a lower speed (Fig. 10).

Another observation that was made about the unfiltered results is that the accuracy is very good when the mobile terminal is in the neighbourhood of a capture point; in these circumstances, applying a filter actually worsens the accuracy instead of improving it. The most obvious option would be to completely disable the filter when the unfiltered position is found to be close enough to a capture point; the chosen option, which gives better results in some extreme cases where the computed position is close to a capture point that is in fact far away from the real position, is to apply an alternative maximum speed rather than completely disabling the filter. This alternative speed must be much higher than the regular maximum speed – five to ten times as much seems to be a reasonable value. In these experiments, it was found that applying the alternative maximum speed when the mobile’s unfiltered position is within 15 m of a capture point gave the best results overall, but of course a different deployment might require a different value.

Table IV and Fig. 11 present the results of the “human” scenario exhibit a decent accuracy, and it was shown that the filter cannot compensate for a prolonged period of bad results.

If the third stretch is ignored (between $CP_2$ and $CP_1$, where the very high error of the unfiltered results is caused by the poor signal received by $CP_2$), then the overall mean error drops from 11.34 to 10.16 metres. If the first stretch is also eliminated, the mean error for the scenario drops to 6.74 metres.

Regarding the “vehicle” scenario, the filter only allows to eliminate the biggest peaks of error, but does not improve significantly the average results: with the filter configured with a maximum speed of 12 km/h, an alternative maximum speed of 120 km/h, and a distance of 15 metres to apply the alternative speed, the maximum error drops from 140.72 to 112.28 metres, and the 90th percentile from 65.68 to 59.71, but the mean error is only lowered by less than a metre. This confirms that the filter is inefficient when the raw positioning results are bad for an extended period of time.

VIII. Conclusion

In this paper, it was shown that OwlIPS, a general-purpose indoor positioning system, could be adapted to the specific environment of underground mining tunnels with limited modifications. Through a series of experiments, the propagation of 2.4 GHz radio waves in the target environment was studied, as well as the accuracy of positioning. The results show that the IEEE 802.11 (Wi-Fi) data network is a good medium for a low-cost underground positioning system, allowing for an acceptable error with a limited number of devices.

Although some work is needed to improve the positioning of a vehicle, the filtered results for the “human” scenario exhibit a decent accuracy, and it was shown that the accuracy could be very much improved by a better deployment. Indeed, the results obtained show that the choices made in this deployment push the limits of the coverage of the capture points, which appears not ideal: a commercial deployment would benefit from a slightly increased number of capture points.

With an error rarely exceeding 20 metres\(^1\), the accuracy obtained is still acceptable for the two use cases

---

\(1\) For comparison, the mean error obtained with the same positioning system in an office environment was below 5 metres in most cases in former studies [7], [16].
envisioned – emergency positioning of workers and asset management. Indeed, while such an accuracy may not be sufficient to provide the workers with a fault-free navigation system, it is enough for the rescue team to know where to go to find a worker in case of an accident. Mine workers are specifically trained to react to the eventuality of a fire, but in that case too a positioning system can help the rescuers to know which refuge each worker could reach, or if they are trapped in a dead-end. As stated in the description of the “human” scenario, the operator walks in the lateral centre of the tunnel; however, a degradation of the accuracy when the mobile terminal is located close to a wall was not observed in preliminary tests, which is important in cases where a worker uses a wall as a guidance in low visibility conditions or when several workers walk side by side.

However, a better accuracy is always desirable, for these use cases but also for other ones such as remote machine guidance. The accuracy could be improved by working on the filter, which is quite basic; for instance, instead of abruptly switching from one maximal speed to another when the mobile is found to be next to a capture point, it would be possible to change the speed proportionally to the distance to the closest capture point. Similarly, it would be possible to detect the variability of the results and change the behaviour of the filter when a high variability is observed. Another way to improve the accuracy would be to enable the positioning server with a degradation of the accuracy when the mobile terminal is located close to a wall was not observed in preliminary tests, which is important in cases where a worker uses a wall as a guidance in low visibility conditions or when several workers walk side by side.

In this work, only the necessary modifications were applied to the autocalibration mechanism to make it work in a network of tunnels, but there is room for improvements. For example, as explained in section III-B, the generated reference points contain measurements for at most two capture points; all the capture points in coverage could be used, which would make the reference points more similar to real measurements and ease the similarity function’s work. It would probably also help in situations such as the east drift, in which CP6 has sometimes troubles receiving the packets, if the three capture points CP1, CP2 and CP3 were all used. Another way to improve the autocalibration would be to allow to define a graph of reference points instead of a path; indeed, currently it is possible to generate reference points all the way from MP51 to MP61 via CP1 and CP3, but for instance the section from CP1 to MP11 cannot be covered at the same time since it would involve a bifurcation in the path of the generated reference points and turn it into a graph.2

All the measurements, as well as further detail about the experiments, are published in a Git repository (see [5]) to allow people from the community to extract their own results, with OwIIPS or other positioning systems and tools.

ACKNOWLEDGEMENTS

The authors would like to thank the laboratory’s technician Ahmad Al-Hogeiri for his help installing the hardware into the mine, as well as Robert Boucher and Denis Gagnon from the CANMET staff for the information, advice and help they provided.

REFERENCES


2 However it is of course possible to generate reference points from MP51 to MP11, or from MP11 to MP61 via CP1 and CP3.
Matteo Cypriani obtained his PhD in 2012 at the Université de Franche-Comté, France, with a thesis on self-calibrated Wi-Fi-based indoor positioning. Prior to his PhD, he completed a computer science engineering curriculum in 2008 at the Université de Technologie de Belfort-Montbéliard, France.

He currently holds a post-doctoral position at the Laboratoire de Recherche Télébé en Communications Souterraines (LRTCS, Underground Communication Laboratory) of the Université du Québec en Abitibi-Témiscamingue (UQAT), Canada; his research activities are focused on complex confined environment problems such as underground propagation, localisation and distributed sensor networks. His research interests lie mainly in radio frequency networks and on positioning techniques and algorithms.

Nadir Hakem is professor at UQAT (Université du Québec en Abitibi-Témiscamingue) since 2009. He obtained his Advanced Studies Diploma and his PhD in Computer Sciences in 1999 and 2004 respectively, both at Blaise Pascal University, Clermont-Ferrand, France. He also holds an Engineer degree from the Science and Technology University of Algiers, Algeria. His PhD thesis was focused on the proposition of appropriate medium access control mechanism based on the IEEE 802.11 standard for wireless home network.

Between 2006 and 2009, he was assistant researcher at the Research Laboratory Télébé in Underground Communications at the Quebec University in Abitibi-Témiscamingue (LRTCS-UQAT). His major field of study was interoperability of heterogeneous communication network in complex confined mine areas. His current research interests are protocol engineering for wireless network, measurements and analytical predictions of RF signal, network design based on millimeter wave signal to multimode communication system based on software defined radio, vehicular network and cognitive radio.

Dr. Hakem is a member of the Order of Engineers of the Province of Québec (OIQ) and member of the Institute of Electrical and Electronics Engineers (IEEE). He is involved in many scientific conference committees. He has supervised the work of over 20 graduate and post-graduate students over the last 5 years. He is involved in many scientific conference and journal committees.

Gilles Delisle is professor at UQAT (Université du Québec en Abitibi-Témiscamingue) since 2010 and president of GDE inc, a consulting organization in telecommunications. He is also Emeritus Professor at Laval University and adjunct professor at four other universities. He was Director, Technology Integration Centre at Technopôle Défense and Security in Valcartier, Québec, Canada from 2008 to 2010. From June 2004 to March 2008, he was Vice-President Research at the International Institute of Telecommunications in Montréal, Canada. Previously, he was Director and Professor at the School of Information Technology and Engineering at the University of Ottawa from 2002 to 2004 and he has been a Professor of Electrical and Computer Engineering at Laval University, Québec, Canada from 1973 to 2001, where he was head of the department from 1977 to 1983. From June 1992 to June 1997, he was also Director of INRS-Telecommunications, a research institute which is a part of the Université du Québec. He is involved in research work in underground mines communication networks, intelligent antenna array, radar cross-section measurements and analytical predictions, mobile radio channel propagation modelling, personal communications and industrial realization of telecommunications equipment.

Dr. Delisle is a member of the Order of Engineers of the Province of Québec and Professional Engineers of Ontario, Fellow of Engineer’s Canada, Past-President of the Canadian Engineering Accreditation Board, Fellow of the Canadian Academy of Engineering, Past Canadian President of URSI. Past President of ACFAS, Life Fellow of the Institute of Electrical and Electronics Engineers (IEEE), Fellow the Canadian Engineering Institute, the Canadian Academy of Engineering and of the Institution of Engineering and Technology (IET-UK).
Learning Algorithm of Neural Networks on Spherical Cap

Zhixiang Chen and Jinjie Hu*
Department of Mathematics, Shaoxing University, Shaoxing, China
*Corresponding author, Email: {czx, jinjiehu}@usx.edu.cn

Feilong Cao
Department of Mathematics, China Jiliang University, Hangzhou, China
Email: feilongcao@gmail.com

Abstract—This paper investigates the learning algorithm of neural network on the spherical cap. Firstly, we construct the inner weights and biases from sample data, such that the network has the interpolation property on the sampling points. Secondly, we construct the BP network and BP learning algorithm. Finally, we analyze the generalization ability for the constructed networks and give the numerical experiments.

Index Terms—Spherical Cap; Neural Network; Learning; Interpolation; BP Algorithm

I. INTRODUCTION

The \((d-1)\)-dimensional unit sphere \(\mathbb{S}^{d-1}\) in \(\mathbb{R}^d\) is defined by
\[
\mathbb{S}^{d-1} = \left\{ (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d : x_1^2 + x_2^2 + \cdots + x_d^2 = 1 \right\}.
\]
In recent years, the construction and approximation of spherical function have attracted the attention of large number of scholars. As the main approximation tools on the unit sphere, spherical polynomials are fundamental, and many results have been explored [1]. The spherical thin-plate splines, as natural analogs of the classical thin-plate splines, have also been constructed for interpolation and approximation on the unit sphere [2]. Moreover, in many applications of geophysics and metrology, we usually need to find some functional models to fit the scattered data collected over the surface of the earth via satellite or ground stations. Recently, a class of so-called spherical positive definite radial basis functions has been used to tackle the problem by interpolating the samples, and lots of results have been obtained. We refer reader to [3] and references therein.

On the other hand, it is well-known that feed-forward neural networks are universal approximator. There has been a lot of research devoted to the topic on the compact subset of Euclidean space \(\mathbb{R}^d\), for example, Cybenko [4], Funahashi [5], Chen and Chen [6], Barron [7], Chen [8], Cao, Xie, and Xu [9], and Chen and Cao [10]. On the unit sphere, Mhaskar et al. [11] introduced the following zonal function network
\[
\sum_{k=1}^{n} a_k \phi(\xi_k \cdot x),
\]
to deal with spherical scattered data approximation, where \(\xi_k, x \in \mathbb{S}^{d-1}, a_k \in \mathbb{R}, \phi\) is a real function defined on \([-1,1]\).

In [12] [13] [14] [15], the following feed-forward networks defined on the unit sphere were considered and some approximation properties were studied:
\[
\sum_{k=1}^{n} a_k \phi(\xi_k \cdot x + c_k),
\]
where \(a_k, c_k \in \mathbb{R}, x \in \mathbb{S}^{d-1}, \xi_k \in \mathbb{R}^d\), and the activation function \(\phi\) was defined on \(\mathbb{R}\).

Since sigmoidal type functions are one important class of activation functions, these functions are usually used to be the activation functions in the hidden layer of neural networks. In fact, a sigmoidal function \(\delta : \mathbb{R} \to \mathbb{R}\) is a bounded function and satisfies
\[
\lim_{x \to \infty} \delta(x) = 1, \lim_{x \to -\infty} \delta(x) = 0.
\]

Recently [16] [17] [18] studied the error of approximation by networks with sigmoidal activation functions. Since we are usually concerned with the target function defined on the local area of \(\mathbb{S}^{d-1}\), on our samples are derived from a cap of \(\mathbb{S}^{d-1}\), so we will discuss the approximation of networks on a spherical cap. It is well known that interpolation is a popular and important approximation method when the samples are obtained. Therefore we construct the interpolation networks to approximate target function. This kind of network can save much training time, naturally, its construction is difficult. On the other hand, although there have been many results concerning the spherical network approximation, the results about learning algorithm are relatively few. In fact, the basic BP (back propagation) learning algorithm has been raising many scholars’ interest, and has been applied to a variety of disciplines, see [19] [20] [21]. Hence we are also to discuss the BP algorithm in the application of network approximation of spherical cap. Considering
\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]
is a typical sigmoidal function, so we will
use $\sigma$ to construct activation functions of interpolation network. And we are also to discuss the BP algorithm of network approximation, and the target functions are defined on a spherical cap.

The paper is organized as follows. Section 2 discusses the existence of spherical interpolation networks, where a constructive method based on scattered (randomly) sampling data will be utilized. In Section 3, we will use classical gradient method and derive the adjusting formulas of weights. And the BP learning algorithm on spherical cap will be established. In Section 4, we will analyse the generalization ability for the constructed networks and give the numerical experiments. Finally, some conclusions are presented in Section 5.

II. CONSTRUCTION OF INTERPOLATION NETWORKS

Let $(X_1, y_1), (X_2, y_2), \cdots, (X_n, y_n)$ be samples, where $X_i \in \mathbb{R}^d, y_i \in \mathbb{R}, i=1,2, \cdots, n$.

$N(x)$ is a network, if it has the following property

$$N(x_i) = y_i, i=1,2, \cdots, n,$$

we say that $N(x)$ is an interpolation network. For given samples above, we will choose properly weights $\xi_i$ and threshold $c_i$ such that the network

$$\sum_{i=1}^{n} a_i \phi(\xi_i \cdot x + c_i)$$

has interpolation property, where weights $a_k$ are derived from linear systems

$$\sum_{i=1}^{n} \frac{a_i}{\xi_i - c_i} = y_i, \quad i=1,2, \cdots, n.$$

Thus, when sample points $X_i \in \mathbb{S}^{d-1}$, the construction of interpolation network is trivial. So we first discuss the general case, that is, the case of $x \in \mathbb{R}^d$.

For two distinct vectors $X_i, X_j$, $X_i = (x_{i1}, x_{i2}, \cdots, x_{id}) \in \mathbb{R}^d, i=1,2$.

If there exists $j_h \in \{1,2, \cdots, d\}$ such that $x_{ih} < x_{jh}$ and $x_{ij} = x_{j2}$ for $j = j_h + 1, j_h + 2, \cdots, d$, we say $X_i \prec X_j$.

To prove the main result of this section, we first give a lemma.

**Lemma 1.** For $n$ distinct vectors: $X_1 \prec X_2 \prec \cdots \prec X_n \in \mathbb{R}^d$, there exists a vector $W \in \mathbb{R}^d$, such that $W \cdot X_1 < W \cdot X_2 < \cdots < W \cdot X_n$.

**Proof.** For $n$ distinct vectors: $X_i = (x_{i1}, x_{i2}, \cdots, x_{id}) \quad (i=1,2, \cdots, n)$, we set

$$w_j = \frac{1}{1 + \max_i \|x_i\|}, \quad j=1,2, \cdots, d.$$

Then

$$x_i' = w_j x_{ij}, \quad i=1,2, \cdots, n, j=1,2, \cdots, d.$$

Let

$$y_i' = x_i' - x_j', \quad i=1,2, \cdots, n-1, j=1,2, \cdots, d.$$

For given $j(1 \leq j \leq d)$, if $y_i' = 0$, for $i=1,2, \cdots, n-1$, then let $n_j = 2$; else, then let $n_j = \min_i \{y_i' \}$. Then

$$w_j' = w_j \prod_{i=1}^{j-1} n_j, \quad j=1,2, \cdots, d,$$

and

$$W = (w_1', w_2', \cdots, w_d').$$

For fixed $i(1 \leq i \leq n-1)$, from $X_i \prec X_i+1$, it follows that there exists $k_0 \in \mathbb{N}$ such that $x_{ki} < x_{i+1,k}$. Now we set $x_i = x_{i+1,k}, j = k_0 + 1, \cdots, d$.

So we have

$$W \cdot X_i = W \cdot X_{i+1} - W \cdot X_i = \sum_{k=1}^{k_0} \prod_{j=1}^{k-1} n_j + (x_{i+1,k} - x_{ik}) \prod_{j=1}^{k_0} n_j = I_i + I_2.$$

Since,

$$2 \prod_{i=1}^{k_0} n_i \left(1 + \frac{1}{n_{k_0-1}} + \cdots + \frac{1}{n_{1\cdots n_{k_0-1}}} \right) \leq 2 \prod_{i=1}^{k_0} n_i \cdot \frac{1}{1 - \frac{1}{2}},$$

and

$$I_1 = \prod_{i=1}^{k_0} n_i \cdot y_i (x_{i+1,k} - x_{ik}) = \prod_{i=1}^{k_0} n_i \cdot y_i' = \prod_{j=1}^{k_0} n_j \cdot y_j'.$$

Hence, $W \cdot X_i \prec X_{i+1}, \quad i.e., W \cdot X_i < W \cdot X_{i+1}$.

The proof of Lemma 1 is complete.

For $\sigma(x) = \frac{1}{1 + e^{-x}}$, we set

$$\tilde{\sigma}(x) \approx \frac{1}{2} (\sigma(x+1) - \sigma(x-1)),$$

and will prove the following theorem.

**Theorem 1.** For $n$ distinct vectors: $X_1 \prec X_2 \prec \cdots \prec X_n$, there exists vectors $W_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$ $(i=1,2, \cdots, n)$, such that

$$\tilde{\sigma}(W_i X_i + b_i) = \tilde{\sigma}(W_i X_{i+1} + b_i), \quad i=1,2, \cdots, n-1,$$

and $\tilde{\sigma}$ is an even and strictly decreasing on $[0, +\infty)$.
Clearly,
\[ \tilde{\sigma}(x) = \frac{1}{2} (e^{-x^2}) \frac{e^x}{(1 + e^{-x^2})(1 + e^x)}, \tilde{\sigma}(0) = \frac{1}{2} e^{-1}. \]

Now we want to prove that there holds
\[ \tilde{\sigma}(x) = \frac{1}{2} (e^{-x^2}) \frac{e^x}{(1 + e^{-x^2})(1 + e^x)} < \frac{1}{2} e^{-1}, \]
that is,
\[ \frac{e^x}{(1 + e^{-x^2})(1 + e^x)} < \frac{e}{n(e+1)^2}, \quad (1) \]

Since
\[ \frac{e^x}{(1 + e^{-x^2})(1 + e^x)} < \frac{e}{n(e+1)^2}, \]
when
\[ e^x > \frac{n(e+1)^2}{e}, \quad (2) \]
the inequality (1) is true. Obviously, when \( x \) satisfies
\[ e^x > \frac{4ne^2}{e} = 4ne, \]
the inequality (2) is valid. So we can choose \( x > \ln 4n^2 + 1 \), and thus when \( |x| > \ln 4n^2 + 1 \), there holds \( \tilde{\sigma}(x) < \frac{\tilde{\sigma}(0)}{n} \).

By Lemma 1, there exists \( W \) such that
\[ W \cdot X_0 < W \cdot X_1 < \cdots < W \cdot X_N, \]
Therefore, we fix \( a > \ln 4n^2 + 1 \), set
\[ k_i = \min\{W \cdot X_0, W \cdot X_1, \ldots, W \cdot X_{n-1}\} \]
\((i = 2, 3, \ldots, n-1)\),
and choose \( W_i = k_i, W(i = 2, 3, \ldots, n) = W_N = W_0, \)
\( W_{n1} = W_{n-1} \), and \( b_i = -W_i \cdot X_0 (i = 1, 2, \ldots, n) \),
then \( \tilde{\sigma}(W_i \cdot X_j) = \tilde{\sigma}(0) \).

On the other hand, for \( i, j = 1, 2, \ldots, n, j \neq i \),
we have
\[ (W_i \cdot X_j + b_i) - (W_i \cdot X_j + b_j) \geq k_i \ | W_i \cdot X_j - W_i \cdot X_j | \geq 2a, \]
and
\[ W_i \cdot X_0 + b_i < W_i \cdot X_1 + b_i < \cdots < W_i \cdot X_{n-1} + b_i \]
\[ < -a < W_i \cdot X_1 + b_i = 0 < a \]
\[ < W_i \cdot X_{n1} + b_i < \cdots < W_i \cdot X_n + b_i. \]

So when \( i \neq j \), it follows that
\[ \tilde{\sigma}(W_i \cdot X_j + b_j) \leq \frac{\tilde{\sigma}(0)}{n}. \]

which leads to
\[ \tilde{\sigma}(0) = \tilde{\sigma}_a > \sum_{j=1}^{n} |\tilde{\sigma}_j|, \]
and illustrates that matrix \( G_a \) is strictly and diagonally dominant. Therefore, \( G_a \) is nonsingular. This finishes the proof of Theorem 1.

From Theorem 1 we know that for samples \((X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)\)
there exists feed-forward neural networks
\[ N_a(X) = \sum_{i=1}^{n} c_i \tilde{\sigma}(W \cdot X + b) , \]
such that \( N_a(X) = y_i (i = 1, 2, \ldots, n) \), which shows that the network \( N_a(x) \) can be an interpolation function for samples \((X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)\).

Since \( S^{d-1} \) is a subset of \( \mathbb{R}^d \), we obtain immediately that

**Corollary 1.** For \( n \) samples
\((X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)\),
there exists a feed-forward neural networks
\[ N_a(X) = \sum_{i=1}^{n} c_i \tilde{\sigma}(W \cdot X + b) , \]
such that \( N_a(X) = y_i (i = 1, 2, \ldots, n) \).

Although function interpolation algorithm is an effective approach for fitting scattered data, the construction of weight is difficult. As an important performance of feed-forward neural networks, the BP algorithm can iteratively adjust the network weights to minimize the least squares objective function by training samples, and thus the networks have a certain generalization ability.

### III. **Spherical BP Neural Network**

The multilayer perceptron employing BP algorithm is one of the most extensive and applicable neural networks. The basic idea of BP algorithm is composed of two processes: the feed-forward propagation of signals and the back-propagation of errors (see [22] for details). In this paper we will use the following (see Figure 1) spherical feed-forward neural network with single-hidden layer, it can be mathematically modeled as

![Figure 1. Neural network with single-hidden layer](image_url)

\[ N_a(x_1, x_2, x_3) = \sum_{i=1}^{n} w_i \tilde{\sigma} \left( \sum_{j=0}^{3} w_j x_j \right), \]
where \( x_0 = -1, (x_1, x_2, x_3) \in S^2 \).

Now we give the definition of network error and the ideas of adjusting weights. When the output \( o \) isn't equal to the sample value \( y \), there exists error \( E \), it is defined as
\[ E = \frac{1}{2}(y - o)^2 = \frac{1}{2}\left[ y - \sum_{i=0}^{3} w_i \sigma \left( \sum_{j=0}^{3} w_j x_j \right) \right]^2. \]

**TABLE I.**

<table>
<thead>
<tr>
<th>Algorithm 1: BP</th>
</tr>
</thead>
<tbody>
<tr>
<td>step 1: For ( i = 1, 2, \cdots, n ), ( j = 0, 1, 2, 3 ), initialize ( w_i, w_j ) with random values, initialize counters ( p, q ) of sample model and training times, both for 1, error ( E = 0 ), and the learning rate ( \eta ) and training accuracy ( E_{\text{err}} ) are set decimals in (0,1) respectively.</td>
</tr>
<tr>
<td>step 2: Using sample data ( X_i ), initialize ( x ) for ( X_i ), and calculate the output ( o ).</td>
</tr>
<tr>
<td>step 3: Using sample data ( y ), calculate error ( E ).</td>
</tr>
<tr>
<td>step 4: Adjust weights ( w_i, w_j ), ( i = 1, 2, \cdots, n ); ( j = 0, 1, 2, 3 ).</td>
</tr>
<tr>
<td>step 5: If ( p &lt; n ), counters ( p, q ) add 1, and return to step 2, or go to step 6.</td>
</tr>
<tr>
<td>step 6: If ( E &lt; E_{\text{err}} ), the training goes over, or, set ( E = 0 ), ( p = 1 ), and go to step 2.</td>
</tr>
<tr>
<td>step 7: Output ( w_i, w_j ), ( i = 1, 2, \cdots, n ); ( j = 0, 1, 2, 3 ).</td>
</tr>
<tr>
<td>Thus we get a network.</td>
</tr>
</tbody>
</table>

Since the principle of adjusting weights is that we should have the error become smaller and smaller, so the adjusting value of weight should be direct proportion to the gradient descent of error, that is

\[ \Delta w_i = -\eta \frac{\partial E}{\partial w_i}, i = 1, 2, \cdots, n. \]

\[ \Delta w_j = -\eta \frac{\partial E}{\partial w_j}, i = 1, 2, \cdots, n; j = 0, 1, 2, 3, \]

where \( \eta \in (0, 1) \) is the learning rate. By standard calculation, we have

\[ \frac{\partial E}{\partial w_i} = (y - o)\sigma \left( \sum_{j=0}^{3} w_j x_j \right), i = 1, 2, \cdots, n. \]

\[ \frac{\partial E}{\partial w_j} = (y - o)\sigma \left( \sum_{j=0}^{3} w_j x_j \right) x_j, i = 1, 2, \cdots, n; j = 0, 1, 2, 3. \]

When \( \sigma(x) = \frac{1}{1 + e^{-x}} \) we have

\[ \sigma'(x) = \sigma(x)(1 - \sigma(x)), \]

and

\[ \sigma'(x) = \frac{1}{2}(\sigma'(x+1) - \sigma'(x-1)) = \frac{1}{2} \left\{ \sigma(x+1)(1 - \sigma(x+1)) - \sigma(x-1)(1 - \sigma(x-1)) \right\} = \frac{1}{2} \left\{ \sigma(x)(1 - \sigma(x+1)) - \sigma(x)(1 - \sigma(x-1)) \right\}. \]

Thus, we derive the following computation formulas:

\[ \Delta w_i = (y - o)\eta \sigma \left( \sum_{j=0}^{3} w_j x_j \right), i = 1, 2, \cdots, n. \]

Now we can give the BP algorithm as follows (see TABLE I).

**IV. NUMERICAL EXPERIMENTS AND ANALYSES**

Now, we apply numerical experiments to discuss the generalization capacity of above two kinds of networks on spherical cap. We choose two functions defined on cap

\[ \{ x = (x_1, x_2, x_3) : d(x, (0,0,1)) \leq \frac{\pi}{4} \} \]

as follows:

\[ f_1(x_1, x_2, x_3) = x_1 x_2 x_3, f_2(x_1, x_2, x_3) = x_1^2 + x_2 x_3 \]

where \( x_1^2 + x_2^2 + x_3^2 = 1 \).

To describe the distribution of samples, we use polar coordinates representations:

\( x_i = \sin(\theta)\sin(\phi), x_2 = \sin(\theta)\cos(\phi), x_3 = \cos(\theta) \),

where \( \theta \in [0, \pi / 4], \phi \in [0, 2\pi] \).

We randomly choose \( N \) samples \((X_1, y_1), (X_2, y_2), \cdots, (X_N, y_N)\) to learn interpolation and BP network respectively. And we test generalization capacity by stochastic points \((TX_1, TY_1), (TX_2, TY_2), \cdots, (TX_N, TY_N)\). The distribution of sample points in experiments are described in Figure 2 and Figure 3.
We define sample test error $E_1$ and generalization test error $E_2$ as follows:

$$E_i = \frac{\sum_{i=1}^{N} |Y_i - y_i|}{N}, E_2 = \frac{\sum_{i=1}^{M} |TY_i - ty_i|}{M},$$

where $Y_i, TY_i$ denote the output values of samples and test points respectively.

From experimental results we find that the sample size affects performances of interpolation network. This network performs well when the sample size is between 10 and 30. So we can call such sample size the valid size. Furthermore, we illustrate the relation between errors and sizes with Figure 6 and Figure 7.

From all above numerical results, Figure 4 - Figure 9, we can see that:

1. When the sample size $N$ is relatively small ($N \leq 10$), these two kinds of networks perform badly in generalization. It is easily understood, for less sample information naturally leads to worse learning effect.
2. As for interpolation network, when the sample size $N$ is relatively large ($N \geq 30$), the network lies in an unstable state, hence, the generalization capacity is undesirable. The reason for such result is that the interpolation $G_\sigma$ is close to singular or badly scaled.  

3. For BP network, the error becomes smaller, and the generalization capacity becomes better as the sample size $N$ increases. However, in order to achieve the same performance, BP network will spend longer time.  

4. When sample size $N$ is at valid size, especially when $18 \leq N \leq 24$) the interpolation network performs very well, and the generalization behaves with very high accuracy. By numerical experiments above we obtain that when the number of sample points on a cap is appropriate, the performance (error and time cost) of the interpolation network excels the BP network. While, if the number of samples gets relatively large, the interpolation network becomes instable easily. In this case the BP network can works, but the time which the network spends increases rapidly.  

To compare the learning effects of above two networks further, we draw 3-dimensional curved surfaces of $f_1$ and $f_2$ so that we can see more clearly the coincidence of tested points. Now, we use above polar coordinates representations to transform $f_1, f_2$ into $g_1(\theta, \phi)$ and $g_2(\theta, \phi)$ . Choosing 20 sample points we get experimental results, see Figure 10 - Figure 13.

![Figure 10. Simulating with interpolation network for $f_1(g_1(\theta, \phi))$](image)

![Figure 11. Simulating with BP network for $f_1(g_1(\theta, \phi))$](image)

![Figure 12. Simulating with interpolation network for $f_2(g_2(\theta, \phi))$](image)

![Figure 13. Simulating with BP network for $f_2(g_2(\theta, \phi))$](image)

It is not difficult to see that, for these two networks, almost all the points marked with “+” are in five-pointed stars, which indicate that the testing effects for original sample points are ideal. As for generalization testing, all the points marked with “*” exactly lie in the center of “o”. This indicates that the generalization of interpolation network is very good.  

In the case of BP network, most points marked with “*” are in the center of “o”, only a few points have fairly big biases.

V. CONCLUSIONS

Interpolation is an important method of data fitting and numerical approximation, so we construct the interpolating network, and the numerical experiments that it can reach the results we want. However, we require constructing weights and solving large scale linear systems to obtain the interpolant of a target function, and when the number of interpolation points is very large, the interpolation matrix may become ill-conditioned. Hence, we also study the BP learning algorithm of neural networks. By properly choosing learning rate and the number of samples the network may have ideal learning effects and generalization capacity.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Nos. 61179041, 61272023).
REFERENCES


Zhixiang Chen, male, was born in Jiangsu Province, China, on March, 1966. He received the M.S. degree in Mathematics from Hangzhou University, China in 1995. In 2002, he received the Ph.D. degree in Mathematics from Zhejiang University, China. He now is a Professor in Shaoxing University. His current research interests include neural networks and approximation theory.

Jinjie Hu received the B.S. degree in Computer Software from Lanzhou University, China in 1999. He received the M.S. degree in Computational Mathematics from Shanghai University, China in 2011. He now is a lecturer in Shaoxing University, China. His current research interests include neural networks and scientific computing.

Feilong Cao received the M.S. degree in Applied Mathematics from Ningxia University, China in 1998. In 2003, he received the Ph.D. degree in institute for information and System Science, Xi’an Jiaotong University, China. He now is a Professor in China Jiliang University. His current research interests include neural networks and approximation theory.
Research on channel selection algorithms in cognitive radio networks

Lin LI\textsuperscript{a,b}, Yi-na DENG\textsuperscript{b}, Yang YUAN\textsuperscript{b}, Wen-jiang FENG\textsuperscript{b}

\textsuperscript{a} Chongqing College of Electronic Engineering, Chongqing, China
Email: mzlsml@126.com
\textsuperscript{b}College of Communication Engineering, Chongqing University, Chongqing, China
Email: yn.deng@foxmail.com

Abstract—To address the secondary users channel selection issue in cognitive radio network, a novel channel selection strategy is proposed. Four typical channel selection models under auction mechanism, machine learning scheme, channel prediction scheme and optimization scheme are compared and analyzed. Based on the optimization theory, the selfish channel selection algorithm and the cooperative channel selection algorithm are proposed in view of the heterogeneity of the channel. The selfish algorithm selects the channel which provides the maximum transmission rate for the secondary users (SU), while the cooperative algorithm selects the channel that benefits overall system throughput. Simulations compare proposed algorithms with random channel selection algorithm, and suggest proposed algorithms outperform random channel selection algorithm in terms of system average throughput, channel utilization, average handoff time and average transmission time.

Index Terms—cognitive radio, channel selection, channel heterogeneity

I. INTRODUCTION

To explore the idle time domain, frequency domain and spatial domain resource of authorized network, through opportunistic dynamic access, cognitive radio can realize spectrum sharing between primary users (PU) and secondary users (SU) and increase resources utilization consequently. Based on the variation of available channels, services types, transmission mode and geographic position inside secondary users working area, and on the limitation of transmission path and related policies, dynamic access will distribute and utilize resources by the near-real-time mode. Cognitive radio dynamic spectrum access consists of spectrum detection and spectrum development. The spectrum detection, composed of spectrum sensing and spectrum analysis, is responsible for the search and attributive analysis of spectrum resource, and the spectrum development, composed of spectrum decision and spectrum handoff, is responsible for the access judgment and access perform of spectrum resource\textsuperscript{[1]}. The selfish algorithm selects the channel which provides the maximum transmission rate for the secondary users (SU), while the cooperative algorithm selects the channel that benefits overall system throughput. Simulations compare proposed algorithms with random channel selection algorithm, and suggest proposed algorithms outperform random channel selection algorithm in terms of system average throughput, channel utilization, average handoff time and average transmission time.

Typical channel selection schemes consists of those based on auction model\textsuperscript{[2,3]}, machine learning\textsuperscript{[4,5]}, optimization\textsuperscript{[6–9]} and channel prediction\textsuperscript{[10]}. Channel selection scheme based on auction model, which takes PU as sellers, SU as buyers and idle channel as products, models the dynamic spectrum sharing of SU according to the game theory, and analyzes SU’s behavior and function in the formalized game structure. The basic idea of the channel selection scheme based on learning is that SU, combined short-term spectral historical state information with long-time historical statistical law, analyzes the success probability of each different channel transmit to estimate the optimal channel and make the channel selection consequently. The channel selection scheme based on optimization takes the most valuable performance index, including system time, throughput, minimum handoff times, maximum channel utilization and minimum system delay, as objective function, then solves the function with optimization method, in this way, the channel selection problem resolves into the constrained optimization problem. While the channel selection scheme based on prediction model, which relies on specific channel model, analyzes PU’s activity routines through environment sensing, and predicts relevant information of available channel through information of present spectrum sensing, such as idle time and spectrum stability, thereby guide secondary users to select channel.

Schemes mentioned above have their respective advantages in different specified conditions, so does their disadvantages. The channel selection scheme based on auction model, which can guarantee the fairness of spectrum practice, applies to the situation in which resource price is uncertain and the price changes according to buyers’ needs, and confirms to the heterogeneous network coexistence which can be meet by different service requirements. However, users’ behavior is not really static and cooperative in practical environment, instead, there are also non-cooperative, selfish and malicious users. Furthermore, delay of auction process is inevitable, therefore this kind of scheme can not satisfy systems that have higher requirements for delay. The channel selection scheme based on learning applies to systems whose PU’s activity is regular, nevertheless in the cognitive radio network coexistence scene, heterogeneous SU’s activity routines must be concerned, thus the machine learning algorithm will face a new challenge. The channel selection scheme based on optimization is simple, direct and targeted, but it is too complex in the process of multi-objective optimiza-

\textsuperscript{[1]} Project supported by Science and Technology Project of Chongqing Education Commission of China(KJ102201), Natural Science Foundation of Hainan Province of China(614237)
tion. The channel selection scheme based on prediction mode is able to pre-select and switch target channel before PU appears through historical information statistics or some prediction technique. It enhances the efficiency of data transmission, and reduces the delay brought by real-time sensing, however, according to the information delay caused by periodic detection, the effectiveness of target channel list can not be guaranteed. In this way, this kind of scheme only applies to systems whose PU’s activity is regular and services which are sensitive to time delay.

SU doesn’t have the unique right, even the preferential right, to use the authorized channel, which will lead to a situation that multiple SU select the same channel simultaneously. If the channel cannot be selected reasonably, the transmission performance will be impacted, and the spectrum efficiency will reduce as well. For SU, different channels have different transmission performances, while the SU tends to select channels which have better transmission performance to deliver data. Therefore the process of channel selection is a trade-off of load-balancing, transmission efficiency and fairness. Combined with the heterogeneity of different SU caused by available channel, this paper proposes the selfish channel selection algorithm and the cooperative channel selection algorithm in the view of channel heterogeneity. The selfish algorithm selects the channel which provides the maximum transmission rate for the SU, while the cooperative algorithm selects the channel that benefits overall system throughput. Simulations compare proposed algorithms with random channel selection algorithm, and suggest proposed algorithms outperform random channel selection algorithm in terms of system average throughput, channel utilization, average handoff time and average transmission time.

II. SYSTEM MODEL

A. System Model

As the coexistence scene of cognitive radio network and primary user network shown in the Fig.1, SU senses the existence of PU periodically in data transmission period. Once PU returns, SU must switch to other available channel or to communication outage. Fusion Center (FC) of cognitive radio network receives channel detection information from SU, and transmits channel allocation results by channel control.

B. Assumption condition

1) ignore the Spectrum sensing error’s impact on channel selection.
2) once SU occupy one channel, no outage in the unit transmission time.
3) the spectrum sharing between SU and PU is based on Overlay.

III. CHANNEL SELECTION ALGORITHM

SU transmits channel sensing information periodically to FC through control channel, then FC provides a channel which has the maximum transmission rate to SU based on the present channel’s availability (the selfish channel selection algorithm), or provides a channel which can enhance its overall system throughput (the cooperative channel selection algorithm).

A. Selfish Channel Selection Algorithm

When FC selects the channel with the maximum transmission rate through channel information, S represents available channel set, and is the average transmission rate when j (SU) occupies i (channel). According to the heterogeneity of different SU caused by different channels, when \( j_1 \neq j_2 \) , \( TR_{i,j_1} \neq TR_{i,j_2} \). For that j can select the channel with the maximum transmission rate, weighting function \( W_{ego}(i,j) \) is defined to represent the transmission rate when j occupies i.

\[
W_{ego}(i,j) = \frac{TR_{i,j} \times G(N_i + 1)}{N_i + 1}
\]  

In the formula, \( G(X) \in [0, 1] \) shows the actual throughput ratio when x users occupy the channel\(^{[11]}\), and decreases as x increases, because more users make the channel competition more intense. \( N_i \) shows the number of SU in the channel i before j enters into i, which means when j enters into i, number of SU turns into \( N_i + 1 \). FC distributes the channel k with the greatest weight to the SU\(_j\):

\[
k = \text{Max}[W_{ego}(i,j)]
\]  

This algorithm distributes the channel with the maximum transmission rate to new users, but it ignores the interference to other SU caused by the scheme. Furthermore, the transmission performance of new SU’s will be influenced by the next user who requests joining.

B. Cooperative Channel Selection Algorithm

Aim of the algorithm is to select channels with higher system throughput. The increase of system throughput can be calculated by the increase throughput of channel i after j enters into i. \( W_{cop}(i,j) \) shows the throughput increase after j enters into i:
As shown in Fig.2, the average throughput changes while the number of SU altered. When the numbers of SU over thirty, the cooperative channel selection algorithm (C-selection) is chosen, and each user’s throughput is larger than that in the selfish channel selection algorithm (S-selection). The reason is that when the cooperative channel selection algorithm is chosen, FC’s each selection is on the premise of increasing system throughput, which can reduce overhead caused by unbalanced load allocation. Although the performance of selfish channel selection algorithm is inferior to cooperative channel selection algorithm, it is still superior to other channel selection schemes, due to that FC selects channels with the maximum transmission rate for each SU to get higher spectrum efficiency. When the number of SU is less than twenty eight, there is no obvious difference among the three selection schemes, because the present channel capacity is unsaturated. According to the simulation environment parameters, if 64QAM, 1/2 encoding rate modulation mode is selected, each channel’s throughput is about 14Mbps, and the overall throughput of ten channels is 140Mbps. If PU occupies half of the channels, all SU share the rest rate of 70Mbps, and if the actual throughput ratio is 0.79, the sharing transmission rate is 55.3Mbps.

As shown in Fig.3, spectrum efficiency can be calculated by ratio of cumulative time in the channel and the overall idle channel time. SU doesn’t have information to send all the time, and switch of SU will cause delay, therefore the idle channel utilization cannot reach 100%. It is can be seen that cooperative channel selection algorithm has the best channel utilization, which benefits from the maximized system throughput. When the number of SU reaches fifty, the channel utilization will reach 90% through cooperative channel selection algorithm, which means the maximum throughput will reach 49.77Mbps. Compared with the throughput in Fig.4, each SU’s average throughput is 0.75Mbps, the overall throughput is 37.5Mbps. It is about 75% of the maximum throughput, because that multiple SU’s access into one channel reduce the actual throughput ratio, and of the spectrum handoff overhead.

Both the selfish channel selection algorithm and the cooperative channel selection algorithm are superior to random channel selection algorithm (R-selection), because both of two algorithms consider the actual condition of channel, and get a better spectrum efficiency.

Fig.4 shows the situation that SU’s average handoff time changes with PU’s average sleep time. In the simulation, the number of SU is assumed to be twenty. The result indicates that with the reduction of PU’s sleep time, SU’s average handoff time increases. The reason is that once PU returns to the present channel, FC will switch SU into other channel compulsively. Each file’s transmission time is far more larger than 100ms, thus when PU’s average sleep time is 100ms, all the channel selection algorithms need relatively long time on switching, and the SU will switch frequently. Even so, the two algorithms
proposed by this paper take less time on switching than random channel selection algorithm. The former considers the increase of spectrum efficiency, and the relatively large throughput reduces data transmission time, thus the handoff probability caused by PU’s return reduces.

Fig. 5 shows the average transmission time of each single file, in this sense the two algorithms proposed by this paper are still superior to random channel selection algorithm. When PU’s sleep time is 200ms and 300ms, SU’s switch period is less than that when PU’s sleep time is 100ms, therefore the average transmission time of each single file reduces. However when PU’s sleep time continues to increase, although SU’s switch period will reduces, it will lose the chance to switch to a more proper channel, to some extent, the average transmission time of each single file increases.

V. CONCLUSIONS

To explore the channel selection schemes of cognitive radio network, this paper firstly analyzes the basic ideas, algorithm description and performance characteristic, and the range of application of channel selection schemes based on auction model, machine learning, optimization and prediction model, then provides two kinds of channel selection algorithms based on channel heterogeneity Selfish Channel Selection Algorithm and Cooperative Channel Selection Algorithm, at the same time, introduces fusion center into system model to execute information fusion, selection and distribution, thereby distributes channel for secondary users by weight decision. Aim of Selfish Channel Selection Algorithm is to get the maximum transmission rate for each single SU, while that of Cooperative Channel Selection Algorithm is to get the greatest increase of system throughput. The Simulation results and analysis show that the proposed algorithms based on channel heterogeneity in this article outperform random channel selection algorithm in terms of system average throughput, channel utilization, average handoff time and average transmission time.

REFERENCES


Gigabit Wireless Networking with IEEE 802.11ac: Technical Overview and Challenges

Farhan Siddiqui
School of Information Systems and Technology, Walden University, USA
Email: farhan.siddiqui@waldenu.edu

Sherali Zeadally 1 and Khaled Salah 2
1. College of Communication and Information, University of Kentucky, USA
2. Electrical and Computer Engineering, Khalifa University of Science and Technology, UAE
Email: szeadally@uky.edu; khaled.salah@kustar.ac.ae

Abstract—The ever-growing proliferation of wireless devices and concurrent deployments of bandwidth intensive applications has been having a significant impact on user experience in high-density wireless areas. IEEE 802.11ac is a recently ratified Wireless Local Area Network (WLAN) standard that promises to improve wireless user experience by delivering gigabit speed to end-user applications. 802.11ac utilizes new technologies such as Channel Bonding, Beamforming, and Multi-User Multiple-Input Multiple-Output (MU-MIMO) to improve wireless performance. This article reviews recent technological advances made in the field of WLANs and then focuses on the recent IEEE 802.11ac standard. We present actual data rates attained by currently available 802.11ac hardware, and we also discuss foreseen technical challenges that still need to be addressed to enable efficient and seamless gigabit wireless networking.

Index Terms—Wireless; MIMO; WLAN; Standards; Beamforming

I. INTRODUCTION

The history of Wireless Local Area Network (WLAN) technology dates back to the year 1997 [9], when the first 802.11 standard was ratified by the Institute of Electrical and Electronics Engineers (IEEE). Since then, WLAN technologies have been continuously evolving, resulting in improvement of transmission rates, coverage areas, security, Quality of Service (QoS), and mobility. At the same time, WLAN usage and applications have underwent remarkable changes. Recent years have also witnessed an exponential growth in the use of mobile devices and a continuous demand from consumers for faster and more robust wireless connectivity. Figure 1 illustrates the growth in the number of wireless devices, increase in wireless data usage and the rising number of households adopting wireless-only services over the last few years [15]. The recently ratified 802.11ac standard promises to deliver gigabit rates to wireless clients. IEEE 802.11ac, also known as Gigabit Wi-Fi and 5G Wi-Fi, is built upon 802.11n and takes advantage of technological advances in communication systems, coding techniques, signal processing, and processing power. 802.11ac offers significant enhancements in data transmission rates, reliability, and Quality of Service (QoS). In this article we review the 802.11ac standard with respect to its new technological features, consumer demands, benefits, and technical performance challenges.

The rest of this paper is organized as follows. In Section 2, we present the evolution of WLAN focusing on the various wireless standards that have been developed to enhance wireless speeds. Section 3 presents some of the main drivers behind the demand for faster and more reliable wireless connections. In Section 4, we describe the modifications made to the Physical (PHY) and Medium Access Control (MAC) layers of the 802.11 protocol stack in order to provide support for gigabit transmission rates. In Section 5, we discuss technical challenges that still need to be addressed to deliver gigabit wireless throughput to end-users. Finally, in Section 6, we present our concluding comments.

II. EVOLUTION OF WLAN STANDARDS

The original IEEE 802.11 standard defines two underlying layers for WLANs: the Physical (PHY) and the Medium Access Control (MAC). The PHY layer is responsible for modulating and transmitting data. The MAC layer is in charge of controlling transmissions among WLAN clients within a single coverage area known as a Basic Service Set (BSS). The PHY layer of the original 802.11 version standardized three transmission techniques: Infrared (IR), Frequency Hopping Spread Spectrum (FHSS), and Direct Sequence
Spread Spectrum (DSSS). Frequency hopping is the process of transmitting on a given frequency for a short interval and switching to another frequency according to a pre-defined frequency-hopping pattern known to both, transmitter and receiver. DSSS systems spread transmissions across a relatively wide band by artificially increasing the used bandwidth. A DSSS transmitter converts an incoming data stream into a symbol stream where each symbol represents a group of 1, 2, or more bits. DSSS transmitter modulates or multiplies each symbol with a pseudorandom sequence, which is called a “chip” sequence. The multiplication operation in a DSSS transmitter artificially increases the used bandwidth based on the length of the chip sequence. This modulation technique is called Quadrature Phase Shift Keying (QPSK). IEEE defines the use of an 11-chip Barker sequence that is actually a sequence of 11 “-1” and “+1” values. The spread signal can undergo as many as 11 phase changes per symbol period where a non-spreaded QPSK signal would undergo a maximum of one phase change per symbol period. The receiver correlates the received signal with the 11-chip sequence to obtain the originally sent data [3]. The Barker modulation technique provides data rates of 1 or 2 Mbps [5].

Table 1 presents a summary of various WLAN standards developed to date, and their main characteristics. The summary focuses on the main technological changes that were introduced in each WLAN standard to achieve higher wireless rates:

<table>
<thead>
<tr>
<th>Wireless Generation</th>
<th>1G</th>
<th>2G</th>
<th>3G</th>
<th>4G</th>
<th>5G</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE WLAN Standard</td>
<td>802.11</td>
<td>802.11b</td>
<td>802.11a</td>
<td>802.11g</td>
<td>802.11ac</td>
</tr>
<tr>
<td>Date Ratified</td>
<td>1997</td>
<td>1999</td>
<td>1999</td>
<td>2003</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2009</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2012</td>
</tr>
</tbody>
</table>

- Max. Theoretical Data Rate
  - 2 Mbps
  - 11 Mbps
  - 54 Mbps
  - 54 Mbps
  - 600 Mbps
  - 1.3 Gbps
  - 6.7 Gbps
  - 7 Gbps

- Typically Achieved Data Rate
  - 1 Mbps
  - 6.5 Mbps
  - 25 Mbps
  - 25 Mbps
  - 200 Mbps
  - 400 – 700 Mbps
  - Yet to be tested
  - Yet to be tested

- Frequency band
  - 2.4 GHz
  - 2.4 GHz
  - 5 GHz
  - 2.4 GHz
  - 2.4 / 5 GHz
  - 5 GHz
  - 5 GHz
  - 60 GHz

- Max. Spatial Streams
  - 1
  - 1
  - 1
  - 1
  - 4
  - 3
  - 8
  - -

- Backward Compatibility
  - -
  - -
  - 802.11b
  - 802.11a/g
  - 802.11n
  - 802.11n
  - -

- Coverage Area
  - -
  - 30 m
  - 30 m
  - 30 m
  - 50 m
  - 70 m
  - 80 m
  - Short range

- Channel Bandwidth (MHz)
  - 20
  - 20
  - 20
  - 20
  - 20
  - 20
  - 40
  - 80
  - 160
  - 2160

- Modulation Scheme
  - FHSS, DSSS
  - HR-DSSS
  - OFDM
  - DSSS,OFDM
  - OFDM, OFDM (256-QAM)
  - OFDM/256-QAM
  - SC OFDM

- Radio Architecture
  - SISO
  - SISO
  - SISO
  - SISO
  - MIMO
  - MIMO
  - MIMO
  - -

- Typical Usage
  - Basic Wireless connectivity
  - Wireless connectivity in homes/offices
  - Web browsing, email, data transfer
  - Video transmission, gaming
  - Real-time video/audio, high density WLAN zones
  - Real-time video/audio, high density WLAN zones, connectivity to multiple devices
  - High Definition Video, connectivity to multiple devices

- Advanced Antenna Technologies
  - -
  - -
  - -
  - MIMO, up to 4 spatial streams
  - SU-MIMO
  - MU-MIMO, Transmit Beamforming
  - Beamforming

Table 1: IEEE 802.11 Standards
improves efficiency and reduces interference between signals by splitting the radio signal into several sub-signals before they reach a receiver. This method uses multiple sub-carriers to transport information between users. In the OFDM method, a high-speed signal is split into multiple lower-speed sub-signals that are transmitted in parallel at varying frequencies. The parallel transmission over multiple sub-carriers allows OFDM WLANs to achieve a higher data rate (about 54 Mbps). The OFDM technique has a lower multi-path distortion as well as lower multi-path delays.

802.11g: IEEE 802.11g [6] employs DSSS, OFDM, or both at the 2.4 GHz frequency band to provide high data rates of up to 54 Mbps. The combined use of both DSSS and OFDM is made possible through the establishment of four different physical layers known as Extended Rate Physicals (ERP). ERPs coexist during a frame exchange, so the sender and the receiver have the option to select and use one of the four layers. The first layer uses DSSS technology with CCK modulation and provides data rates similar to 802.11b. The second layer (new to 802.11g) uses OFDM at the 2.4 GHz band. The third layer uses DSSS with a coding algorithm called Packet Binary Convolutional Coding (PBCC) and provides 22 and 33 Mbps. The fourth layer uses a combination of DSSS (for transmitting packet header) and OFDM (for transmitting packet payload).

802.11n: IEEE 802.11n [8] brought in significant improvements to the application throughput by introducing new PHY and MAC layer features. IEEE 802.11n provided an approximate transmission rate of 130 Mbps. 802.11n adopted the OFDM modulation with 52 data sub-carriers in 20-MHz channel (instead of 48 sub-carriers used in IEEE 802.11g). The higher data sub-carriers helped to improve the highest data rate per stream to 65 Mbps as compared to 54 Mbps supported in IEEE 802.11g. 802.11n utilizes the Multiple-Input Multiple-Output (MIMO) technology. A MIMO system (represented by N x M) has N transmitters and M receivers. MIMO utilizes Spatial Multiplexing to transmit two or more parallel data streams in the same frequency channel. Using MIMO and Spatial Multiplexing 802.11n doubles its transmission capacity (130 Mbps). The transmission capability was improved by transmitting and receiving two parallel spatial data streams over two transmitters at the same time.

IEEE 802.11n also implemented the efficient frame aggregation and block acknowledgement mechanisms to improve throughput. 802.11n introduced frame aggregation mechanisms called Aggregated MAC Service Data Unit (A-MSDU) and Aggregated Multi-Protocol Data Unit (A-MPDU) to reduce the overhead of IEEE 802.11n packets. Frame aggregation aims at combining payloads of various PHY/MAC frames such that header size is considerably less than payload size. A-MSDU aggregates MAC frame payloads. IEEE 802.11n also provides optional features that can be used to further improve the performance in the PHY layer. For example, a two adjacent 20 MHz channels can be bonded and a 40 MHz channel may be used (instead of 20 MHz) if desired.

IEEE 802.11n also expands the number of sub-carriers in 40-MHz channel to 108 sub-carriers.

802.11ac: The 802.11ac standard operates only in the 5 GHz band. Theoretically, 802.11ac proposes data rates over 1 Gbps. The new specifications are based on the 802.11n standard, by expanding the channel bandwidth to 80 MHz and adding optional 160 MHz channels. In addition, 802.11ac utilizes MIMO with up to 8 spatial streams and a higher order modulation scheme called 256-Quadrature Amplitude Modulation (256-QAM). The standard also provides support for other advanced features such as beamforming and Low Density Parity Check (LDPC). In the beamforming process, an access point utilizes more than one antenna to transmit a signal. The multiple signals are sent to client devices to receive feedback about the best transmission path to the client. Beamforming is used to create a directional Radio Frequency (RF) beam while LDPC is utilized by 802.11ac to improve the Signal-to-Noise Ratio (SNR). We discuss further details about beamforming in section 4.

802.11ad: The 802.11ad standard operates in the unlicensed 60 GHz band. The standard proposes a theoretical data rate of 7 Gbps with low power consumption. 802.11ad is expected to employ wide channels of 2.16 GHz. This standard is intended to support high-performance wireless implementations such as high-definition video.

III. DRIVERS OF GIGABIT WIRELESS SPEEDS

The rapid proliferation of wireless devices at homes and work places has given rise to huge demands for higher wireless speeds and wider coverage areas. The ever-increasing need for mobile connectivity and high throughput requirements of present-day wireless devices and applications drive the need for deploying wireless technology with significant enhancements in terms of speed, reliability and Quality of Service (QoS). IEEE 802.11ac has been developed to meet the requirements of new mobile devices such as smartphones, tablets, laptops, and smart TVs. 802.11ac has also been established to fulfill the demands of bandwidth-intensive and latency-
sensitive applications. Figure 2 illustrates a typical scenario of how the 802.11ac technology can be employed in homes. The main drivers behind the demand for gigabit wireless speeds can be summarized as follows:

Increased Usage of Video Streaming Applications: the use of video applications has increased exponentially in the last decade. Many video streaming applications are being run by wireless users. Typical video usage includes gaming applications, digital home entertainment, videos for patient care, instructional videos in education, video presentations in corporate environments, etc. Demand for the ever-increasing, bandwidth-intensive applications has given rise to the need for gigabit wireless speeds.

High-density WLANs: the rise in the usage of all kinds of wireless devices such as smartphones and tablets has resulted in increased traffic in wireless networks. The heavy wireless traffic plays a major role in driving the demand for gigabit speed technology. Furthermore, with the latest trend of corporations to switch from having employees to bring single computing device to BYOD (Bring Your Own Device) paradigm, networks are experiencing increased load due to each person connecting multiple gadgets to the network. In order to provide high WLAN capacity, corporations are looking forward to the deployment of gigabit WLANs.

Popular Use of Latency-Sensitive Applications: various such as Voice over IP (VoIP), audio and video streaming, real-time videoconferencing, etc. require stringent low network latencies. With newer wireless technology that can offer higher bandwidths, these latency-sensitive applications can run more efficiently and provide enhanced user experience. Hence, there is a demand for improved transmission techniques that can support increasing network loads and reduced transmission delays.

Increased Usage of WLAN-enabled devices in Education: the heavy usage of WLANs at schools and university campuses is yet another driving force behind the deployment of gigabit wireless technologies. With the advent of mobile devices, students now carry multiple WLAN enabled devices laptops, tablets, smartphones on campus and frequently run bandwidth-intensive applications such as streaming videos, making audio/video calls, or playing live or pre-recorded lecture materials and videos. The high consumption of network bandwidth calls for implementing high-speed wireless access technologies that can deliver higher network throughput in the future.

Use of WLAN-enabled devices in Healthcare: many healthcare applications heavily depend on reliable and high-speed wireless connectivity include cardiac and radiology imaging, telemedicine, health scanners, etc. Running these healthcare applications efficiently requires high network capacity, sustained performance, and ubiquitous wireless coverage. The availability of gigabit wireless speeds will allow medical applications to stream multimedia content smoothly to physician’s wireless devices. Doctors, nurses and caregivers want to use gigabit wireless technology more often because it improves efficiency and provides faster access to healthcare services.

IV. TECHNIQUES FOR ENHANCING WIRELESS SPEEDS IN IEEE 802.11AC

In this section, we describe briefly the advances and improvements made to IEEE 802.11ac, with further improvements. The following are the main enhancements in the PHY layer of 802.11ac [13][14]:

Improved Channel Width: 802.11ac supports channel widths of 20, 40, 80, and 160 MHz. The maximum channel width supported in previous 802.11 standards did not exceed 40 MHz (as in the case of 802.11n). With more than a two-fold increase in the channel width, 802.11ac has the capability to offer much higher data rates.

Denser Modulation Method: 802.11ac adds 256 Quadrature Amplitude Modulation (QAM) to OFDM and achieves increased throughput as compared to 64-QAM.

Enhancement of Spatial Streams: 802.11ac supports up to 8 spatial streams and therefore increases the cumulative data speed. For example, an aggregate data rate of 1.3Gbps can be achieved using three spatial streams in 802.11ac.

Improvements to MIMO Technology: Multi-User, Multiple-Input Multiple-Output (MU-MIMO) [10] is an enhancement of the MIMO technology that allows simultaneous downlink transmission of separate streams to different clients in the same channel. MU-MIMO can only be used when the access point transmits to the client devices (downlink). This feature cannot be used when client devices transmit to the access point (uplink). Downlink MU-MIMO improves throughput when transmitting to multiple single-stream clients (e.g., smartphones) or multiple multi-stream clients (e.g., PCs). Simultaneous downlink transmission in MU-MIMO is illustrated in Figure 3.

Figure 3. Multi-User Multiple-Input Multiple-Output (MU-MIMO) in 802.11ac

Beamforming: 802.11ac specifies a standard implementation procedure for the beamforming technology. Beamforming is expected to improve
bandwidth utilization and increase the range of the wireless network. Conventional 802.11 access points used antennas that were omnidirectional. Omnidirectional antennas keep the radio channel busy in all directions. Beamforming focuses energy toward a client, as shown in Figure 4. The colored path shows the area where beamforming focus increases power, and thereby contributes to the improvements in the signal-to-noise ratio and data rate.

Figure 4. Beamforming (Left) versus Omnidirectional (Right) Wi-Fi Antennas

B. MAC Layer Enhancements

The original MAC layer of WLANs was designed to only support one-on-one communication. Later and in order to enable a WLAN Access Point (AP) to transmit simultaneously to multiple WLAN clients, certain modifications have been implemented at the MAC layer. Some of these main changes include the implementation of Transmit Opportunity (TXOP) Sharing and Enhanced Aggregation.

1) TXOP Sharing

The MAC protocol employs a mandatory contention-based channel access function called Distributed Coordination Function (DCF), which is based on the Carrier Sense Multiple Access (CSMA) mechanism. Mobile stations deliver MAC Service Data Units (MSDUs) after detecting that there is no other transmission on the same wireless medium.

The Enhanced Distributed Channel Access Method (EDCA) is an IEEE 802.11 MAC access method in which a station with high priority traffic has a lower waiting time (for transmission) than a station with low-priority traffic. EDCA provides contention-free access to the channel for a period called a transmit opportunity (TXOP) [1]. A TXOP is a bounded time interval during which a station can send as many frames as possible (as long as the duration of the transmissions does not extend beyond the maximum duration of the TXOP). If a frame is too large to be transmitted in a single TXOP, it is fragmented into smaller frames.

In the case of the legacy TXOP, only frames belonging to the same Access Category (AC) are transmitted and multiple frames belonging to different ACs cannot be simultaneously transmitted. Access Categories (ACs) are four QoS classifications (voice, video, best effort and background) made by Wireless Multimedia Extensions (WME) [4].

TXOP sharing, also referred to as the Multi-User Transmit Opportunity (MU-TXOP) mechanism, extends the legacy 802.11 TXOP concept to support Down-Link (DL) MU-MIMO. TXOP Sharing defines two types of Access Categories (primary and secondary), and two types of destinations (primary and secondary). A Primary AC is the AC that wins the TXOP for channel access after both external and internal competitions. There is only one primary AC at any time. A Secondary AC is an AC that does not win a TXOP but wants to share the TXOP obtained by the primary AC for simultaneous transmissions. There could be multiple secondary ACs at any time. Primary destinations are those targeted by the frames belonging to the primary AC. There could be one or more Primary destinations at any time. Secondary destinations are those targeted by the frames belonging to secondary ACs. There could be one or more Secondary destinations at any time.

For each AC, an enhanced version of the DCF, called Enhanced Distributed Channel Access function (EDCAF), contends for TXOPs using a set of EDCA parameters. Once an EDCAF wins a TXOP it becomes the owner of the TXOP. Its corresponding AC becomes the Primary AC, and its corresponding destination(s) become Primary destination(s). The Primary AC can share the TXOP with other ACs that did not win the medium access – the TXOP becomes the MU-TXOP. These other ACs become Secondary ACs and their corresponding destinations become Secondary destinations. The AP then groups the Secondary destinations together with the Primary destination(s) for simultaneous transmission (e.g. using GroupID). When the primary AC allows secondary ACs to share the TXOP for simultaneous transmissions, it results in Multi-User TXOP (MU-TXOP).

Figure 5. Multi-User Transmit Opportunity (MU-TXOP) Sharing
Figure 5 shows an example of MU-TXOP Sharing [7]. Here AC_VI is the primary AC and has 2 MSDU frames ready for transmission. The first MSDU is destined for STA-1 and the second MSDU is destined for STA-3. Since, STA-1 and STA-3 are destinations for a MSDU belonging to the primary AC, STA-1 and STA-3 become the Primary destinations. AC_VO and AC_BE are secondary ACs and STA-2 is the Secondary destination. Higher priority traffic (AC_VI) is transmitted earlier than lower priority traffic (AC_VO, AC_BE).

2) Enhanced Aggregation

Two forms of aggregation were included in IEEE 802.11n. These include the Aggregation MAC Service Data Unit (A-MSDU) (shown in Figure 6(a)) and the Aggregation MAC Protocol Data Unit (A-MPDU) (shown in Figure 6(b)).

As shown in Figure 6, several MSDUs with the same destination address are concatenated to form an A-MSDU. The A-MSDU is encapsulated within a MPDU. Several MPDUs are aggregated to form an A-MPDU. For each MSDU sub-frame in an A-MSDU frame, the MSDU sub-frame includes the Sub-frame Header, the MSDU data payload and the Padding field. The Sub-frame Header includes three fields: the Destination Address (DA), the Source Address (SA) and Length that indicates the MSDU data payload. A-MSDU aggregation can only be done for packets with the same SA and DA. A single A-MSDU contains multiple MSDU sub-frames. A single A-MSDU frame is transmitted after adding the Physical Header, the MAC header and the FCS field. The principle of A-MPDU is to send multiple MPDU sub-frames with a unique PHY header so as to reduce the overhead of the PHY header. For each A-MPDU, every MPDU sub-frame includes an MPDU frame, the MPDU delimiter and the padding bytes. Multiple MPDU sub-frames are concatenated into one larger A-MPDU frame. All the MPDU sub-frames within an A-MPDU should be addressed to the same receiver, but the MPDU sub-frame could have a different source address. The aggregation process was proposed to improve the MAC efficiency in 802.11n. 802.11ac further improves the MAC efficiency by implementing Enhanced Aggregation wherein A-MSDU and A-MPDU have further extensions in their lengths. In the case of 802.11ac, an A-MSDU can have a maximum length of 11426 bytes, and an A-MPDU can have a maximum length of 1048579 bytes [2].

V. CURRENT CHALLENGES IN ACHIEVING GIGABIT THROUGHPUT WITH 802.11AC

The new characteristics of the 802.11ac are significant and constitute a major advancement in the development of WLAN technologies. However, the theoretically calculated data rate of 1.3 Gbps at the physical layer has not been practically achieved thus far by the end-user. Several vendors have tested their 802.11ac products and identified factors that can vary the performance of 802.11ac.

Table II. Observed Throughput using an 802.11ac Access Point with Different Devices

<table>
<thead>
<tr>
<th>Device</th>
<th>Achieved Throughput (Mbps)</th>
<th>No. of Spatial Streams supported by device</th>
</tr>
</thead>
<tbody>
<tr>
<td>MacBook Air Laptop</td>
<td>400</td>
<td>2</td>
</tr>
<tr>
<td>Samsung Galaxy S4</td>
<td>218</td>
<td>3</td>
</tr>
<tr>
<td>Smart Phone</td>
<td>218</td>
<td>1</td>
</tr>
<tr>
<td>Wireless Bridge</td>
<td>722</td>
<td>3</td>
</tr>
</tbody>
</table>

Tests conducted by an Aruba Networks [11] showed that when the wireless network was loaded with 120 clients, an 802.11ac equipped MacBook Air laptop achieved a throughput of 400 Mbps. Similarly, results of tests conducted by Cisco [13] using the 802.11ac access point are shown in table 2. It can be observed that a Samsung Galaxy S4 smart phone (with one spatial stream) client yielded a throughput of about 218 Mbps while a wireless bridge (with three spatial streams) yielded a throughput of about 722 Mbps. The observed throughputs show that increase in throughput can be achieved in devices supporting more than one spatial stream. The tests also reveal that although the throughput obtained using 802.11ac is definitely better than that obtained via 802.11n, however the measured throughput using 802.11ac has not practically achieved the theoretical gigabit rates of greater than or equal to 1Gbps.

The maximum achievable data rates for 802.11ac are dependent on the following factors:
1. Number of spatial streams used
2. Use of 80 MHz wide channels
3. Use of the 256-QAM modulation scheme
4. Use of the Transmit Beamforming
5. Use of the Low Density Parity Check (LDPC)

There are various reasons why the first generation of 802.11ac devices (referred to as Wave 1) cannot deliver the theoretical maximum physical data rate of 1.3 Gbps:

**Limitations of client devices:** to achieve the maximum speed, all three spatial streams should be utilized. However a large percentage of client devices are smart phones and they support only a single spatial stream. Most 802.11ac access points being deployed currently support up to three spatial streams. The 802.11ac enabled client devices support several spatial streams (simultaneously). For example, the Apple’s MacBook Air and Samsung Galaxy S4 can only support one spatial stream.

**Issues in deploying 80 MHz and 160 MHz channel bonding:** WLAN bandwidth can be increased by bonding multiple channels together. 802.11ac allows the creation of 20, 40, 80, or 160 MHz wide channels. The 160 MHz channel can also be a combination of two non-contiguous 80 MHz channels (80+80). From a Radio Frequency (RF) planning perspective, there are a few obstacles to using these wider channels. For example, a wider channel is more susceptible to RF interference from neighboring wireless networks. Although channel bonding increases bandwidth, wider channels are more susceptible to signal interference that may lead to reduced range and poorer signal quality.

**Complexities involved in utilizing 256-QAM Modulation:** to enable higher data rates, 802.11ac has introduced a novel and efficient modulation technique. In contrast to the conventional 64-QAM modulation, 256-QAM enhances efficiency by approximately 33%. 256-QAM modulation increases modulation complexity by another order of magnitude, representing 8 bits with each constellation point. However, 256-QAM modulation can only be used in scenarios that have high Signal-to-Noise (SNR) ratios, or in very favorable channel conditions (low interference) such devices operating close by in a home. Furthermore, in order to support 256-QAM, the transmitter and receiver need to be designed such that transmit and receive Error Vector Magnitude (EVM) is able to support the higher constellation. Therefore, RF design of a system supporting 256-QAM remains a significant challenge [12]. With currently available products, 256-QAM works only within a very short range (10-20m). Hence, the data rate gains proposed by 802.11ac depend on the deployment scenario.

**Beamforming and Low Density Parity Check (LDPC):** these features in 802.11ac can be used when the client device is in close range with the 802.11ac access point. When the client device moves farther away from the access point, the 802.11ac radios switch to narrower channels and lower modulations.

In addition to the aforementioned technical limitations that still need to be overcome to achieve high 802.11ac throughput, there are also challenges related to security, capacity planning and deployment, and interoperability. For security, Galois/Counter Mode Protocol (GCMP) is the recommended encryption protocol to be used for 802.11ac. GCMP gives much higher performance than its widely adopted predecessor Counter Mode with CBC-MAC Protocol CCMP that is used in 802.11n. The challenge with GCMP is still computational overhead as the data rate approaches the gigabit rates. This leads to additional cost in chipset design to accommodate for newer generation of crypto accelerator processors to speed up the internal encryption algorithm of GCMP. More importantly, GCMP protocol is not backward compatible with existing GCMP protocol, and that poses a major deployment and migration challenge that needs to be addressed.

Capacity planning and deployment are also major challenges to ensure the seamless adoption of 802.11ac. The wired core network capacity has to be adequate enough to meet the new multi-gigabit Wi-Fi clients connected to 802.11ac APs. Proper capacity planning for the core links, switches, and routers should be at least upgraded to 10 Gbps. Above all, legacy network middleboxes of firewalls, web servers, and intrusion detection should also be upgraded to meet the enormous expected increase of incoming and outgoing gigabit traffic to end users; otherwise, user experience will be relatively poor as these network middleboxes will not keep up with the multi-gigabit network demand. In addition, gradual deployment and migration are more practical than a full 802.11ac rollout to ensure smooth and seamless transition.

Finally, interoperability in terms of performance and management features of the different 802.11ac vendor chipset and product offerings can be a big challenge. This may have as much impact on the overall system performance as 802.11ac itself. So, it is advisable to perform thorough and comprehensive interoperability testing among different vendor products and solutions prior to making such products and chipsets available on the market.

VI. CONCLUSION

In this paper, we have presented recent technological advancement and developments that have enhanced wireless speeds and led to the deployment of various WLAN standards. We have highlighted the primary drivers that have created the need for gigabit wireless speeds. Early evaluations of the latest 802.11ac products (Wave 1) have demonstrated that IEEE 802.11ac provides better throughput than its predecessor (802.11n). However, it has been observed that several new features (beamforming, support for spatial streams, channel bonding, advanced modulation) designed for 802.11ac with the intent of maximizing throughout beyond 1 Gbps are currently not being utilized to their full capability. Therefore, delivering gigabit user throughput to IEEE 802.11ac users currently remains a challenge. Despite the present limitations, the 802.11ac standard shows significant potential to achieve gigabit user throughput in the right environmental conditions and with the deployment of mobile devices that support higher number of spatial streams.
REFERENCES


Farhan Siddiqui received a B.E. (2000) degree in Computer Science from Osmania University, India, and M.S. (2003) and Ph.D. (2007) degrees in Computer Science from Wayne State University, MI, USA. She was a full-time faculty member at Bradley University, IL, USA during Fall 2008. Since March 2009, she is a faculty member in the School of Information Systems and Technology at Walden University, MN, USA. Her current research interests are in Wireless Networks, Mobile Computing, Internet-of-Things (IoT), and Network Security.

Sherali Zeadally is an Associate Professor at the College of Communication and Information, University of Kentucky, Lexington, KY, USA. He received his bachelor and doctorate degrees in computer science from the University of Cambridge, Cambridge, England, and the University of Buckingham, Buckingham, England, respectively. He is a Fellow of the British Computer Society and a Fellow of the Institution of Engineering Technology, England.

Khaled Salah is an Associate Professor in the Electrical and Computer Engineering Department, Khalifa University of Science, Technology and Research (KUSTAR). He received the B.S. degree in Computer Engineering with a minor in Computer Science from Iowa State University, USA, in 1990, the M.S. degree in Computer Systems Engineering from Illinois Institute of Technology, USA, in 1994, and the Ph.D. degree in Computer Science from the same institution in 2000. He has over 10 years of industrial experience in software and firmware development. He joined KUSTAR in August 2010. Prior to joining KUSTAR, Khaled was an associate professor in the Department of Information and Computer Science, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia. Khaled has been teaching graduate and undergraduate courses and has over 100 publications in the areas of cloud computing, computer and network security, operating systems, computer networks, and performance evaluation. Khaled is an Editorial Board member of a number of prestigious international journals including IET Communications, IET Networks, Elsevier JNCA, Wiley IJNM, Wiley SCN, and J.UCS. Khaled was the recipient of Khalifa University Excellence in Research Award of 2008/09, KFUPM University Excellence in Research Award of 2008/09, and KFUPM Best Research Project Award of 2009/10, and also the recipient of the departmental awards for Distinguished Research and Teaching in prior years.