



**HAL**  
open science

# Defining the communication architecture for data aggregation in Wireless Sensor Networks: application to communicating concrete design

Hang Wan, Michael David, William Derigent

## ► To cite this version:

Hang Wan, Michael David, William Derigent. Defining the communication architecture for data aggregation in Wireless Sensor Networks: application to communicating concrete design. 7th International Conference on Future Internet of Things and Cloud, FiCloud 2019, Aug 2019, Istanbul, Turkey. 10.1109/FiCloud.2019.00022 . hal-02421238

**HAL Id: hal-02421238**

**<https://hal.science/hal-02421238v1>**

Submitted on 13 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Defining the communication architecture for data aggregation in Wireless Sensor Networks: application to communicating concrete design

Hang WAN, Michael DAVID, William DERIGENT  
Université de Lorraine, CRAN, UMR 7039  
Vandoeuvre-lès-Nancy Cedex, 54516, France.  
{hang.wan, michael.david, william.derigent}@univ-lorraine.fr

**Abstract**—Wireless sensor networks consist of different sensor nodes, able to monitor inaccessible environments. Our objective is to improve the lifetime of specific WSNs in the context of communicating material. And we apply it to smart concrete design. The McBIM project, its objectives and constraints are introduced with some existing solutions for data collection. Protocols including aggregation features have been preferred, in order to reduce the number of messages and optimize network lifetime. Three main methods in data aggregation are presented with their main advantages and drawbacks: chain-based, cluster-based and tree-based methods. Furthermore, a comparison of the different associated algorithms is proposed based on required performance metrics, related to our context of application. Finally, some perspectives for the future work are presented.

**Keywords**—Internet of Things, wireless sensor networks, communicating material, data aggregation, energy efficiency

## I. INTRODUCTION

The advancement of technology in microprocessor and communication network makes the Internet of Things (IoT) possible. In 2009, the Research Center for Automatic Control (CRAN) began to study the “communicating materials” concept [1], defining a class of material able to process, store and communicate data with its environment. This concept has been first studied by Kubler [2] and further applied by Mekki [3] to the construction industry. The McBIM Project [4] (Material communicating with the BIM - Building Information Modelling) aims to design a “communicating concrete”. Namely, it is the concrete equipped with embedded wireless micro-sensor network (WSN), which can measure the physical environment, store the information and exchange data with BIM platforms. Another objective is to demonstrate the usefulness of this new concept for different phases of a building's lifecycle, especially the construction and exploitation phases.

The next section introduces the McBIM project and explains the data collection problems occurring in this context. The constraints taken into account for the data collection are defined and the problem statement is clearly enounced. Section III presents the methods classically used for structuring the data collection in WSN. Section IV provides a synthesis of these solutions and discusses their usability to our application case. Last section concludes and gives some perspectives for the development of this work.

## II. PROBLEM STATEMENT

### A. On the McBIM context

The concept of “communicating materials” is applied in the McBIM project's framework. The communicating concrete consists of many sensing and communicating nodes

as illustrated in Fig. 1. The sensing nodes will periodically (by example every hour) monitor the physical parameters (like temperature, humidity ...) of the concrete. Communicating nodes aggregate received data and transmit them to monitoring center thanks to BIM standard. Besides, manufacturing data (like its physical properties or manufacturing actors) may be stored locally. The main constraint is that the concrete element has to live several decades from the manufacturing phase to the latest of the exploitation phase.

Both sensing nodes and communicating nodes are inserted and initialized during manufacturing phase. Communicating nodes are placed in line within concretes, each one communicates with its surrounding sensor nodes via a star network. The detected information can be used during different phases of the concrete lifecycle i.e. the manufacturing, the inventory management and the logistics of those concretes. To capture this information, some data collection algorithms must be implemented in the embedded WSN. In the framework of the McBIM project, concrete experts declared that concrete beams can be used several years after their manufacturing. In this case, an energy-efficient data collection technique is then required for ensuring the concrete's long-time service, at least twenty or thirty years.

In order to select suitable algorithms, we first study the communicating concrete's behaviors along its lifecycle (see Fig. 2) in order to specify some performance metrics for selecting potential solutions. After communicating concretes are produced, they are stored in an orderly manner according to their batch. They may both have similar physical status before reaching available state for delivery. Therefore, the embedded WSN of these concretes can work together with a static 2D network to report its status instead of working separately. These monitoring data are accessible directly via a reader device or remotely via the internet. During the construction phase, communicating concretes will be assembled together. In this case, auto-organization is then needed to dynamically define a 3D network [5] to achieve energy savings over this phase. When the construction is

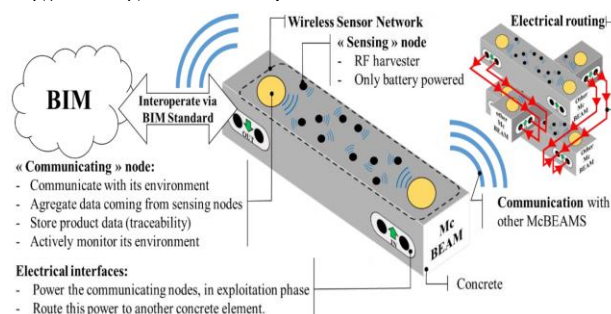


Fig. 1. McBIM (Material communicating with the BIM)

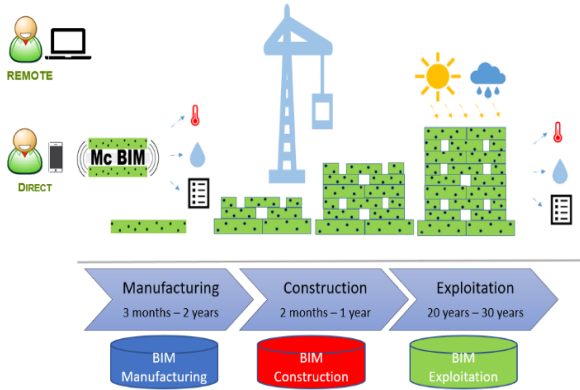


Fig. 2. McBIM communicating concrete lifecycle

completed, a large 3D static WSN will regularly monitor struct health data (such as cracks, temperature, corrosion, etc.) to ensure the safety of the building.

As mentioned above, during manufacturing, the communicating concrete is mainly used to monitor its status. The network is small and static. Therefore, a *simple* and *efficient* data aggregation protocol will be preferred. For construction, the problems are not only the minimization of energy within the concrete beam, but also the re-organization of network when new concrete beams arrive. Thus, a data aggregation technique favoring *dynamic* and *fast* network reconfiguration should be privileged. After the construction, a large network will be completed and used for the exploitation phase. Consequently, a *robust* with a *low maintaining cost* protocol will be adapted.

To sum up, the required performance metrics for data aggregation routing protocols are described as follows: *energy efficiency*, represents the algorithm efficiency for the data collection. It is computed by the amount of consumed energy by the number of received messages; *Maintaining cost* shows its energy consumption for reconstruction of the architecture; *Resilience to link failure* indicates its robustness facing the loss of link when a node dies; *Scalability* stands for its capability to scale the scenario;

### B. On WSN data collection

As depicted in Fig. 3, depending on application, density of nodes and communication protocol, a WSN architecture can be *one-hop* or *multi-hops* from a sensing node to a sink node or Base Station (BS). If a WSN is preliminary organized, it can also use coordinator nodes to deliver data to the BS.

In a clustered architecture, the nodes triggered by the supervisor organize themselves into clusters, with one node per cluster acting as Cluster Head (CH). Other nodes transmit packets to its CHs, which then transmit the received packets plus the one generated by themselves to the supervisor via a direct link. Therefore, being a CH is much more energy intensive than being a non-CH node. If the CHs were chosen a priori and fixed throughout the network lifetime, these nodes would quickly use up their limited energy: once the CH runs out of energy, it is no longer operational. Thus, it is assumed that the algorithm incorporates a randomized rotation of the CH role among the nodes in the network: at each round, a node autonomously decides to elect itself CH with a certain probability. Decisions taken in different rounds are uncorrelated. In this way, the energy load of being a CH is evenly distributed among nodes during network lifetime.

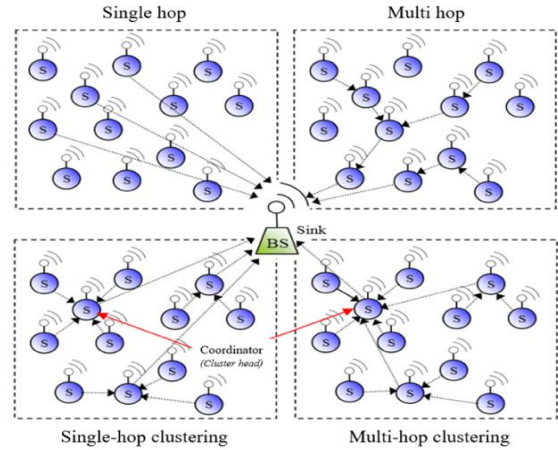


Fig. 3. Different WSN architectures for data collection

### C. On data aggregation for WSN

In order to reduce the overall energy consumption due to the transmission of messages, it is useful to partially decentralize the processing task. Considering WSN architecture, most nodes are not directly connected to the sink, its data reach the final destination through intermediate nodes. These intermediate nodes can be used to perform partial processing. Data aggregation techniques can be adopted to reduce the amount of transmitted data. Typical aggregation techniques take advantage of the correlation among adjacent samples. In general, this operation depends on the process statistics, spatial correlation, data semantic correlation.

Aggregation in WSN, as described in [6], needs three elements to be effective: an adapted structure of the network (for routing the data); aggregation functions (for fusing the data); a data representation (for exploiting the data). Second and third elements are problems mainly depending on the application case. “Aggregation functions” depends on the service which is offered by the WSN and “Data representation” consists of the final exploitation of data by users (can be considered as an external problem for the WSN). In this work we just focus on the preliminary step for aggregation, the WSN structure.

## III. RELATED WORK

In order to ensure a long-time service, we look for energy efficiency solutions for WSN data collection. To do this, several strategies exist, and among them, we focus on data aggregation techniques to minimize the number of messages. There are three main data aggregation architectures (chain-based, cluster-based and tree-based) detailed hereafter.

### A. Chain-based architecture

A chain-based architecture is a WSN architecture where nodes communicate via a line structure. To construct this structure, several routing protocols have been proposed and are detailed as follows.

PEGASIS (Power-Efficient gathering in sensor information systems) [7] is the first chain-based routing protocol proposed for WSN. As illustrated in Fig. 4, each node only communicates with its closest neighbors, and takes turn transmitting data to the BS. In this case, energy consumption of transmission per round is reduced and the network lifetime is increased with a random selection mechanism of the leader (in Fig. 4, c2 is selected as the leader for this round).

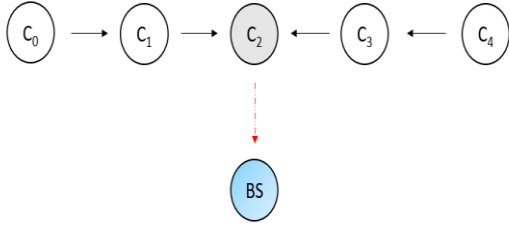


Fig. 4. Data collection in PEGASIS

In this approach, all nodes have global knowledge of the network and employ the same greedy algorithm. Each node can directly transmit to the BS. The computation of chain can be done at the base station, it then broadcasts the chain information to nodes. Or it can be constructed in a dynamic manner by all the sensor nodes. The dynamic construction begins with the furthest node. Each node performs the greedy algorithm to select its closest unvisited neighbor as next. In addition, they take turns to act as the leader with the probability:  $i \bmod N$  where  $i$  is the node number and  $N$  is the number of nodes. The chain will be reconstructed if a node dies.

Once the chain is constructed, the BS synchronizes all sensor nodes for each round. A time slot is used for transmitting data in PEGASIS. The data collection is shown as Fig. 4. Where node  $c_2$  is the chain leader. Data collection begins with one of the end nodes. For example, the end node  $c_0$  transmits its data to node  $c_1$  in slot one, then  $c_1$  performs data fusion and transmit data to node  $c_2$  in slot two. It then continues with another side of this chain, where the end node  $c_4$  transmit its messages until all messages arrive the leader. At the end, the leader  $c_2$  transmits the aggregated data of this chain to BS.

Since data fusion is performed at nodes, PEGASIS reduces the energy consumption in transmission per round. Moreover, its random leader selection mechanism balances the consumption for the whole network and improves the lifetime. However, it is hard to ensure all nodes can connect with the BS in large WSNs. A long chain may have an important impact on the delay of transmission. Moreover, the greedy algorithm in this protocol only considers the distance between nodes. The remaining energy at each node and the delay of transmission should be taken in account.

Based on PEGASIS, many improved algorithms have been proposed. A large part of these algorithms are related to the chain-cluster based approach, such as CCM (Chain-Cluster based Mixed routing) [8] in Fig. 5, CRBCC (Chain routing based on coordinates-oriented cluster) [9] in Fig. 6. This kind of protocol combines the advantages of the chain-based protocols and the cluster-based. It usually divides the networks into clusters. In each cluster, nodes form into a chain, a node will be selected as the chain leader. Between the different clusters, the selected cluster leaders then form a high-level chain, and one of them is selected to transmit information to the BS.

The differences between those algorithms are usually in the cluster partition and the construction of the chain routing. In CCM, the reference network is an evenly distributed two-dimension network, each node with coordinates  $(x, y)$ . The

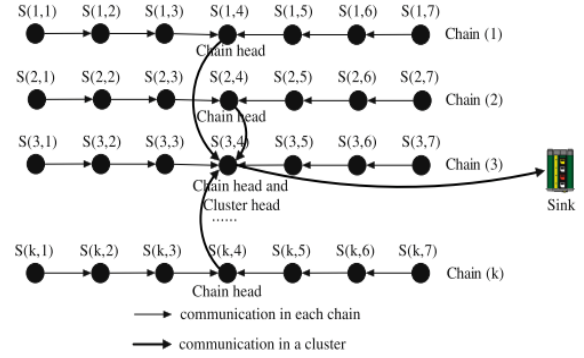


Fig. 5. CCM protocol [8]

cluster partition only depends on the  $x$  coordinate. During the  $j^{\text{th}}$  round, the nodes  $S(i, j)$  are elected as leaders where  $i, j$  are the horizontal and vertical coordinates. In a chain, a node fuses its own data with the received data from its neighbors, then transmits it to next hop until the chain head. After the completion of chain-based routing in a cluster, the leaders for each basic chain form a high-level chain, where the high-level leader is chosen by its residual energy. In this high-level chain, all nodes transmit its data directly to the leader. Then the leader fuses all data and transmits to the BS.

Compared with CCM, nodes are randomly distributed in CRBCC. As illustrated in Fig. 6, the network is divided according to the  $y$  coordinate and the numbers of nodes where each cluster have almost the same number of nodes. Instead of using the greedy algorithm, CRBCC uses Simulated Annealing (SA) algorithm to form the low-level chain (into a cluster) as well as the high-level chain (between clusters). The initial objective of SA algorithm is to compute the shortest way. In this algorithm, SA is used to calculate the chain routing with the lowest energy consumption. This kind of chain-cluster approach avoids the long chain disadvantage and reduces the delay. However, it increases the clustering overhead (i.e. the number of messages exchanged to construct the clustering), and the reconstruction of network is more complex than with the single chain.

Another kind of approach is the sub-network chain-based routing protocol. A Balanced Chain-Based Routing Protocol (BCBRP) has been proposed in [10]. It is shown in Fig. 7. The idea behind this protocol is that it divides the network into different areas to reduce the energy consumption in maintaining of the WSN. The chain construction in BCBRP has four phases. First, the network is divided into different

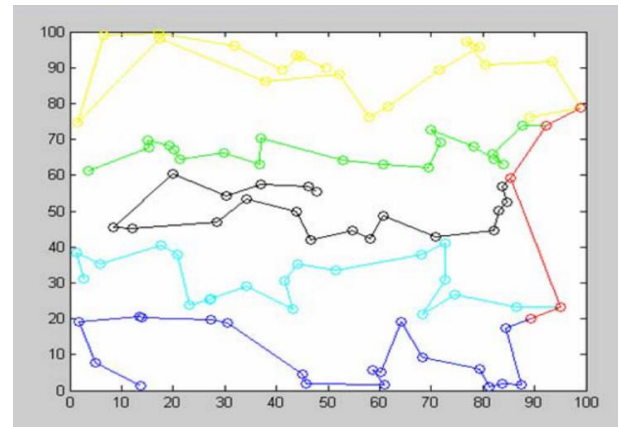


Fig. 6. CRBCC Protocol [9]



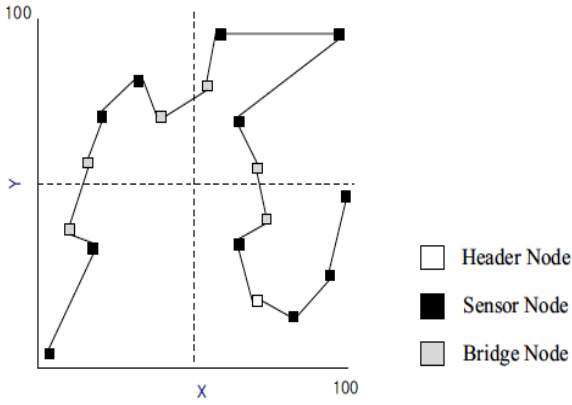


Fig. 7. BCBRP protocol [10]

equal parts. Second, the closest nodes to the boundary are elected as bridge nodes which connect with the others sub-networks. Third, an algorithm similar to the minimum spanning tree is used to construct the shortest chain in each part. Contrary to the greedy algorithm, the minimum spanning tree algorithm gives a global decision to connect all the nodes together with the minimum total route weight. Fourth, the chain leader is randomly selected from the area which has the largest number of nodes.

In BCBRP, when a node dies, the reconstruction of chain takes place only in a sub-network. Hence, the energy consumption used for maintaining the WSN structure is significantly reduced. It avoids the long chain disadvantages. However, the single chain will still cause more delay. Moreover, it is hard to ensure that all nodes could directly transmit to the BS.

### B. Cluster-based architecture

In a large WSN, the drawbacks of a single chain approach are obvious. Not only in terms of delays, but also in terms of network maintenance. In the literature, another structure is proposed: the cluster-based approach where nodes are regrouped into clusters. Within a cluster, a node is elected as the cluster head (CH) which receives information from its cluster members, aggregates and sends it to the sink.

LEACH (Low Energy Adaptive Clustering Hierarchy) is the first proposed cluster-based protocol [11]. As illustrated in Fig. 8, it is a distributed and self-organized clustering protocol. A random mechanism is used to rotate the CHs in order to balance the energy consumption. LEACH works with two phases for one round: *setup phase* and *steady state phase*. During the setup phase, all nodes organize themselves into clusters. A distributed probabilistic approach is used to elect the cluster leader. The threshold  $P_i(t)$  is defined as (1):

$$P_i(t) = \frac{k}{N - k * (r \bmod (N/k))} * c_i(t) \quad (1)$$

Where  $k$  is a predetermined cluster number,  $N$  is the total number of nodes,  $r$  is the current round number,  $i$  is the node's id and  $c_i(t)$  is a function which indicates if the node has already been a CH in a previous round. For each round, sensor nodes randomly pick a number from 0 to 1, and then compare the value with the threshold. It becomes a leader if the number is bigger than the threshold. This probabilistic approach balances the energy at each node. When a CH is elected, it broadcasts a message to announce that it is the cluster leader.

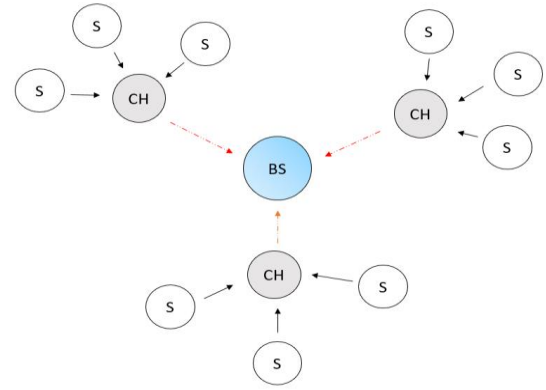


Fig. 8. LEACH approach

Surrounding nodes then decide which cluster to join depending on the signal strength of the received messages.

The second phase is the data collection as illustrated in Fig. 8. All non-headers transmit data to its CH. The leaders then transmit data directly to the sink after receiving all data. This algorithm improves the lifetime of network, but the single-hop routing is inefficient for a large network. The leader consumes a large amount of energy when it is far from the sink. Furthermore, aggregation function could be implemented at the cluster-heads to reduce the transmission data.

LEACH-C is an improved version. It centralizes the cluster formation in a centralized manner at the sink to improve the performance of LEACH in term of the number of received data at the sink. TL-LEACH [12] is a multi-hop algorithm, which creates an inter-communication between clusters. Instead of a direct one-hop transmission in LEACH, the CHs pass its data through other CHs until the sink. This algorithm reduces the communication between CHs to the BS and improves the energy efficiency.

Unlike the random election of leader in LEACH, Hybrid Energy-Efficient Distributed Clustering (HEED) [13] takes residual energy and communication cost into account. An Average Minimum Reachability Power (AMRP) is defined for the estimation of communication cost. AMRP is the average of minimum power levels required for the non-leader nodes to connect with its CH. For each round of transmission in HEED, every node computes its probability  $P_{CH}$  which is defined as (2), and becomes the leader if its  $P_{CH}$  reaches 1.

$$P_{CH} = C * \frac{E_{residual}}{E_{max}} \quad (2)$$

Where  $C$  is the defined percentage of cluster heads,  $E_{residual}$  is the current residual energy, and  $E_{max}$  is the initial energy. Since the leaders are elected, the surrounding nodes then choose a CH with the lowest AMRP cost. Compared with LEACH, this residual energy-based election mechanism improves the network lifetime. An improved algorithm DWEHC (Distributed Weight-based Energy-efficient Hierarchical Clustering protocol) is proposed by Gupta in [14]. Like HEED, the residual energy is accounted in DWEHC with the cluster range for calculating the weight of each node. Within cluster, the no-leader nodes use the knowledge of distance to decide to be a *single-level* or *multi-level* member. This location awareness balances the cluster size and improves the energy efficiency.

### C. Tree-based architecture

Another widely used approach in WSN is the tree-based architecture, where the sensor nodes transmit data following a tree structure network via intermediate nodes to the sink. Processing operations such as data aggregation can be performed at intermediate nodes to reduce the data transmission and save energy.

An Energy-Aware Distributed Aggregation Tree (EADAT) is proposed in [15]. This approach computes the residual energy and the number of hops to the sink to construct the tree. In addition, a control message which contains five fields (ID, parent, power, status, hopCnt) is defined for the construction. Where *ID* and *parent* records the node's identification, and its parent node respectively; *power* represents its residual power; *status* indicates the state of the sending node in the network (undefined state, leaf node, non-leaf node, or danger state); *hopCnt* is the number of hops to reach the sink.

At the beginning of the construction process, the sink broadcasts initial control message. A node  $v$  sets up its timer after receiving the first message. The timer counts down when the channel is idle. However, the timer will be reinitialized if the node  $v$  receives another message. During this period, the node  $v$  chooses a neighbor which has higher residual power and shortest path to the sink as its parent. It then broadcasts a message with its own *hopCnt*. Another node  $u$  will make itself as a non-leaf node if it receives a message from  $v$  indicating its *parent<sub>v</sub>* value is  $u$ . Otherwise, it will be a leaf node of  $v$ . Once all nodes broadcast their message and find their corresponding status in the network, the construction of tree is completed.

For maintaining the network structure, an active node broadcasts "help" messages each  $T_d$  time units when the residual energy is smaller than a given threshold. It will then shut down its radio. If a node receives the help message from its parent, it chooses another node from its neighbors as its new parent if it exists. Otherwise, it enters in the danger state. Meanwhile, the sleep nodes periodically wake up and broadcast a "hello" message that records its path length to the sink. The node that is in danger state will invite a surrounding sleep node into the tree with the shortest length to the sink. The use of shortest path in this approach saves energy and extends the lifetime of network.

A Power Efficient Data gathering and Aggregation Protocol (PEDAP) was proposed by Tan in [16]. It is a near optimal minimum spanning tree-based (MST) protocol. PEDAP aims to improve the lifetime by minimizing the total energy consumption in communication per round. In this approach the computation of the routing tree is completed in a centralized manner at the BS. A Prim's MST algorithm that is a greedy algorithm for constructing MST with the minimum edge weight is used for the computation. In this approach, each edge  $e_{ij}$  between nodes  $i$  and  $j$  of the MST has a weight which is defined as (3), it is equal to the communication cost between these two nodes.

$$C_{ij} = 2 \times E_{elec} \times k + E_{amp} \times k \times d^2 \quad (3)$$

Where  $E_{elec}$  is the energy consumption of transmission or reception,  $E_{amp}$  is the consumption of the transmit amplification,  $k$  is the size of the packet and  $d$  is the distance.

The construction of the tree begins with the root. MST example is shown in Fig. 9. During several iterations, it adds

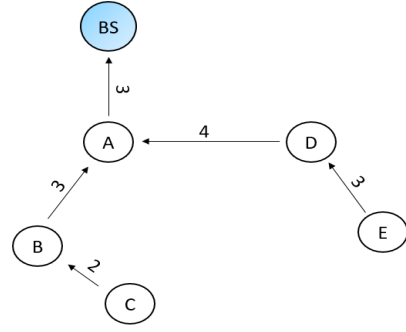


Fig. 9. Minimum spanning tree-based routing protocol

the vertex which has the minimum weighted edge into the tree. MST ensures a minimal weight route for the whole network. However, the running time complexity is  $O(n^2)$  where  $n$  is the number of nodes in the network.

In this protocol, the BS can estimate the remaining energy by knowing the location of all nodes among the network. It will recompute the routing after a certain number of rounds. A power aware version of this protocol PEDAP-PA takes the residual energy into account. In PEDAP-PA, the weight is replaced by  $EC_{ij}(k) = C_{ij}(k)/e_i$ . Where  $e_i$  is the normalized residual energy of node  $i$ . Compared with LEACH and PEGASIS, the authors prove that PEDAP and PEADP-PA significantly improve the time of the first dead node.

The Tiny AGgregation (TAG) approach is a data-centric tree-based protocol [17]. A declarative interface is used for data collection and aggregation. With a given query, the relevant information is sent to the sink periodically. This algorithm has two phases: the distributed phase and the collection phase. The construction of tree is established in the first phase. The sink first broadcasts a message which contains the distance to the sink. When a node (not already in the tree) receives a message, it sets the sending node as its parent. The node then broadcasts again the same message with its own ID and distance. After the tree is constructed, a semantic query which specifies the required type of data (such as temperature or humidity) will be sent by the sink to the network following the tree. The form of the query is shown as follows:

```

SELECT {agg(expr), attrs} from SENSOR
WHERE {selPreds}
GROUP BY {attrs}
HAVING {havingPreds}
EPOCH DURATION i
  
```

The query contains the desired data type (attrs), the type of aggregation for those data, and the time. During the collection phase, relevant nodes send its data to its parents then go to sleep. After received all data from its children, a parent node then fuses the data and transmits the aggregate message within a communicating slot. It equals to the time divided by the total level of the network. This in-network aggregation processing algorithm reduces the transmission at intermediate nodes and improves the network lifetime. However, it is not robust to link failure and re-organization is highly energy consuming.

## IV. SYNTHESIS AND DISCUSSION

The vast majority of research on WSN is based on simulations to optimize the network structure. The particular context of communicating material, which implies variable

applications needs over time, requires to have a more pragmatic approach. In this section, we compare useful algorithms from the three main architectures.

We have chosen protocols that using aggregation because it reduces the number of messages and therefore optimizes the network lifetime. Our strategy is therefore to identify the best data collection solutions within our specific context. An originality of our proposal is to be able to adapt the communication structure to meet the variable application needs. This is why we are looking to compare the mentioned protocols for each phase of the concrete element's lifecycle.

With the different approaches presented above, we create a classification to select suitable algorithms for our application case. In Table. I , protocols are compared with the required performance metrics defined in section II. A. Where energy efficiency depends on the algorithm's complexity for the phase of data collection, maintaining cost focus on the cost of re-organization of network and the initialization process. Resilience to link failure presents the protocol robustness. Scalability stands for its capability to scale the scenario. Thus, the plus and minus are used to represent the strength and weakness of these routing protocols. These plus and minus can also be converted in a numerical value for quantitative analysis using the following correspondence: ++→1; +→0.5; -→-0.5; --→-1.

As mentioned above, the requirements for data collection are different for different phases of the communicating concrete's lifecycle. We can then apply different coefficients to the performance metrics for each lifecycle phase. During the phase of manufacturing, the network is small and static. Its main job is the monitoring of the concrete's status. A simple and efficient protocol will be preferred. Therefore, the energy efficiency is the most important performance metric, the coefficients for the four-performance metrics are: 0.4; 0.2; 0.1; 0.3. During construction, the reorganization of network often occurs. A low maintenance cost routing protocol with a good scalability must be privileged. Therefore, the coefficient of scalability is higher than the others; meanwhile, the energy efficiency is still important. As consequence, the coefficients are: 0.3; 0.2; 0.1; 0.4. At the end, a large static 3D network will be constructed for exploitation phase, an efficient and robust routing protocol will be a good choice. For this reason, the energy efficiency and robustness take important places in this phase. Therefore, the coefficients of performance metrics for this phase are: 0.4; 0.2; 0.3; 0.1.

Table I. SUMMARY OF DATA AGGREGATION PROTOCOLS

Protocol name (Type)	Energy efficiency	Maintaining cost	Resilience to link failure	Scalability
PEGASIS (Ch)	-	--	-	--
CCM (Ch)	+	-	-	--
CRBCC (Ch)	+	+	+	-
BCBRP (Ch)	-	--	++	--
LEACH (Cl)	--	+	+	--
TL-LEACH (Cl)	-	+	+	-
HEED (Cl)	+	-	+	+
DWEHC (Cl)	++	-	++	+
EADAT (Tr)	-	--	-	+
PEDAP (Tr)	++	+	-	+
TAG (Tr)	+	++	-	+

With the performance metrics table and the coefficients, we can calculate the performance of each protocol for the three different phases. For example, the performance of PEGASIS is the sum of its performance metrics multiply the coefficients of manufacturing. In this manner, we obtain the indicator table as Table II. All mentioned protocols are compared for finding a better solution for each phase of the communicating concrete lifecycle. Besides, the mean column is used to indicate the protocols' average performance over the three phases. The larger the value is, better is the algorithm.

**During the manufacturing phase**, since the drawback of long single chain in maintaining cost and scalability, the chain-based protocols have a lower value than the others. As the energy efficiency is the most important performance metrics in this step. **DWEHC and PEDAP have similar and higher performance values in manufacturing.** DWEHC is more robust than PEDAP. However, maintaining cost for DWEHC is higher than PEDAP. Therefore, both of them may be considered as good choices before construction. TAG gives also a good performance indicator, but its energy efficiency is weaker.

**During the construction phase**, scalability is the most important performance metrics. Different communicating concretes are regrouped into a 3D network. Both cluster-based and tree-based protocols give good results. DWEHC has a high energy efficiency and a good resilience to link failure, but maintaining cost is high. TAG uses the semantic tree to improve energy efficiency. PEDAP achieves a better energy efficiency with MST. Therefore, **DWEHC, PEDAP and TAG seems good choices for the construction phase.**

**During the exploitation phase**, the communicating concretes work together in a large 3D static WSN. At this phase, the energy efficiency, maintaining cost and resilience to link failure are both important for data exploitation. The multi-level mechanism in **DWEHC provides a robust network**, taking the residual energy into account prevents the loss of links. It may be a good choice for periodical monitoring. Although CRBCC may have good result, it cannot be used in the context of communicating material, because it needs that all the nodes can directly communicate with the BS. In addition, different physical phenomena will be monitored in this phase, such as temperature, humidity, etc. A data-centric tree-based routing protocol like **TAG will also be considered. Besides, PEDAP can also be a good choice because of its low maintaining cost.**

Table II. PERFORMANCE METRICS INDICATOR OF DATA AGGREGATION PROTOCOLS IN WSN

Protocol name (Type)	Manufacturing	Construction	Exploitation	Mean
PEGASIS (Ch)	-0.75	-0.80	-0.65	-0.73
CCM (Ch)	-0.25	-0.40	-0.15	-0.27
CRBCC (Ch)	0.20	0.10	<b>0.40</b>	0.23
BCBRP (Ch)	-0.60	-0.65	-0.20	-0.48
LEACH (Cl)	-0.55	-0.55	-0.25	-0.45
TL-LEACH (Cl)	-0.20	-0.20	0.00	-0.13
HEED (Cl)	0.30	0.30	0.30	0.30
DWEHC (Cl)	<b>0.55</b>	<b>0.50</b>	<b>0.65</b>	<b>0.57</b>
EADAT (Tr)	-0.30	-0.20	-0.50	-0.33
PEDAP (Tr)	<b>0.60</b>	<b>0.55</b>	<b>0.40</b>	<b>0.52</b>
TAG (Tr)	<b>0.50</b>	<b>0.50</b>	<b>0.30</b>	0.43

From the last column, we find that the **DWEHC and PEDAP have better performance for the whole lifecycle of the communicating concrete**. Using one of these two protocols for our application will give a good result. For optimizing the service time, we can also test PEDAP for manufacturing and construction, and DWEHC for exploitation. Even the performance indicator of **TAG** is lower, it can also be interesting to test its data-centric capacity for efficiently monitoring data like temperature or humidity.

## V. CONCLUSION

In this paper, the concept of communicating material is declined. It is applied to design communicating concrete in the framework of the McBIM project. The required performance metrics for data collection, from the manufacturing of the concrete elements until its exploitation in a building, are also detailed. Three main data aggregation architectures are presented with some typical routing protocols, as well as its advantages and drawbacks. Finally, we compare these algorithms with the required performance metrics for our context. Briefly, the chain-based approach is easy to implement and very adapted for a small network, but it may cause a long delay. Moreover, chain-based protocols imply that all nodes can directly communicate with the BS. In the context of communicating material, we cannot ensure that this hypothesis is always true. The strong scalability of tree-based protocol makes the auto-organization of communicating concretes possible for the construction phase. However, low resilience to the link failure, and high maintaining cost are unavoidable drawbacks. Cluster-based approach provides a robust network and low delay for a large network, it may be a good choice for the exploitation phase.

We focus in this paper on the structure of the network to optimize the energy consumed by a concrete element. There is another interest in determining this structure. Indeed, the physicists who will design this type of communicating material will be able to orient their development according to this optimal structure. This work is the first step toward the final choice of one or more protocols. Today we are working on both real tests and simulation software. But it is by coupling the WSN with an external control that we will be able to guarantee the flexibility of the communication structure.

The future experiments will aim to test and analyze the energy consumption of the data aggregation algorithms that would have been previously selected. These results will confirm if the aggregation approach could significantly improve the lifetime of the WSN for the McBIM project. It will also be possible to decide if one technique could be used during all the lifecycle phases of McBIM elements or if a combination of different techniques should be used regarding the lifecycle phase of an element.

## ACKNOWLEDGMENT

Authors thank financial support from the French National Research Agency (ANR) under the McBIM (Material communicating with the BIM) project, grant number ANR-17-CE10-0014.

## REFERENCES

- [1] A. Thomas, "RFID et nouvelles technologies de communication; enjeux économiques incontournables et problèmes d'éthique," 6ème Conférence Int. Concept. Prod. Intégrées, 2009.
- [2] S. Kubler, K. Främling, W. Derigent, and M. J. Rabita, "Premiers travaux relatifs au concept de matière communicante : Processus de dissémination des informations relatives au produit," Thèse de doctorat, Université de Lorraine, 2012.
- [3] K. Mekki, "Gestion de l'information embarquée dans des matériaux communicants à l'aide de protocoles de réseaux de capteurs," Thèse de doctorat, Université de Lorraine, 2016.
- [4] "ANR Project McBIM," 2017. [Online]. Available: <http://mcbim.cran.univ-lorraine.fr/Members>.
- [5] C. D. Ortiz, J. M. Puig, C. E. Palau, and M. Esteve, "3D wireless sensor network modeling and simulation," 2007 Int. Conf. Sens. Technol. Appl., pp. 307–312, 2007.
- [6] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "in-Network Aggregation Techniques for Wireless Sensor Networks: a Survey," IEEE Wireless Communications, vol. 14, no. 2, pp. 70–87, April 2007.
- [7] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," IEEE Trans. Parallel Distrib. Syst., vol. 13, no. 9, pp. 924–935, 2002.
- [8] F. Tang, I. You, S. Guo, M. Guo, and Y. Ma, "A chain-cluster based routing algorithm for wireless sensor networks," J. Intell. Manuf., vol. 23, no. 4, pp. 1305–1313, 2012.
- [9] G. Zheng and Z. Hu, "Chain Routing Based on Coordinates-Oriented Clustering Strategy in WSNs," Int. Symp. Comput. Netw. Multimed. Technol., vol. 2, no. 1, pp. 1–4, 2010.
- [10] K. S. Ahn, D. G. Kim, B. S. Sim, H. Y. Youn, and O. Song, "Balanced chain-based routing protocol (BCBRP) for energy efficient wireless sensor networks," IEEE Int. Symp. Parallel Distrib. Process. with Appl. Work., pp. 227–231, 2011.
- [11] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," 33rd Annu. Hawaii Int. Conf. Syst. Sci., vol. 8, p. 8020, 2000.
- [12] V. Loscri, G. Morabito, and S. Marano, "A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH)," IEEE Veh. Technol. Conf., vol. 3, pp. 1809–1813, 2006.
- [13] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," IEEE Trans. Mob. Comput., vol. 3, no. 4, pp. 366–379, 2004.
- [14] P. Ding, J. Holliday, and A. Celik, "Distributed Energy-Efficient Hierarchical Clustering for," IEEE international conference on Distributed Computing in Sensor Systems, pp. 322–339, 2005.
- [15] M. Ding, X. Cheng, and G. Xue, "Aggregation tree construction in sensor networks," IEEE 58th VTC, p. 2168–2172 Vol.4, 2004.
- [16] H. Ö. Tan and I. Körpeoğlu, "Power efficient data gathering and aggregation in wireless sensor networks," ACM SIGMOD Rec., vol. 32, no. 4, p. 66, 2005.
- [17] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: A Tiny AGgregation Service for Ad-hoc Sensor Networks," Osdi, vol. 36, no. SI, pp. 131–146, 2002.