Fault-Tolerant Flooding Through Formal Concept Analysis for Wireless Sensor Networks

Tarek Moulahi[1,2,3], Salem Nasri[2] and Hervé Guyennet[3]
[1]CBE, Qassim University, Al-Qassim, KSA
[2]ENIS, University of Sfax, Sfax, Tunisia
[3]University of Besançon, Besançon, France
tarek.moulahi@lifc.univ-fcomte.fr
salem.nasri@enim.rnu.tn
herve.guyennet@lifc.univ-fcomte.fr

Abstract— **Flooding and broadcasting are basic and fundamental operations in Wireless Sensor Networks (WSNs). In fact, they are used for data dissemination, time synchronization, key distribution, node localization, and routing. On the other hand, sensors are prone to failure such as exhaustion of energy, environmental hazards, and software or hardware malfunctioning. In this paper, we make an enhancement of the latest proposed relay-based broadcasting method by considering a trust index model to achieve a fault tolerant relay node selection. In addition, Formal Concept Analysis is used instead of the linear function. The effectiveness of the proposed scheme is confirmed through a simulation study using NS-2.**

Keywords-component; Fault tolerance, Flooding, Formal Concept Analysis, WSNs.

I. Introduction

Wireless Sensor Networks (WSNs) are sets of many sensor nodes managed by a base station also called sink node. Each sensor is an energy-constrained small device able to sense, compute and communicate via wireless channel. The ability to collect, to compute and to communicate data of these devices compared with its low cost has attracted the attention of the researchers' community. WSNs are therefore used in several domains such us military, health care, environment and others [1].

Nodes in WSNs are deployed without any predefined topologies. Consequently, the network is formed randomly using wireless channels. A sensor node is constrained by its limited memory size, low computing capacity, and weak energy, so that many methods, algorithms and protocols were introduced and developed taking into consideration this constraint.

On the other hand, broadcasting is a fundamental and operation in such. In fact, broadcasting and flooding can be used for:

- Data dissemination [2].
- Time synchronization [3].
- Key distribution [4].
- Node localization [5].
- Routing [6]

To perform the broadcasting task, many methods and protocols have been proposed: using relay approach or using a special set as a backbone like trees or CDS. In this paper relay based broadcasting is chosen to be enhanced. In fact, relay based method are based on the use of partial 2-hop information, each node chooses a small set of forward neighbors to relay messages and this set covers the nodes 2-hop neighbor set.

The latest relay-based broadcasting method RDS-MPR [7] selects relay nodes according to their remaining energy, degree, and probability of reception. Unfortunately, RDS-MPR does not take into consideration node failure. Furthermore, this method uses a linear function to compute node weight. We show next in this paper the imperfection of such a choice.

The aim of this research paper is to enhance RDS-MPR by considering a trust index model in order to make a fault tolerant selection of relay node. This fact increases notably the success ratio of RDS-MPR. Moreover, Formal Concept Analysis (FCA) is used as a technique to select relay nodes instead of the linear function. In fact, FCA provides a perfect organization of nodes according to their characteristics (remaining energy, degree, probability of reception, and trust index). This sort helps to make a quick selection of relay nodes and provide an effective performance of the proposed broadcasting method.

The remainder of this paper is organized as follows. In the next section, there is a study of the related work. Section III makes a review of RDS-MPR method and FCA. In section IV, the contribution of this work is presented. Section V contains a measurement and experimental study to illustrate our improvement. Finally, section VI recapitulates this research work.

II. Related Work

Relay broadcasting methods are based on the use of partial 2-hop information, each node chooses a small set of forward neighbors to relay messages and this set covers the nodes 2-hop neighbor set. This kind of method was proposed in [2] by the Multipoint Relay (MPR) method. In MPR, the broadcasting is performed by levels aiming to minimize the redundant packets and the total cost of broadcasting. Since its proposition MPR was the topic of many research papers.

In [8] MPR was used, for the first time, to compute the Connected Dominating Set (CDS) used in both routing and broadcasting. Then, a proposition of GMPR (Getaway MPR) aiming to improve the broadcasting by combining MPR and the maximal independent set (MIS) concepts was proposed to compute a connected dominating set (CDS) in the network [9].

In [10], authors provide an extension of MPR to compute connected dominating sets, by proposing several extensions the authors' aim was to select a smaller CDS using complete 2-hop information to cover each node's 2-hop neighbor set.

In [11] authors show the disadvantage of MPR in a realistic environment. So, they propose a modification of this method to improve its reachability with a realistic physical layer.

An improvement of MPR was also proposed in [12] aiming to decrease the total amount of energy to accomplish the broadcasting task in WSNs using a cluster method for avoiding redundant receptions.

In [7] there is a proposition of a new relay-based broadcasting method by combining the computation of CDS with MPR and the use of a realistic physical layer. This proposition has added a realistic behavior to the method firstly proposed in [8]. This proposition has led to improve the original heuristic in term of accessibility i.e. success ratio of nodes which have received correctly a broadcasted message from one node in networks. The proposed method was called RDS-MPR.

To our knowledge, previous relay-based flooding methods have not take into consideration the node failure in spite of the importance of such a criterion.

In this research work a new fault tolerant scheme is proposed where FCA is used as a weighted function instead of linear summation.

III. Preliminaries

In this section, we first describe our network model. Next, a review of RDS-MPR scheme is performed in addition to a brief presentation of FCA method.

A. *Network model*

In this paper we use sensors with the same coverage area; the probability of reception between sensors is performed according to the lognormal shadowing model [7]. WSNs can be considered as a graph *G(V;E),* where *V* is the set of vertex representing nodes and *E* is the set of edges between vertexes. An edge exists between two nodes *A* and *B* in *G* only if *A* is in the coverage area of *B* and vice-versa.

B. *RDS-MPR review*

RDS-MPR [7] is a broadcasting technique applied to non clustered WSNs. This method is based on the use of partial 2-hop information, each node chooses a small set of forward neighbors to relay messages and this set covers the nodes 2-hop neighbor set. The choice of relay node is performed according to a weight function, which depends on node remaining energy, node degree and probability of reception between nodes according to the lognormal model. The selected set consists of nodes having the greatest weight and should provide good accessibility. In this method, the weight function is called *RW*: Realistic Weight (1),(2) which is computed according to the node remaining energy, the node degree and the reception probability according to the lognormal shadowing model.

$$RW(u)=a\times RW_{degree}(u)+b\times RW_{energy}(u)+c\times RW_{probability}(u) \quad (1)$$

Where:

$$a+b+c=1 \quad (2)$$

a, b, and c are priority factors. In Fig. 1, *n1* and *n3* are chosen to relay packets sent from *u* in order to complete the broadcasting. In blind flooding *n1, n2*, and *n3* will relay information sent from *u*.

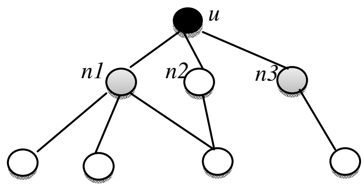

Figure 1. Selection of relaying nodes using RDS-MPR

C. *Formal Concept Analysis*

FCA is a data analysis tool applied in many different domains like psychology, sociology, anthropology, medicine, biology, linguistics, computer sciences, mathematics and industrial engineering [13]. The aim of FCA is to find automatically groups of objects or entities that have a group of attributes in common.

- Definition 1: Context

A context is a specific type of data, which consists of objects (O) described by several attributes (A), represented as an n*m table with objects as rows and attributes as columns. If an object (in row i) has the attribute (in column j), this will be indicated by a cross in cell(i,j) of the table. Otherwise, the cell will be empty. The data in the table can also be represented by Boolean values (1 if there is a cross and 0 otherwise).

- Definition 2: Formal concept

Let A be a set of attributes, O be a set of objects where A and O describe a context C.

(*a*,*o*) is a formal concept of C, if and only if:

(1) $a \subset$ A,

(2) $o \subset$ O and

(3) the set of attributes *a* has in common the set of objects *o*.

- Definition 3: FCA

FCA is a technique applied to the previously described context, and performed in two steps: (1) extract all formal concepts from a context, (2) find groups of objects that share groups of attributes.

- Definition 4: Concept lattice

The classification of the entire concept forms a concept lattice, i.e. it is the result of the application of FCA on a context.

IV. Paper Contribution

In this section, the contribution of this research work is presented

A. Trust index model

In [14], there is a presentation of a trust index model, which can be used as a factor of distinguishing correct nodes from fault nodes. Each node has a trust index (*TI*) between 0 and 1; initially, for all nodes the *TI* is equal to 1. A node is more reliable if its trust index is higher. This weight of a node *u* can be calculated as:

$$TI(u) = e^{-\lambda v(u)} \qquad (3)$$

Where:

λ : is a modifiable variable to recalculate $TI(u)$, and it is used to decide on the fastness of changing TI(u) when v(u) increases or decreases.

$v(u)$: initially equal to 0, because initially the *TI* is equal to 1 for all nodes. $v(u)$ is incremented by the sink node by a step of 0.1 in case of receiving a report from *u* estimated as fault. Otherwise (in case of receiving a report from *u* estimated as correct), $v(u)$ is decremented by a step of 0.1. So that v(u) can be described by:

$$\begin{cases} v_0(u)=0 \\ v(u)= \begin{cases} v(u)+0.1 & \text{if } u \text{ is estimated fault} \\ Max(v(u)-0.1;\,0) & \text{if } u \text{ is estimated correct} \end{cases} \end{cases}$$

B. Relaying context

RDS-MPR method selects relay node according to their energy weight, probability of reception weight and degree weight. In our contribution we add a new weight which is fault tolerant weight defined previously as a trust index.

An example of WSNs is given in fig.2 where we propose to select best relaying nodes.

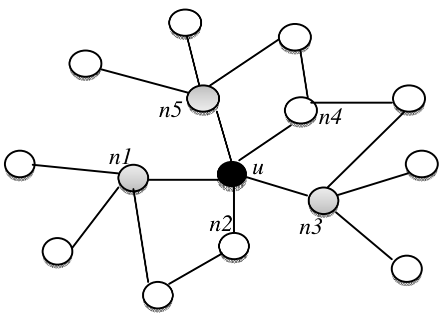

Figure 2. Selection of relaying nodes

To define the relaying context we use nodes as objects and weights as attributes. We define a threshold τ for each weight w_i of a node n_i then if $w_i(n_i) > \tau$, consequently the node will be considered as having this weight i.e. we make a cross in cell(n_i,w_i) of the relaying context table. Table1 describes the relaying context of the networks given in fig.2.

Table I. Relaying Context

node	Energy Weight (EW)	Degree Weight (DW)	Probability Weight (PW)	Fault tolerance Weight (FW)
n1	x	x	x	x
n2			x	x
n3	x	x	x	x
n4	x		x	
n5	x	x		x

The previous table describes characteristics of the node *u* neighbor's:

- The cell (*n1*,EW)=x, means that the remaining energy in the node *n1* is greater than the predefined energy threshold.
- The cell (*n2*,EW)=' ', means that the remaining energy in the node *n2* is less than the predefined energy threshold.
- The cell (*n1*,FW)=x, means that the trust index of the node *n1* is greater than the predefined trust index threshold.
- The cell (*n1*,PW)=x, means that the probability of a correct reception from *u* to *n1* is greater than the predefined probability threshold.
- The cell (*n3*,DW)=x, means that the number of *n3* neighbor's is greater than the predefined number of neighbor's threshold.
- The cell (*n4*,FW)=' ', means that the trust index of the node *n4* is less than the predefined trust index threshold.

And similarly for other cells in the table.

C. Relaying lattice

FCA provides a classification of objects (nodes) according to their weights (attributes). Many tools can be used to generate concept lattice like in [15]. Fig. 3 describes the relaying lattice where bottom level gives node having the maximum common attributes, and top level maximize objects.

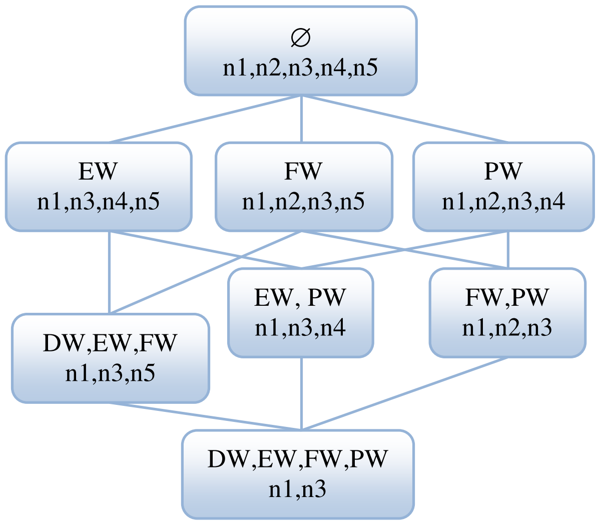

Figure 3. Relaying lattice

D. *Algorithm*

The proposed algorithm uses the previously described relaying context and relaying lattice. The choice of using FCA instead of linear function is due to: the summing of many weights can provide wrong choice because weak weights can be hidden by higher weights. On the contrary FCA requires a specific threshold in each weight.

Algorithm: Fault-tolerant Flooding through FCA

Input: a node *u* and 2-hop nodes information from *u* in a connected graph *G(V,E)*
Output: the set of relaying nodes *RN*(u)
1: put 1-hop nodes from *u* in a set called *N1(u)*
2: put 2-hop nodes from *u* in a set called *N2(u)*
3: remove all nodes in *N2(u)* having single parent in *N1(u)*, add their parent to *RN*(u), and remove them from *N1(u)*.
4: Generate the relaying context of node in *N1(u)*
5: Generate the relaying lattice of previous relaying context
6: **while** *N2(u)* is not empty
7: start by node in the bottom level of the relaying lattice, add it to *RN*(u) and remove its child from *N2(u)*
8: if *N2(u)* still not empty
Go up in the relaying lattice
9: **End while**

V. EXPERIMENTAL STUDY

The test and the measurement of the performance are realized using NS-2 [16], and according to table 2.

TABLE II. SIMULATION PARAMETERS

Parameter	Value
Size of filed	500/500m
Number of nodes	50:300 randomly deployed in the filed
Radius of transmission	50m
Routing protocol	AODV
Node intial energy	10kJ
Tx power	2.0 mW
Rx power	1.0 mW
Idle power	0.2 mW
Sleeping power	0.02 mW
Transition power from sleeping to idle	2.0 mW
Transition time from sleeping to idle	0.05 second
Weights thresholds	
Energy threshold	5.0 mW
Degree threshold	2
Probability threshold	0.5
Trust index threshold	0.5

In the simulation we execute both our proposed algorithm and RDS-MPR five times. Furthermore, following assumptions are used:

- *TI(u)* is a probability between 0 and 1 which is affected randomly to nodes in the filed. In fact, we assume that these values are updated when gathering data by the sink node according to the node behavior, and each node has the trust index of its 1-hop neighbors.
- The probability of reception between nodes is computed according to lognormal shadowing model.
- Nodes are deployed randomly and forming a connected graph.

In fig. 4, the number of nodes in the network is fixed at 150, and a variation of failure node ratio from 10% to 50% is performed i.e. node with low *TI* will be considered as failed. Then, we calculate the success ratio of each method which is the total number of nodes having received data sent from sink node. In our proposed algorithm node with low *TI* will be avoided to be chosen as relay node. Consequently, the total success ratio will be greater than RDS-MPR success ratio. According to fig.4, our proposed method enhances the success ratio by an average of 36.7% in the best cases and no less than RDS-MPR success ratio in the worst cases .

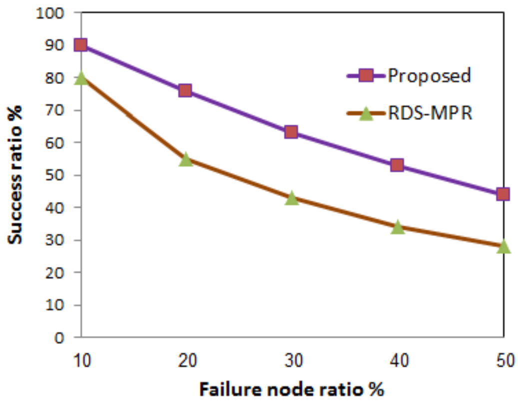

Figure 4. Comparison of success ratio for a fixed number of nodes (150) and different failure node ratio

In fig. 5, the node failure rate is fixed at 20%, and we make a variation of number of nodes in the network from 50 to 300.

We can see that the previous result is confirmed and the improvement of our proposed method can be shown. The rate of improvement is 50% in the best cases and no less than RDS-MPR success ratio in the worst cases.

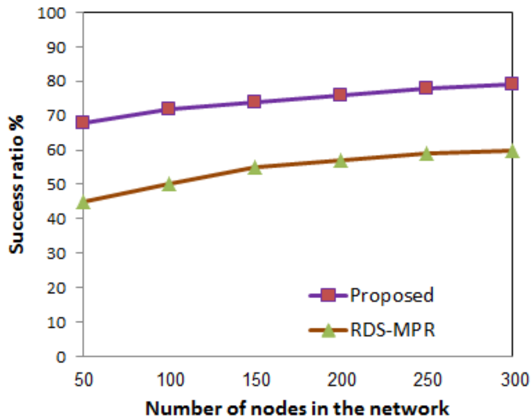

Figure 5. Comparison of success ratio for a fixed failure node ratio (20%) and different numbers of node network

The consumed energy to select relay nodes using RDS-MPR or using our proposed scheme is the same because the same number of requests is realized to gather data from nodes. The collected data are:

- Node position in order to compute probability of reception according to the lognormal shadowing model
- Node remaining energy

The node degree is computed according to the number of received message as a response to gathering data requests.

We consider only number of message because the consumed energy for internal processing can be neglected compared to the consumed energy for communication.

VI. CONCLUSION

This work defines a new relay-based flooding method aiming to enhance an existing technique by considering a node trust index. The use of trust index induces a fault-tolerant behavior to the proposed method. This aspect is demonstrated by the improvement of our scheme success ratio, in the best cases, by an average of 36.7% compared to RDS-MPR and using the same amount of energy. In addition, the limit of using a linear weighted function is shown. Furthermore, FCA is used instead. In fact, FCA is a powerful technique which is widely used in several scientific domains. To our knowledge, it is the first time FCA is used to enrich flooding methods in WSNs.

REFERENCES

[1] Jennifer Yick, Biswanath Mukherjee and Dipak Ghosal. "Wireless sensor network survey", Computer Networks 52, pp.2292-2330, 2008.

[2] Qayyum, A.; Viennot, L.; Laouiti, A., "Multipoint relaying for flooding broadcast messages in mobile wireless networks," System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on , vol., no., pp.3866,3875, 7-10 Jan. 2002.

[3] C. Lenzen, P. Sommer, and R. Wattenhofer. "Optimal clock synchronization in networks". In Proceedings of the 7th ACM Conference in Embedded Networked Sensor Systems, pages 225–238. ACM, 2009

[4] Chen, C.-Y. and Chao, H.-C. (2011), A survey of key distribution in wireless sensor networks. Security Comm. Networks.

[5] Sheng Wang; Zude Zhou, "An Improved Nodes Localization Scheme with Energy Saving Flooding Base on MAS in WSN", 4th International Conference on Wireless Communications, Networking and Mobile computing2008. WiCOM '08,pp.1,4, 12-14 Oct. 2008.

[6] Miklos Maroti, "Directed flood-routing framework for wireless sensor networks", Proceedings of the 5th ACM/IFIP/USENIX international conference on Middleware, pp. 99-114. 2004.

[7] Tarek Moulahi, Salem Nasri, Hervé Guyennet, "Broadcasting Based on Dominated Connecting Sets with MPR in a Realistic Environment for WSNs & AD-HOC", Journal of Network and Computer Applications, Elsevier, November 2012. pp 1720-1727.

[8] C. Adjih, P. Jaquet, and L. Viennot. "Computing Connected Dominated Sets with Multi-point Relays", International Journal Ad Hoc & Sensor Wireless Networks, vol.1no.1-2, pp.27-39, 2005.

[9] Ou Liang, Y. Ahmet Sekercioglu and Nallasamy Mani." Gateway multipoint relays an-MPR-based broadcast algorithm for ad hoc networks", IEEE Singapore International Conference on Communication systems,2007.

[10] Jie Wu, Wei Lou, and Fei Dai. "Extended Multipoint Relays to DetermineConnected Dominating Sets in MANETs", IEEE transactions on computers, 2006, vol. 55 no. 3.

[11] Francois Ingelrest and David Simplot-Ryl. "Maximizing the Probability of Delivery of Multipoint Relay Broadcast Protocol in Wireless Ad Hoc Networks with a Realistic Physical Layer", Mobile Ad hoc and Sensor Networks, Lecture Notes in Computer Science (Springer), 2006,pp.143-154.

[12] Tarek Moulahi, Salem Nasri and Hervé Guyennet. "Amalioration of MPR by a Backbone-based Brodcasting Algorithm for WSNs", International Conference on Information Technology and e-services (ICITeS), pp.506-511, 2012.

[13] Wolff, Karl Erich, "A first course in Formal Concept Analysis", in F.Faulbaum, StatSoft '93, Gustav Fischer Verlag, pp. 429-438, 1994.

[14] Xianghua XU1, Chunxiong FAN, Jian WAN, Wei ZHANG, Naixue XIONG,"TIM-SNAP: Fault Tolerant Event Location Estimation in Sensor Network",Journal of Computational Information Systems 8: 8 (2012) 3507–3517.

[15] Galois lattice interactive constructor, http://www.iro.umontreal.ca/~galicia/goals.html

[16] The Network Simulator - ns-2, http://www.isi.edu/nsnam/ns/.