Improvement on K-means and TSP Based Mobility Protocol of Wireless Sensor Network

Afef Ghabri[1], Monia Bellalouna[2], Walid Khaznaji[3]
[1]Laboratory RIADI-GDL, [23]Laboratory CRISTAL POLE GRIFT
National School of Computer Sciences (ENSI)
Manouba, Tunisia
e-mail: [1]afef2108@hotmail.com, [2]monia.bellalouna@ensi.rnu.tn, [3]mwkhaznaji@yahoo.fr

***Abstract*—Wireless sensor networks are widely used in different environments in order to execute diverse tasks and applications. In this paper, we will consider problems related to the fault tolerance issue and we will present a new version of K-means And Traveling Salesman Problem based mobility protocol which aims to provide not only better energy efficiency within the wireless network, but also better reliability compared with the conventional method based on K-means clustering and the approximate solution for Traveling Salesman Problem by using the simple local search algorithm "2-Opt". This problem is Non-deterministic Polynomial-time hard (NP-hard), so we propose the new approach that navigates the mobile sink to go through the cluster centers according to the optimized route by implementing the local search method "Tabu". Simulation results have demonstrated that the solution given by the Tabu heuristic outperforms the original solution of K-means and Traveling Salesman Problem based mobility protocol in terms of quality. Our goal is to propose a much more realistic model that provides less execution time than the conventional strategy and an effective improvement once the problem is disturbed by the breakdowns of some nodes.**

Keywords-Wireless sensor networks; fault-tolerance; failures; Tabu; quality; realistic.

I. INTRODUCTION

The technological advances carried out in the past years have allowed the development of new varieties of sensors, which can be configured and used for wireless communication in order to form autonomous networks. Indeed, the sensors can send data to all the nodes which are connected to the network. They are enormously employed in various environments in a random way to perform different monitoring tasks such as the delivery, the searching, the help in case of disasters and the target tracking [1]. Generally, a sensor network is primarily composed of several collector nodes that are scattered throughout a sensing field, a data processing center and sinks [2]. Each node, randomly distributed, is responsible for collecting obtained data from its coverage area. The sinks are particular nodes that must always be active and whose role is to route recovered data coming from sensor nodes. Their number depends on the network load and size. These collection points represent the interface between the sensing field and the data processing center that facilitates data recovery and ensures their treatment in order to extract for the user useful information [3]. Figure 1 shows this principle.

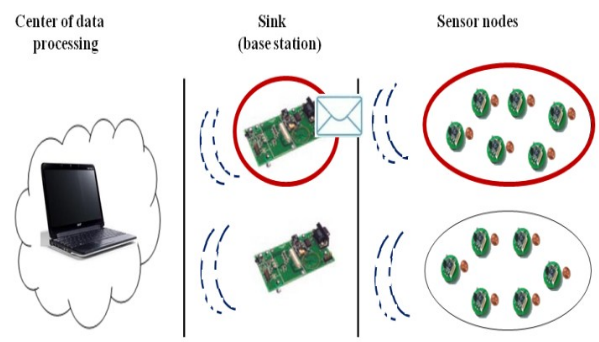

Figure 1. Wireless sensor network architecture

Generally, in the case of a large-scale network, data is periodically transmitted from source nodes to the sink up the tree- structure. Data routing is ensured by intermediate nodes that forward the messages until reaching the destination. Then, there exist different possible paths in order to derive information [4]. This process imposes energy conservation and network lifetime prolongation problems. It is thus necessary to find the shortest and optimal route in order to provide the data routing, within the energy constraints and the time of transferring packages [5]. In dealing with this problem, we noticed that it leads to a Combinatorial Optimization Problem (COP) [6]. In other words, several problems to be solved in the Wireless Sensor Network (WSN), such as the routing problem which returns to find a shorter path, or the minimization problem of energy, consumed by the sensors, taking into account several constraints as their displacement and their communications, can be reduced to optimization problems [7]. Unfortunately, the sensors are prone to many intrusions and failures, mainly due to physical destructions, depletion of batteries, communication link errors, malicious attacks and environmental interferences [8]. Thus, the sensor breakdown causes the loss of communication links that leads to a significant change in the overall network topology. This can affect the network connectivity and decrease its life [7]. Failures of sensor nodes should not affect the overall network performance. This is a problem of reliability or fault tolerance. Fault tolerance is the ability of a network to keep its functionality without causing interruptions if a sensor node stops working. We can say that its objective is to avoid the total system flaw despite the existence of errors in a subset of its elementary components [9]. The studied problem is then reduced to a COP, where the number of data is a random variable instead of a deterministic value, and thereafter it was more interesting to consider it as being a Probabilistic Combinatorial Optimization Problem (PCOP) [10].

The purpose of this paper is to propose a new strategy that guarantees the obtaining of a solution in real time once the problem is disturbed by the failures of some nodes, by considering the fault tolerance as a PCOP through the K-means and Traveling Salesman Problem-based mobility (KAT-mobility) protocol [11]. The rest of the paper will be organized as follows: In Section II, we will present the PCOP and particularly the Probabilistic Traveling Salesman Problem (PTSP). Section 3 will describe its solving methods. A description of the KAT-mobility protocol principle will be shown in Section 4. In Section 5, we will propose our improved strategy. Section 6 presents the simulation results and analysis. Finally, we will conclude the paper in Section 7.

II. Probabilistic Combinatorial Optimization Problems

In recent years, COPs are considered as a critical issue and a very interesting subject of research. The study of these problems became one of the most exciting and active areas in the field of discrete mathematics [12]. In fact, a COP is defined as follows: given n data, a set of configurations S and an objective function defined on S in $\Re$, the goal is to find a configuration s* that minimizes the function f.

$$\begin{aligned} f &: S \to \Re \\ s &\mapsto f(s) \end{aligned} \tag{1}$$

This deterministic model is inadequate with reality, where the number of data of the studied problem is frequently a random variable between 0 and n. Recently, a specific family of COPs, characterized by the fact that probabilistic elements are included explicitly in the problem definitions, has appeared and has been investigated [13]. For this reason, they were named PCOPs. There are many motivations for studying the effect of probabilistic elements inclusion in COPs: among them two are the most significant. The first one is the desire of formulating and analyzing models which are more convenient for real world problems, where randomness exists [12]. The second motivation is the possibility to analyze the robustness of optimal solutions for deterministic problems, considering the modification of the instances for which these problems are solved: disruption by the absence of certain data [14]. The PCOP is defined as follows: given a random number of data varying between 0 and n (0 when all the tasks are absent and n when all the tasks are present), a set of configurations S, an objective function defined on S in $\Re$ and a modification strategy which adapts the feasible solution through S to the new subset in order to obtain the realizable solution through the set of present tasks, the objective is to find a configuration s* which minimizes the functional of f according to this strategy.

$$E[f(s^*)] = \min(E[f(s)]) \tag{2}$$

The first studied problem in PCOPs, initially introduced by Jaillet [15], was the PTSP [16]. This approach has been then extended to other different problems and studies have continued in several domains, such as the Probabilistic Traveling Salesman Facility Location Problem (PTSFLP) [13], the probabilistic longest path problem [17] and the probabilistic minimum vertex covering problem [18]. The probabilistic approach was extended to combinatorial problems that are not defined on graphs, such as the Probabilistic Scheduling Problem (PSP) [19] and the Probabilistic Bin Packing Problem (PBPP) [20]. There are many interesting and important applications of PCOPs, particularly in the context of communication systems, strategic planning, job scheduling, etc. [21]. To deal with these problems, two strategies can be used: the re-optimization strategy [22] and the a priori strategy [13].

A. *Solving Strategies*

Frequently in applications, after having solved a particular instance of a given combinatorial optimization problem, we must solve repeatedly many copies of the same problem. These additional instances are generally simple variations of the original problem; however, they are sufficiently different to require an individual treatment [18]. The most natural approach used to address this kind of situation consists in solving in an optimal way the different potential copies. We call this strategy the "re-optimization strategy" [23]. However, it has many disadvantages and the most important one is the high cost. For example, if the considered COP is NP-hard [24], one could have to solve an exponential number of instances of a very hard problem. Moreover, in several applications it is required to find a solution to each new instance promptly, but one could not have the necessary computing or other resources to carry out such a task [13]. It is therefore necessary to adopt a different strategy. Rather than re-optimizing each successive exemplary, we can try to determine a priori solution of the initial problem that can be successively modified in a simple way to solve the following instances. We call this strategy, introduced by Jaillet [15][10] and which is less costly in terms of computations, the "a priori strategy" [6].

Different algorithms were implemented for PCOP resolution and have shown a satisfactory performance. Approximate methods have been used to solve large scale PCOPs. Among them, we mention the work of Bertsimas [25], in which he proposed and analyzed heuristics for PVRP. Many algorithms based on classic heuristics were also proposed for the PBPP, in the works of Bellalouna [6][22]. Metaheuristics were also employed to solve PCOPs, such as simulated annealing and Tabu search, implemented in the case of PTSP [22]. A Tabu Search was implemented by Gendreau et al. [26] for solving the vehicle routing problem with stochastic demands and customers [14].

 ISBN: 978-1-61208-421-3

Furthermore, exact methods have been used, such as branch & bound algorithm developed by Rosenow [27], to solve PCOPs and especially the probabilistic traveling salesman problem.

B. Probabilistic Traveling Salesman Problem

The PTSP is a variation of the standard Traveling Salesman Problem (TSP) [28]. It is essentially a TSP, in which the number of nodes that require being visited in each problem is a random variable [13]. In fact, for a given finite set of points, the TSP consists of finding a tour through the points of minimum total length [12], and in the probabilistic version of this problem, introduced for the first time by Jaillet [15]; only a subset of the nodes may be present in any given instance of it. The TSP can then be considered and treated as a special case of PTSP. Indeed, the principal difference between the two problems is that in TSP the probability of each visited node is 1, while in PTSP the probability of each visited node is between 0 and 1 [14].

In other words, consider a routing problem through a set of n known points. On any given instance of this problem, only a random subset S has to be visited. In many cases, we can not have resources to reoptimize the tour for every instance, and even if we had them, re-optimization may become too time consuming [13]. We wish to find a priori tour through all the points. On any given instance, the subset of present points will then be visited in the same order as they appear in the a priori tour, i.e., we simply skip the absent points in that problem instance [21]. The problem of finding such a priori tour of minimum expected length according to this skipping strategy is defined as a PTSP [12]. Figure 2 shows this principle.

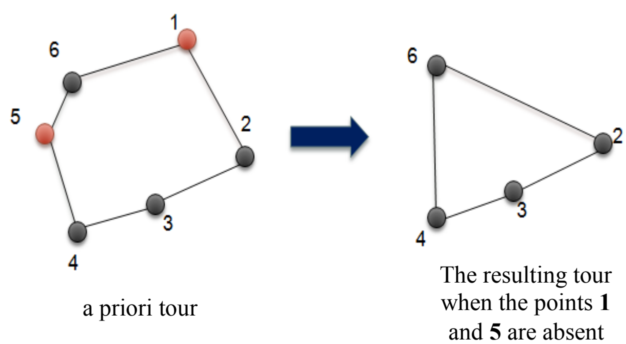

Figure 2. PTSP methodology

What distinguishes the PTSP from other problems is the probability distribution (law) that specifies the number and the identity of present points that require to be visited on any given problem instance. The PTSP can be formulated as follows: Consider a complete graph G = (V, E) on n nodes and a priori tour R through these points. If every possible subset of the node set V = {1, 2…, n} may or may not be present on any given instance of the optimization problem, for example, on any given day, the traveling salesman may have to visit only a subset S (S ⊆ V), then there are 2^n possible instances of the problem, i.e., 2^n is the number of all the possible subsets of V [29]. Let P(S) be the probability that instance S occurs. Given a method μ for updating a priori solution R to this optimization problem on G, μ will then produce for S, a feasible solution with the value (cost) $L_R(S)$. In fact, this solution is a tour through the subset S of nodes and $L_R(S)$ represents the length of that tour [15][29]. Then, given that we have already selected the updating strategy μ, the natural choice for the a priori solution is to select a tour through all potential nodes that minimizes the expected cost with the summation over all subsets of V [29].

$$\mathbb{E}\left[L_{(R,\mu)}\right]=\sum_{S\subset V} P(S)L_{(R,\mu)}(S) \quad (3)$$

Otherwise, each point i has a probability of presence p_i, independently of the others. We assume that d (i, j) is the distance between points i and j and we suppose that the a priori tour is R = (1, 2….., n-1, n, 1), then our problem is finding a tour through n nodes, which minimizes the expected length of a determined a priori PTSP tour R, denoted $\mathbb{E}\,[L_{(R,\mu)}]$ [14].

$$\mathbb{E}\left[L_{(R,\mu)}\right]=\sum_{i=1}^{n}\sum_{j=i+1}^{n} d_{ij}p_i p_j \prod_{k=i+1}^{j-1}(1-p_k)+\sum_{j=1}^{n}\sum_{i=1}^{j-1} d_{ji}p_i p_j \prod_{k=j+1}^{n}\prod_{k=1}^{i-1}(1-p_k) \quad (4)$$

If all the points have the same probability of presence ($p_i = p$, ∀ i), the expected length $\mathbb{E}\,[L_{(R,\mu)}]$ can be expressed using the following formula [6]:

$$\mathbb{E}\left[L_{(R,\mu)}\right]=p^2\sum_{\tau}^{n-2}(1-p)^{\tau} L_{(R,\mu)^{\tau}} \quad (5)$$

Where

$$L_{(R,\mu)^{\tau}}=\sum_{i=1}^{n} d(i,R^{\tau}(i)) \quad (6)$$

R^{τ} comprises pgcd (n, τ+1) sub-tours. It consists in jumping τ points from the initial tour R. $R^{\tau}(i)$ represents the point after i along the permutation R^{τ}, so $L_{(R,\mu)}{}^{\tau}$ is the permutation length. We note that R^0 is the tour R and $L_{(R,\mu)}{}^{0}$ is the length of the tour R [14][15].

There are several methods for solving PTSPs. A wide variety of approximate and exact algorithms have been proposed for solving it. Exact methods can only solve relatively small problems. Concerning the approximate algorithms, a number of heuristics and metaheuristics have been effectively used in the case of large problems. Nowadays, approximate methods are considered as a very interesting subject for many researchers who still trying to find the best algorithm that gives a very good approximate solution in appropriate running time [14].

III. SOLVING METHODS OF PTSP

The PTSP was introduced by Jaillet [15], who studied some of its properties and proposed different heuristics and also an exact branch & bound algorithm to solve this problem [30]. It represents an optimization problem for which, there is no known algorithm that ensures the obtaining of an exact solution in polynomial time. The algorithms used for solving this problem can be classified into two categories: exact algorithms and heuristics. The exact algorithms permit finding an optimal solution, but they

 ISBN: 978-1-61208-421-3

are characterized by a complexity that increases exponentially with the problem size (the number of cities). However, the approximation algorithms or heuristics permit the obtaining of good solutions, but can not guarantee their optimality. Thus, these methods approach the optimal solution in reasonable time. Among the most commonly used techniques for PTSP resolution, we can mention the local search algorithms 2-Opt and Tabu and the branch & bound technique [9].

A. 2-Opt Algorithm

The 2-Opt algorithm was proposed by Lin-Kernighan [31] in 1973, with a purpose of ensuring the obtained starting solution improvement [32]. Indeed, a possible transformation consists in removing 2 edges of the tour and recomposing another trajectory by reconnecting remaining segments in another way, through the use of new links and by reversing their course direction if possible (suppression of 2 edges then testing all the ways to reintegrate them). The 2-Opt algorithm provides an elementary transformation consisting in selecting two nonadjacent present edges in the Hamiltonian cycle and carrying out their exchange by two other edges in a manner to obtain a new cycle. One starts with a random tour and systematically tries to find a better tour by replacing two arcs of the tour by two other arcs [31]. Figure 3 illustrates this operation.

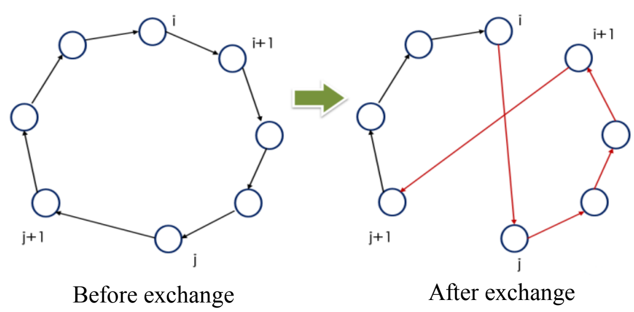

Figure 3. 2-Opt procedure

B. Tabu Algorithm

Tabu Search (TS) is a local search method combined with a set of techniques to avoid being trapped in a local minimum or cycle repetition. TS was mainly introduced by Glover [33] in 1986. This method has proved high efficiency for the resolution of difficult optimization problems. Indeed, from an initial solution in a set of local solutions, subsets of solutions that belong to this set neighborhood are generated. Through the evaluation function, we retain the solution which improves its value, selected among the set of neighboring solutions. The algorithm may accept sometimes solutions that do not always improve the current solution [26]. We implement a Tabu list containing the last visited solutions, which does not give the possibility for a solution, which is already found, to be accepted and stored in the Tabu list. Then, the choice of the next solution is carried out on a set of neighboring solutions outside of the elements of this list. When the list length is reached, each new selected solution replaces the oldest in the list. The list construction is based on the First In First Out (FIFO) principle. As a stopping criterion, one can for example fix a maximum number of iterations, or fix a limited time after which the search should stop.

TS is a local search method, and the structure of its basic algorithm is similar to that of simulated annealing, with the advantage of having a simplified parameter setting: the parameter setting consists initially in finding an indicative value of iterations during which the movements are prohibited. It will be necessary to choose a memorization strategy. However, the Tabu method requires a memory management, more and more intense, by putting complex memorizing strategies [33]. The effectiveness of the Tabu method provides its use in several classical COPs such as the TSP, the scheduling problem, the problem of vehicles, etc.

IV. KAT-MOBILITY PROTOCOL

Nakayama et al. [34] proposed the KAT-mobility protocol based on optimization methods of routing and aggregation [9]. It is a model of mobility of sinks that can effectively gather detected data employed in a WSN, even if some nodes are destroyed [35]. Indeed, the implementation of this mobility concept provides better energy management and increases the network lifetime. The system is composed of two modules: the clustering algorithm and the approximate solution for the TSP [34].

A. Clustering Algorithm

The KAT-mobility protocol is based on clustering and especially on the K-means approach [36] defined by McQueen [37] in 1967 and illustrated in Figure 4. The goal of this method is the division of the sensors set into k partitions, where each sensor will belong to the appropriate partition with the nearest average. The implementation of this algorithm is carried out by following these steps: the first step consists in choosing k that represents the number of groups to be created and then generating k initial centers randomly. The following step is to browse through all the nodes in order to assign them to the proper group whose center is the closest, on the basis of the computation of the distance between this center and each node. Then, we calculate the new centers associated with the new partition by seeking the average value of all the sensors in the respective group (centroid). Thereafter, we repeat the second step to perform the reassignment and the third step to update the average of each group. These last two steps will be then repeated until convergence to a stable partition: this convergence will be reached when there is no change. This algorithm convergence can be slow, so it is advised to add a stopping criterion, which is the case of KAT-mobility protocol. This process can be regarded as an optimization problem, as it aims to minimize the sum of the distances between group centers and the points located inside each group (the sum of the approximate errors). In fact, the group cost is estimated by the approximation error between the sensors and the center. This algorithm divides the set of nodes virtually into k clusters (C_1, C_2 … C_k) geographically close. We denote by n the nodes number in the network, usually n >> k. Let m_j (j = 1, 2 ... k) be a group center and x_i

 ISBN: 978-1-61208-421-3

(i = 1, 2 ... n) be a sensor, which is represented by a 2-dimensional vector (i.e., sensor position) [9]. d (x_i, m_j), which is indicated by the Euclidian distance between the group center and the sensor, represents the approximation error. The goal is to assign each sensor to a cluster C_j by reducing the total error of the clusters in order to reduce batteries consumption and communications energy [38]. The final objective is to assure the configuration of Cj such that the sum of approximate errors is minimized.

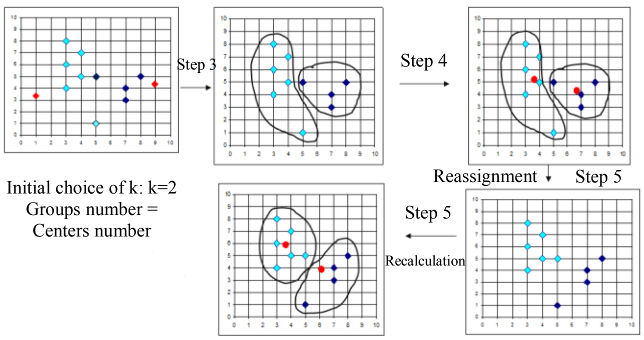

Figure 4. Operating principle of K-means algorithm

After the grouping of nodes, the K-means method reaches the mobile sink to make a course through the centers of groups, determined as anchor points, according to the trajectory of an optimized route [9]. This route is set as an approximate solution of the TSP.

B. Route Optimization

The second procedure of KAT-mobility protocol consists in optimizing the routing path of the collector (sink). Finding the best path for the mobile node is identical to the TSP. Then, a collector represents the traveling salesman and cluster centroids (centers) define cities. The route optimization of the mobile collector to visit once and only once every cluster centroid is equivalent to searching for the shortest trip of the traveling salesman in order to visit each city once [11]. Here, it is supposed that an administrator distributes nodes to supervise the targeted zone, and they are scattered at random locations and do not move afterwards. Indeed, this method reaches the mobile sink to make a course through the centers of groups according to the trajectory of an optimized path. The sink collects then the data coming from nodes on the level of the visited groups [9]. Figure 5 illustrates this principle [34].

The goal is to find the trajectory π which minimizes the tour lengths, where the initial position of the mobile sink is indicated by m_0. The quantity mentioned in (7) shows the tour length of a mobile collector that will be realized by visiting the centers in the specified order according to the permutation, while returning finally to the starting position.

$$\sum_{j=0}^{k-1} d\left(m_{\Pi(j)}, m_{\Pi(j+1)}\right) + d\left(m_{\Pi(k)}, m_{\Pi(0)}\right) \quad (7)$$

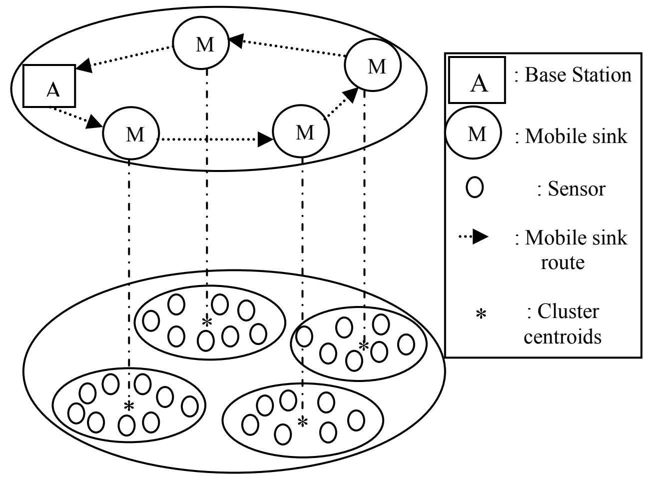

Figure 5. KAT-mobility protocol

The TSP belongs to the category of NP-hard problems and it represents an optimization problem, for which there does not exist a known algorithm that ensures the obtaining of an exact solution in polynomial time. Then, the conclusion of the optimal trajectory can not be executed in an easy manner. To solve this problem, a particular family of algorithms called the heuristics, providing the obtaining of almost optimal solutions, is proposed [9]. For the KAT-mobility protocol, the local search algorithm 2-Opt, based on the modification of a current solution to TSP by heuristics, is implemented. Fortunately, for a WSN, sensors can communicate together, and the mobile sink does not need to visit all the sensors. Therefore, it is only needed to optimize the routes among group centers. The KAT-mobility approach assumes that the collector has a priori knowledge of the positions of its member sensors [11]. It is possible that it loses communications with its blocked or failed members. In this case, it can stay at the center of its cluster to discover broken sensors and its trajectory can then be recalculated as soon as it reaches the access point [11]. Accordingly, a new algorithm was proposed in [9], with the aim of providing a better and practical solution by improving the route optimization of the sink in the presence of faults and breakdowns. In order to ameliorate this algorithm performance and the conventional KAT-mobility algorithm, our purpose in this paper is to propose a priori strategy that ensures transmission reliability, achieves better energy efficiency and guarantees the obtaining of a solution in real time once the problem is disturbed by the failures of some sensors.

V. NKAT-MOBILITY PROTOCOL

We propose through this paper to perform a transmission of information in real time and in an automated manner via a WSN. Our proposal consists in presenting the New K-means and Traveling Salesman Problem based mobility (NKAT-mobility) protocol, based on K-means algorithm and the routing path optimization by using the Tabu method. The aim of this paper is to present this a priori strategy which ensures the reliability of data transmission in the presence of faults, so we model the wireless network by a graph G (V, E), where V is the set of group centers and E represents the set of edges reflecting the possible communications between these points. We denote by k the number of vertices in the graph, which is considering here as the problem size. In wireless sensor networks, one or several sensors may not function correctly, so the number of functional vertices varies between 0 and k. We are working in a situation where the vertices do not exist in a deterministic manner in this graph, but they are present in a probabilistic way. In other words, a probability of presence p_j will be associated with each vertex m_j (j = 1, 2 ... k). Moreover, on any instance of the problem, we avoid the total flaw of the system by changing the graph structure and transforming it into a subgraph, according to a modification strategy that will be specified in advance.

For the KAT-mobility protocol, the local search algorithm 2-Opt, based on the modification of a current solution to TSP by heuristics, is implemented. To ameliorate this work in the deterministic case, which is a particular case of this probabilistic problem, i.e., when p_j (j = 1, 2 ... k) = 1, we propose the improvement of the route optimization by implementing the Tabu algorithm instead of the used method 2-Opt. This algorithm depends on an initial solution, the neighborhood and a Tabu list. It is possible that the mobile sink lose communications with its failed or broken member nodes. In this case, it can stay at the center of its group to discover failed nodes and its trajectory can be recalculated as soon as it reaches the access point [11]. Updating this route is preceded by a modification of the centroids locations in the WSN. In other words, a new clustering procedure through the subset of operational sensors is performed. Consequently, we propose that only some centers among all the vertices will really necessitate a visit according to their probabilities of appearing (remaining functional) and we assume that when all a cluster members are failed, its center will be regarded as absent. We specify then a modification strategy μ, which removes absent centers from the initial a priori trajectory. This is a problem of finding a priori trajectory that minimizes the functional of covered distances. Our aim consists in obtaining a tour among the initial vertices such that G is transformed into the subgraph G'= G [V'], where V' ⊆ V is the set of present group centers and the new path through its vertices will be in the same order as that created by the a priori tour. Given the set of cluster centers, the probability law ℙ, the set of all the subsets of V, i.e., each instance V' ⊆ V has a probability of presence ℙ (V'). For a given trajectory R through the vertices defined on V, the modification strategy μ used for generating a realizable solution through V', consists in gumming or eliminating those who are absent from the a priori trajectory. Let $L_{(R,\mu)}$ be the random variable defined on 2^V, which for a trajectory R and for all V' ⊆ V, associates the length $L_{(R,\mu)}$ (V') through V', induced of the trajectory R by the modification method. Therefore, the route optimization of the mobile sink to visit once and just once every node is equivalent to find the trajectory that minimizes the functional of $L_{(R,\mu)}$ [21]. To solve our cited problem, this a priori strategy is modeled with the aim to maintain network functionality and to avoid the total system flaw despite of the existence of faults in a subset of its elementary components by minimizing the objective function of this problem. So, we propose the implementation of the Tabu algorithm for optimizing the routing path of the sink, as a second procedure of NKAT-mobility protocol, in the deterministic case as well as in the probabilistic case (in presence of breakdowns, i.e., when p_j (j = 1, 2 ... k) ≠ 1). The functional of covered distances depending on R and μ, can be expressed by:

$$\min_{R}\left(\mathbb{E}\,(L_{(R,\mu)}) = \sum_{V'\subseteq V} \mathbb{P}\,(V')\;\; L_{(R,\mu)}(V')\right) \quad (8)$$

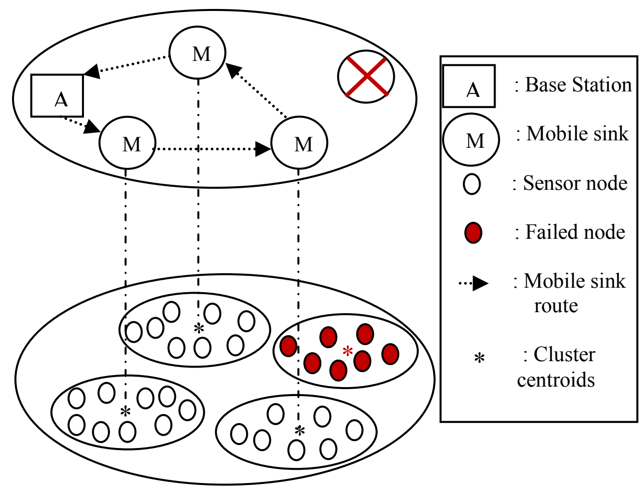

Figure 6. NKAT-mobility protocol

NKAT-mobility protocol allows us to have a real time solution, which is valid in all situations. This a priori strategy is used to achieve better performance than other existing methods and the new route obtained through the set of functional centroids is illustrated in Figure 6.

VI. SIMULATION RESULTS AND ANALYSIS

In this section, simulation results are presented and analyzed. In fact, simulation is a very important step that enables us to analyze a WSN performance [39]. In fact, it is necessary to reduce the various possible conception errors by performing a validation step. To achieve our simulations and to study our approach (NKAT-mobility) performance, we have used the software Java. We have simulated the KAT-mobility protocol, the existed strategy proposed by Bellalouna et al. [9] and the NKAT-mobility protocol. In fact, we have implemented the K-means algorithm that represents the first module of each protocol and thereafter the second module of each one, which differs from one method to the other.

The WSN that we have simulated comprises 200 sensor nodes, which are dispersed inside the simulation area. They are deployed according to a random distribution in order to test our proposed strategy performance under conditions that are more similar to those real. We consider the metric tour cost (expressed in meter) for estimating the tour lengths.

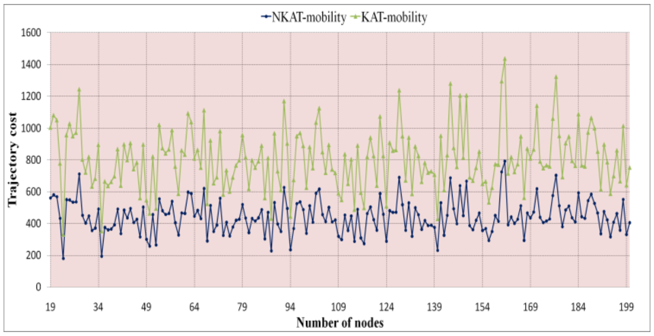

Figure 7. Variation of the trajectory cost according to the number of nodes in deterministic case

Through Figure 7 that represents the trajectory cost depending on the number of nodes, varying from 19 to 200, we can see that the tour length obtained by NKAT-mobility protocol for any used distribution is lower than that obtained by KAT-mobility protocol. Then, we can conclude that the new method based on K-means clustering and the approximate solution for TSP by using the local search method Tabu achieves better performance than the conventional method based on K-means clustering and the approximate solution for TSP by employing the simple local search algorithm 2-Opt, in deterministic case. In other words, NKAT-mobility protocol ensures the minimization of the trajectory length for any employed distribution, when the probability of presence p_j, associated with each vertex m_j (j = 1, 2 ... k), is equal to 1 (all the nodes are operational).

The nodes number may vary from one day to the other because of the possible failures and attacks, so the NKAT-mobility protocol was proposed to deal with this problem (probabilistic case). To confirm our a priori strategy performance, we have executed the simulations of the three mentioned algorithms: KAT-mobility, NKAT-mobility and the existed algorithm proposed by Bellalouna et al. [9].

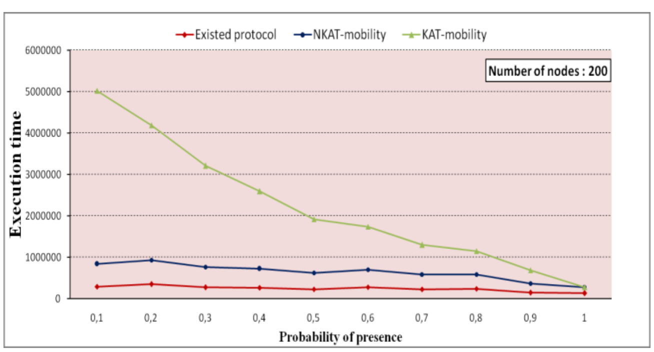

Figure 8. Variation of the execution time according to the probability of presence in probabilistic case

It is shown from the simulation results of Figure 8, representing the time execution, which is expressed in nanoseconds, depending on the presence probability of cluster centroids, that the NKAT-mobility protocol provides less execution time than the KAT-mobility protocol. Our proposed probabilistic model is much more realistic, once the problem is disturbed by the failures of some nodes (i.e., the probability of remaining functional is less than 1). It is clear from the simulation results that the reduction in terms of time, compared with the conventional method, is very important. However, we can notice a small gap between our protocol and the existed protocol. Therefore, it was interesting to compare these two methods by evaluating the functional of covered distances.

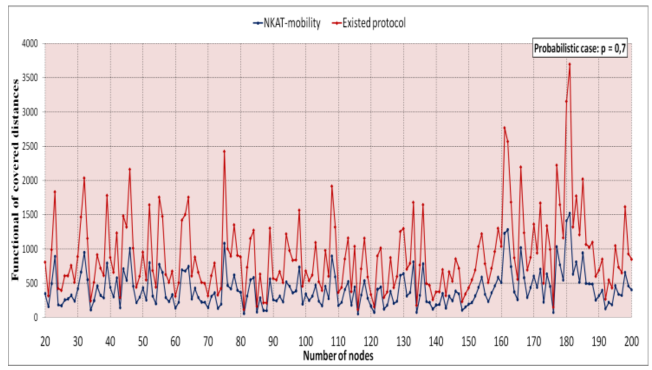

Figure 9. Performance comparison with failed nodes (30% absent)

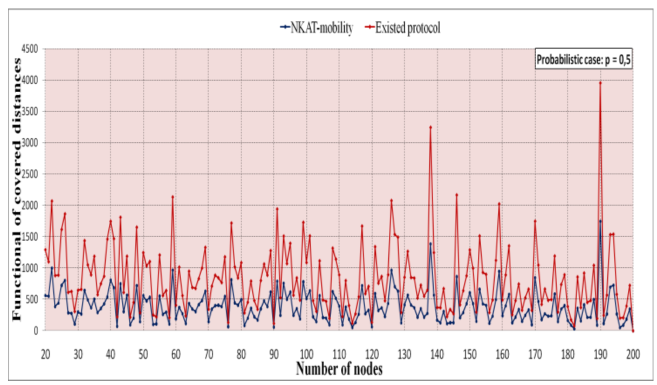

Figure 10. Performance comparison with failed nodes (50% absent)

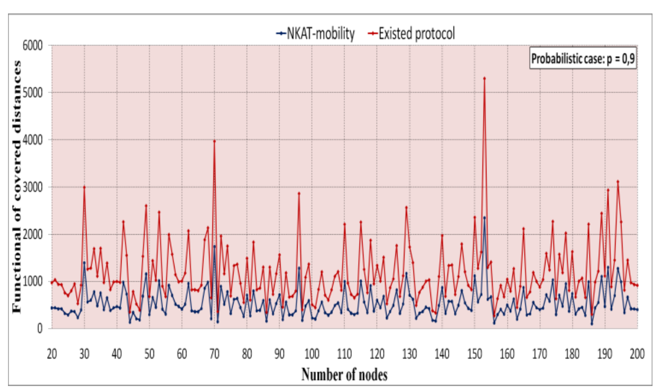

Figure 11. Performance comparison with failed nodes (10% absent)

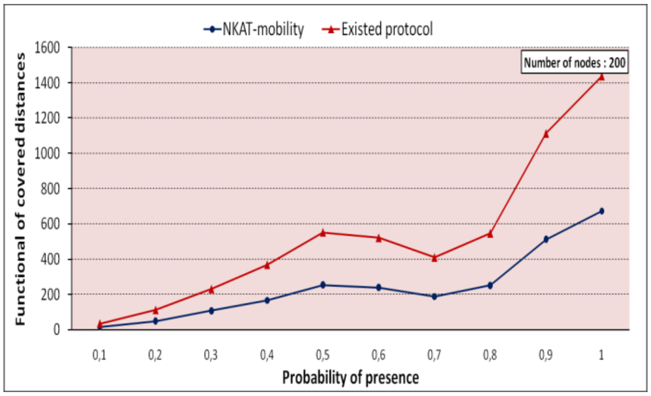

Figure 12. Variation of the functional according to the probability of presence in probabilistic case

Through Figures 9, 10 and 11, we can see that the functional of covered distances obtained by our method (NKAT-mobility) is lower than that of the existed strategy proposed by Bellalouna et al. [9], for any used distribution and for any tested values of p_j. In fact, these figures representing the functional variation, with values of n ranging from 20 to 200 and with different values of the presence probability (the percentages of centroids absence are 10%, 30%, 50%), shows that NKAT-mobility achieves better performance than the existed protocol and guarantees the functional minimization. This idea is also confirmed in Figure 12 that represents the functional variation when p_j varies and when n is fixed at 200. It is clear from our simulations that the improvement provided by our new method is effective and that our algorithm is more realistic than the fault tolerant KAT-mobility protocol, valid in all situations and better than the existed protocol in terms of solution quality (optimized routing path).

VII. Conclusion and Future Works

In this paper, a new strategy of data collection in a WSN has been proposed. The novelty of this scheme is that the combinatorial optimization provides applicable approaches in the context of wireless sensor networks. In other words, the WSN can be modeled as a probabilistic combinatorial optimization problem, particularly in presence of breakdowns. Then, we have presented the fault tolerant NKAT- mobility protocol that assures finding a real time solution to each new problem instance, once it is disturbed by the absence of some nodes. This a priori strategy aims to provide better practical solution compared with two conventional methods based on K-means clustering and the approximate solution for TSP, in probabilistic case. Simulation results demonstrated that the performance differences between our probabilistic model and the two methods are significant. We can conclude that NKAT-mobility ensures the improvement of the collector route optimization. To accomplish this work, we aim to use a new strategy by implementing various algorithms employed for probabilistic combinatorial optimization problems resolution, such as approximate methods and this will be so interesting when the risk of breakdowns becomes considerable and the number of cluster centroids will be reduced.

References

[1] S. Clement, “Some contributions in wireless sensors networks: Localization and Routing,” Ph.D thesis, Avignon University, France, 2008.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” Computer Networks, vol. 38, issue 4, 2002, pp. 393-422, doi: 10.1016/S1389-1286(01)00302-4.

[3] H. Bousetta, “Multiphysics modeling and global simulation of autonomous systems on chip,” Ph.D thesis, National Polytechnic Institute of Grenoble, France, 2010.

[4] K. Bouchakour, “Hierarchical routing in wireless sensor networks: KhLCH protocol (K-hop Layered Clustering Hierarchy),” Ph.D thesis, National High School of Computer Sciences, Algeria, 2012.

[5] M. Hempstead, M. J. Lyons, D. Brooks, and G. Wei, “Survey of hardware systems for wireless sensor networks,” Low Power Electronics, vol. 4, issue 1, 2008, pp. 11-20, doi: 10.1166/jolpe.2008.156. 2008.

[6] M. Bellalouna, “Probabilistic combinatorial optimization problems,” Ph.D thesis, National School of Bridges and Roads, Paris, 1993.

[7] M. Bellalouna and A. Ghabri, “A priori methods for fault tolerance in wireless sensor networks,” World Congress on Computer and Information Technologies (WCCIT), IEEE, Sousse, 2013, pp. 1-6, doi: 10.1109/WCCIT.2013.6618654.

[8] C. Karlof and D. Wagner, “Secure routing in wireless sensor networks: attacks and countermeasures,” Ad Hoc Networks, vol. 1, 2003, pp. 293-315, doi: 10.1016/S1570-8705(03)00008-8.

[9] M. Bellalouna, A. Ghabri, and W. Khaznaji, “The K-means and TSP Based Mobility Protocol Modeling as a Probabilistic Combinatorial Optimization Problem,” in proceedings of the Ninth International Conference on Systems and Networks Communications (ICSNC), IARIA, Nice, France, October 2014, pp. 48-53.

[10] D. J. Bertsimas, “Probabilistic combinatorial optimization problems,” Ph.D thesis, MIT, Cambridge, Mass, 1988.

[11] H. Nakayama, N. Ansari, A. Jamalipour, and N. Kato, “Fault-resilient sensing in wireless sensor networks,” Computer communications, vol. 30, issue 11, September 2007, pp. 2375-2384, doi: 10.1016/j.comcom.2007.04.023. 2007.

[12] P. Jaillet, “Analysis of Probabilistic Combinatorial Optimization Problems in Euclidean Spaces,” Mathematics of Operations Research, vol. 18, issue 1, February 1993, pp. 51-70, doi: 10.1287/moor.18.1.51.

[13] D. Bertsimas, P. Jaillet, and A. Odoni, “A priori Optimization,” Operations Research, vol. 38, issue 6, December 1990, pp. 1019-1033, doi: 10.1287/opre.38.6.1019.

[14] A. Henchiri, M. Bellalouna, and W. Khaznaji, “Probabilistic traveling salesman problem: a survey,” in proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS), Warsaw, September 2014, pp. 55-60, doi: 10.15439/2014F381.

[15] P. Jaillet, “Probabilistic Traveling Salesman Problems,” Ph.D thesis, MIT, Cambridge, Mass, USA, 1985.

[16] P. Jaillet, “Analysis of the probabilistic traveling salesman problem in the plane,” Ricerca Operativa, vol. 36, January 1986, pp. 57-68.

[17] C. Murat and V. Paschos, “The probabilistic longest path problem,” Networks, vol. 33, issue 3, 1999, pp. 207-219, doi: 10.1002/(SICI)1097-0037(199905)33:3<207::AID-NET7>3.0.CO;2-7.

[18] C. Murat and V. Paschos, “The probabilistic minimum vertex covering problem,” International Transactions in Operational Research, vol. 9, issue 1, January 2002, pp. 19-32, doi: 10.1111/1475-3995.00338.

[19] Z. Bouyahia, M. Bellalouna, P. Jaillet, and K. Ghedira, “A Priori Parallel Machines Scheduling,” Computers & Industrial Engineering, vol. 58, issue 3, April 2010, pp. 488-500, doi: 10.1016/j.cie.2009.11.009.

[20] M. Bellalouna and L. Horchani, “The Two-Dimensional Probabilistic Bin-Packing Problem: An average case analysis,” International Journal of mathematics and computer in simulation, vol. 2, issue 1, March 2008, pp. 42-46.

[21] P. Jaillet, “A priori solution of a Travelling Salesman Problem in which a random subset of the customers are visited,” Operations research, vol. 36, issue 6, Maryland, US, 1988, pp. 929-936, doi: 10.1287/opre.36.6.929.

[22] M. Bellalouna, C. Murat, and V. Paschos, “Probabilistic combinatorial optimization problems on graphs: A new domain in operational research,” European Journal of Operational Research, vol. 87, issue 3, December 1995, pp. 693-706, doi: 10.1016/0377-2217(95)00240-5.

[23] L. Forlizzi, J. Hromkovic, G. Proietti, and S. Seibert, “On the stability of approximation for hamiltonian path problem,” in proceedings of the Conference on Current Trends in Theory and Practice of Computer Science, Berlin, 2005, pp. 147-156, doi: 10.1007/978-3-540-30577-4_18.

[24] D. J. Bertsimas and L. H. Howell, “Further results on the probabilistic traveling salesman problem,” European Journal of Operational Research, vol. 65, issue 1, February 1993, pp. 68-95, doi: 10.1016/0377-2217(93)90145-D.

[25] D. J. Bertsimas, “A Vehicule Routing Problem with Stochastic Demand,” Operational Research, vol. 40, issue 3, June 1992, pp. 574-585, doi: 10.1287/opre.40.3.574.

[26] M. Gendreau, G. Laporte, and R. Séguin, “A Tabu Search Heuristic for the Vehicle Routing Problem with Stochastic Demands and Customers,” Operations Research, vol. 44, issue 3, May 1996, pp. 469-477, doi: 10.1287/opre.44.3.469.

[27] S. Rosenow, “Comparison of an Exact Branch-and-Bound and an Approximative Evolutionary Algorithm for the Probabilistic Traveling Salesman Problem, Operations Research, August 1998, pp. 168-174, doi: 10.1007/978-3-642-58409-1_16.

[28] D. S. Johnson and L. A. McGeoch, “The Traveling Salesman Problem: A Case Study in Local Optimization,” John Wiley & Sons, New York, NY, USA, 1997, pp. 215-310.

[29] M. Bellalouna, V. Paschos, and W. Khaznaji, “Well solved cases of probabilistic traveling salesman problem,” 42nd days of statistics, INRIA, Marseille, France, June 2010, pp. 1-7.

[30] P. Jaillet, “On some probabilistic combinatorial optimization problems defined on graphs,” Flow Control of Congested Network, vol. 38, 1987, pp. 255-267, doi: 978-3-642-86726-2_16.

[31] L. Bianchi, J. Knowles, and N. Bowler, “Local search for the probabilistic traveling salesman problem: Correction to the 2-p-opt and 1- shift algorithms,” European Journal of Operational Research, vol. 162, issue 1, February 2005, pp. 206-219, doi: 10.1016/j.ejor.2003.10.016.

[32] L. Bianchi and A. M. Campbell, “Extension of the 2-p-opt and 1-shift algorithms to the heterogeneous probabilistic traveling salesman problem,” European Journal of Operational Research, vol. 176, issue 1, January 2007, pp. 131-144, doi: 10.1016/j.ejor.2005.05.027.

[33] F. Glover, “Future paths for integer programming and links to artificial intelligence,” Computer & Operations Research, vol. 13, issue 5, 1986, pp. 533-549, doi: 10.1016/0305-0548(86)90048-1.

[34] H. Nakayama, N. Ansari, A. Jamalipour, Y. Nemoto, and N. Kato, “On data gathering and security in wireless sensor networks,” in proceedings of the IEEE Sarnoff symposium, IEEE, Nassau Inn, Princeton, NJ, 2007, pp. 1-7, doi: 10.1109/SARNOF.2007.4567337.

[35] J. P. Jafrin and P. A. C. Angelin, “A comparative study of data gathering algorithms for a mobile sink in wireless sensor network,” International journal of advanced research in computer engineering and technology, vol. 1, issue 9, November 2012, pp. 250-255.

[36] Z. Guellil and L. Zaoui, “Proposition of a solution to the initialization problem K-means Case,” in proceedings of the second International Conference on Informatics and Applications (CIIA), Saida, Algeria, 2009, pp. 1-9.

[37] P. Mathur, S. Saxena, and M. Bhardwaj, “Node clustering using K Means Clustering in Wireless Sensor Networking,” in proceedings of the second National Conference in Intelligent Computing Communication (EICC), Department of Computer Science and Engineering, Indian Institute of Technology, Uttar Pradesh, India, 2013, pp. 1-6.

[38] T. Fukabori, H. Nakayama, H. Nishiyama, N. Ansari, and N. Kato, “An efficient data aggregation scheme using degree of dependence on clusters in WSNs,” in proceedings of the IEEE International Conference on Communications, May 2010, pp. 23-27, doi: 10.1109/ICC.2010.5502285.

[39] A. Makhoul, “Sensor Networks: localization, coverage and data fusion,” Ph.D thesis, University of Franche-Comte, France, 2008.