Optimal Self Boundary Recognition with Two-Hop Information for Ad Hoc Networks

Yen-Hsu Chen[1], Wei-Ho Chung[2], Guo-Kai Ni[1], Hongke Zhang[3], Sy-Yen Kuo[1,3]

Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan[1]
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan[2]
College of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China[3]

{sykuo@cc.ee.ntu.edu.tw}

***Abstract*—The ad hoc network is composed of multiple sensor nodes to serve various applications, such as data collection or environmental monitoring. In many applications, the sensor nodes near the boundary of the deployment region provide biased or low-quality information because they have limited number of neighboring nodes and only partial information is available. Hence, the boundary recognition is an important issue in the ad hoc networks. By the statistical approach in high node density networks, Fekete's pioneer work identified the boundary node by number of neighboring nodes and using a specific threshold. By exploiting the number of nodes in the two-hop region, our proposed algorithm has significant improvement of boundary recognition contrasted with Fekete's algorithm in the low-density network. Given the information topology and the cost function, the analyses provide a framework to obtain the optimal threshold for boundary recognition. Besides, the simulation results reveal the proposed algorithm has greater than 90% detection rate and lower than 10% false alarm rate.**

Keywords-ad hoc networks, boiundary recognition, decision, optimal algorithm

I. INTRODUCTION

The ad hoc networks are composed of numerous low-cost wireless nodes. The nodes with limited accuracy sensors, battery lifetime, and computing power are used to monitor the region of interest (ROI). Certain applications also use sensors to measure or detect the interesting events, such as animal or human motions. Due to the limitation of the hardware, the researchers developed energy-efficient and distributed algorithms to solve many specific applications in the ad hoc networks. Through the cooperation of sensor nodes, the nodes perform specific application in distributed schemes, such as the fusion of data measurement or the event detection. However, the performance of the applications degrades in the boundary region of ROI. Because the nodes in the boundary have fewer neighboring nodes, the nodes obtain only partial information related to the goal of the application. Therefore, the recognition of boundary nodes becomes an important issue in the ad hoc networks.

The related works can be classified into three categories: the geometrical scheme, statistical scheme and topology scheme.

In the geometrical scheme, the geometrical information is available to the sensor nodes. Fang et al. [1] transform the greedy geographical routing into a local minima problem. The geographical localization is assumed available in the algorithm, and the boundary nodes route through longer path. Localization-aware devices cause the extra energy cost to a sensor node and reduce the lifetime of the sensors.

In the topology scheme, the algorithms utilize the connectivity information among the nodes and the topology constraints to detect the boundaries and holes in ROI. S.Funke et al. [2] constructed ios-lines from selected nodes. If the ios-line is broken at both edges, the node is recognized as the boundary node or near boundary node. The algorithm [3] improves the Funke's work by iteratively collecting the information of 2-hop neighbor ios-lines. Every node checks whether the ios-line is broken by itself. By Y. Wang et al. [4], an root node is selected to build a shortest path tree for checking the holes. However, the process of network floods requires the synchronization of nodes. The algorithm proposed by Kroller et al. [5] recognizes a topological structure called flowers and augmented cycle to detect the boundary. O.Saukh et al.[6] also search a special pattern of geometric structure. The researchers [7] partition the network into clusters, the algorithm searches the boundary nodes in the overlapping regions of the clusters.

In statistical scheme, the nodes are identified as the boundary nodes or interior nodes by the characteristic of the networks, such as distribution of the nodes. Fekete et al.[8] assumes that nodes on the boundaries have lower average number of neighboring nodes than the other nodes in the interior of network. The algorithm requires high node density, e.g., 100 nodes in one communication range. Fekete et al.[9] designed a threshold depending on the area of boundary node and interior nodes, and then classified the nodes as boundary and interior nodes. The algorithm also requires high node density. Another issue in [9] is that the decision by one-hop information has higher variance, so Bi. et al. [10] improved the threshold by averaging the number of one-hop neighboring nodes in the two-hop region. The algorithm proposed by Ghrist et al. [11] detects the hole via homology. In order to construct the homology in mathematics, the sensing and communication range must be carefully regulated. G. Destino et al. [12] constructed clusters to recognize the boundary nodes and the interior nodes.

In this paper, we propose a novel boundary recognition scheme in the uniformly and randomly deployed ad hoc

*Acknowledgment: This research was supported by National Science Council, Taiwan under Grant NSC 99-2221-E-002-108-MY3, National Natural Science Foundation of China (NSFC) under Grant 60833002, and "111" Project under Grant B08002.

networks. By using the number of nodes in two-hop region, our algorithm achieve high detection rate for the boundary recognition in low node density compared with [8]. The main contribution of this work is to transform the boundary recognition problem into the binary hypothesis test problem, and analyze the detection rate and false alarm rate. Given the topology information of the network and the cost function, our proposed approach provides a framework for designing the optimal threshold of the binary hypothesis test problem.

II. THE BAIC CONCEPT AND OUR ALGORITHM

We assume the distribution of the nodes in the ad hoc network follows a Poisson point process (PPP) with the intensity λ_1, and all nodes are equipped with the same communication range (CR) in the ROI F. Let $b(.)$ represent the function of a boundary of a network, e.g. the borders of a network. Whether a node D is a boundary node or not is transformed into a binary hypothesis test problem written as

$$H = \begin{cases} H_1: \exists b, dist(b, D) \leq CR; \\ H_0: \forall b, dist(b, D) > CR. \end{cases} \quad (1)$$

where $dist(.)$ is the shortest Euclidean distance between a boundary and a node. If a node is not a boundary node, we say the node is an interior node.

The average number of nodes is λ_1 over a unit area, so the average number of nodes is $\lambda = \lambda_1 CR^2 \pi$ over the circle with radius CR. In the PPP constructed nodes, we define $N(R) = |R|/\lambda$ as the average number of nodes in the region R where $|.|$ is the area operation. Therefore, we define the $N(C) = \lambda$ as the average number of nodes in the circle C centered an interior node with radius CR. For a boundary node, we define the $N(B)$ as the average number of nodes in the intersection region $B = C \cap F$. As in [8], the authors employed the relation $N(B) < N(C)$ and a threshold to classify the boundary node and interior node. We extends the relation, $N(B) < N(C)$, to two-hop region. The two-hop region $R_{C,2}$ is defined as the region covered by a node and its neighboring nodes. Let $N(R_{i,2})$ denote the average number of nodes in $R_{C,2}$ of an interior node and $N(R_{b,2})$ denote the average number of nodes in two-hop region of a boundary node. The relation $N(R_{i,2}) > N(R_{b,2})$ still holds if $R_{C,2}$ is available, but $R_{C,2}$ is an irregular area and is difficult to be calculated. Hence, we derive the approximate circular area called two-hop circle to serve as the equivalent area for the $R_{C,2}$. In [13], the excepted distance from a source node to a destination node per hop is derived as

$$D_{hop} = CR\left(1 + e^{-\lambda} - \int_{-1}^{1} e^{\frac{\lambda}{\pi}\left(acos(t) - t\sqrt{1-t^2}\right)} dt\right). \quad (2)$$

The distance $\beta CR = 2D_{hop}$ is taken as the radius of the equivalent two-hop circle where $1<\beta \leq 2$, and the number of nodes in the two-hop circle is expected to be close to the number of nodes in $R_{C,2}$. Fig. 1 demonstrates the highly accurate approximation between the number of nodes in the equivalent two-hop circle and the number of nodes in the true $R_{C,2}$. Let λ_β denote the average number of nodes in the two-hop circle of an interior node, and λ_b denote the average number of nodes in the two-hop circle of a boundary node. The result in Fig. 1 also reveals the circular area of diameter λ_β approximates the area constructed by two-hop neighboring nodes. Because λ_β is greater than λ_b, a node decides whether itself is a boundary node or not by setting a specific threshold on the number of nodes in its two-hop circle, and the threshold is related to λ_β and λ_b. Given a cost function and the information of the topology of the network, the threshold can be designed to achieve the optimal cost. The design of the threshold is discussed in the section III. The rationale of our algorithm is shown in Fig. 2. Because we have $(\lambda_\beta - \lambda_b) > (\lambda - N(B))$, our algorithm can easily find a threshold to classify a node as a boundary node or an interior node. Since λ_β is greater than λ, our algorithm is more suitable to sensor deployment with low node density than the algorithm in [8].

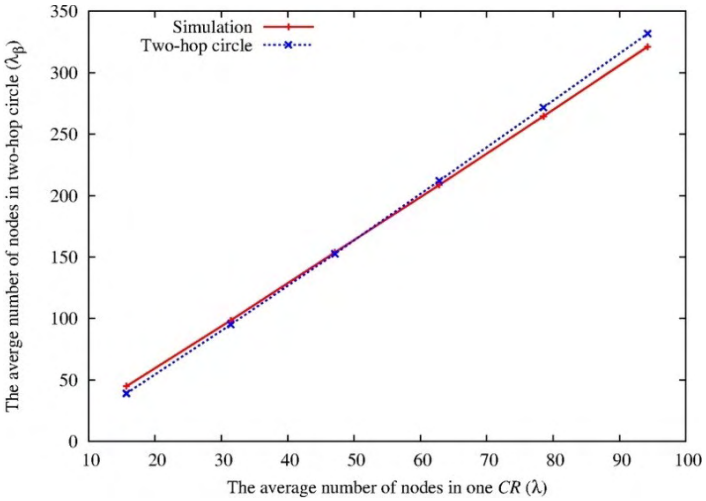

Fig. 1. The average number of nodes in two-hop circle and two-hop region

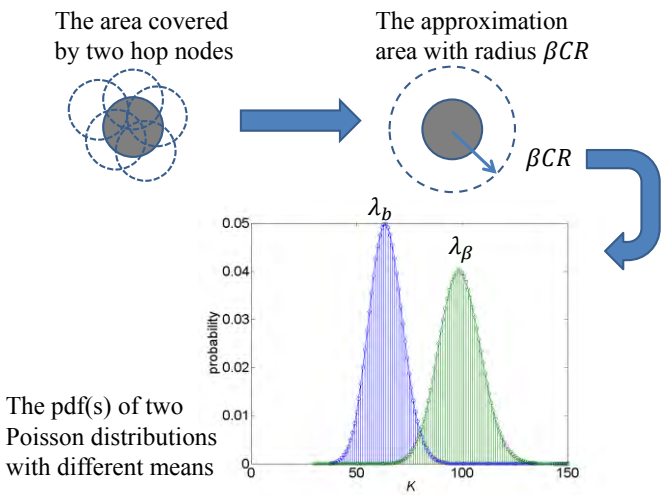

Fig. 2. The concept of the proposed algorithm

The value λ is the prior knowledge available to every node, or a packet with λ is broadcasted to the whole network after the nodes deployment process. Our algorithm of a node classification is simplified into three steps:

1) Collecting the number of nodes in the $R_{C,2}$;
2) Estimating the radius of two-hop circle;
3) Classifying the node as a boundary node or not by the number of nodes $Nodes(R_{C,2})$ in $R_{C,2}$ with a threshold T, where T is an positive integer and the decision rule is written as

$$Nodes(R_{C,2}) \underset{H_0}{\overset{H_1}{\gtrless}} T \tag{3}$$

III. THE ANALYSIS OF DETECTION RATE AND FALSE ALARM RATE IN THE SQUARE FIELD

A. The two-hop regions of the network

The system detection rate is related to the ratio of the boundary region and the non-boundary region of the ROI. In other words, the system detection rate is topology dependent. In order to analyze the system detection rate and false alarm rate, we consider the ROI of a $b \times b$ square area and assume CR=1 without loss generality. The ROI is partitioned into certain regions shown in Fig. 3. The boundary regions are divided into two kinds. For a node in black region or gray region, the $|R_{C,2}|$ of this node is constructed by the two-hop circle and a boundary. For a node in the slash line region or dot region, the $|R_{C,2}|$ of this node is constructed by the two-hop circle and two boundaries. For a node D in the black or gray region of the top in the network, the shortest distance from the node to a boundary is $y = dist(b, D)$ as shown in Fig. 4. The difference between the dark region and gray region is the range of y. For a node in the dark region, the y is in [0,1]. For a node in gray region, the y is in $(1, \beta]$.

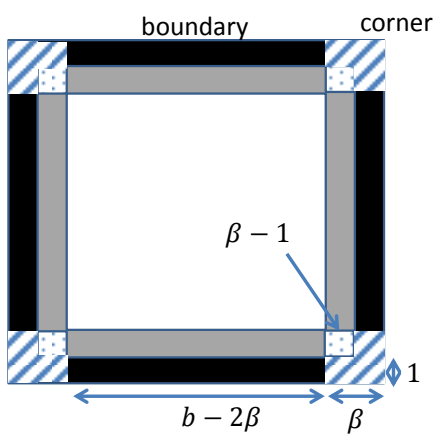

Fig. 3. The partitions of the ROI.

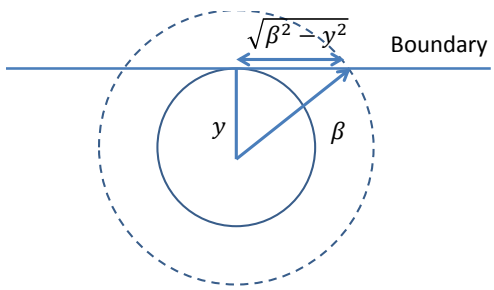

Fig. 4. The area constructed by the two-hop circle and the boundary.

After triangular manipulations, the area constructed by tow-hop circle and a boundary can be written as

$$f(\beta, y) = \left(\pi - \operatorname{acos}\left(\frac{y}{\beta}\right)\right)\beta^2 + \sqrt{\beta^2 - y^2}\, y. \tag{4}$$

The average number of nodes in the two-hop circle in the black or gray region is written as

$$\lambda_y = \frac{\lambda f(\beta, y)}{\pi}. \tag{5}$$

If a node A falls into slash line region at the upper and left corner of the square field, the distances to two boundaries are denoted as x and y where $0 \le x, y \le \beta$ respectively. We have

$$\sin(\theta_1) = \frac{y}{\beta}, \sin(\theta_2) = \frac{x}{\beta}. \tag{6}$$

Therefore, the area constructed by the two-hop circle of A and the boundaries can be written as

$$g(\beta, x, y) = \frac{y(x+\beta) + (y+\beta)x}{2} + \left(\frac{\operatorname{asin}\left(\frac{y}{\beta}\right) + \operatorname{asin}\left(\frac{x}{\beta}\right)}{2\pi} + \frac{1}{4}\right) * \beta^2\pi. \tag{7}$$

The average number of nodes in the area constructed by the two-hop circle of A and the boundaries is written as

$$\lambda_c = \frac{\lambda g(\beta, x, y)}{\pi}. \tag{8}$$

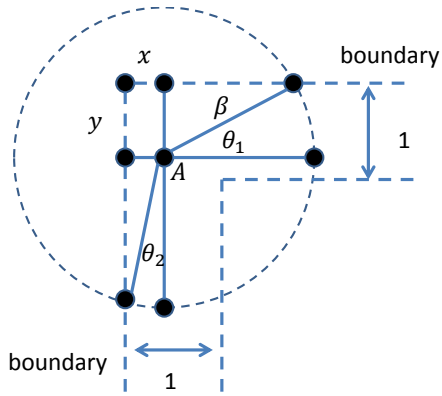

Fig. 5. A node in the corner region

B. Detection rate and false alarm rate

Given the threshold T, the detection rate can be obtained by computing the cumulative distribution function of PPP conditioned on the average number of nodes in two-hop region $\lambda_a = N(R_{C,2})$. Therefore, the detection rate of a boundary node can be written as

$$P(k \le T|\lambda_a) = \sum_{k=0}^{T} \frac{\lambda_a^k}{k!} e^{-\lambda_a}, \tag{9}$$

where λ_a is the number of nodes in the area constructed by the boundary and the two-hop circle. In PPP, a boundary node in any coordinate of the boundary region is $1/(b - 2\beta)$. We consider the detection rate in the upper black region, which is written as

$$P_{b,d}(\beta, T) = \frac{1}{b - 2\beta} \int_{\beta}^{b-\beta} \int_{0}^{1} P(k \le T|\lambda_y) dy dx. \tag{10}$$

Considering the slash lines region at the upper and left corner of the square field, the detection rate of the nodes is written as

$$P_{c,d}(\beta,T)=\frac{1}{2\beta-1}\left\{\int_0^{\beta}\int_0^{1}P(k\le T|\lambda_c)dydx+\int_0^{1}\int_1^{\beta}P(k\le T|\lambda_c)dydx\right\}. \quad (11)$$

Since the detection rates of all black regions are the same and the detection rates of all gray regions are the same, the overall system detection rate is written as

$$P_d=\frac{4(b-2\beta)}{4b-4}P_{b,d}+\frac{4(2\beta-1)}{4b-4}P_{c,d}. \quad (12)$$

With the growth of λ, $P_{b,d}$ and $P_{c,d}$ are converged to an upper bound. In addition, β is close to 2. When λ and b are large e.g. $\lambda=60$ and b=10, the P_d is dominated by the first term of (12). The result indicates that system detection rate is independent of the network scale with large λ and b.

Similarly, the false alarm rate equals the sum of false alarm rate in corner region of the gray regions, the dot regions and the white regions. The false alarm rate in white region is

$$P_{w,fa}(\beta,T)=\frac{1}{(b-4)^2}P\left(k\le T|\lambda_\beta\right). \quad (13)$$

The false alarm in the gray region is

$$P_{g,fa}(\beta,T)=\frac{1}{b-2\beta}\int_{\beta}^{b-\beta}\int_{1}^{\beta-1}P\left(k\le T|\lambda_y\right)dydx. \quad (14)$$

The false alarm rate in the dot region is

$$P_{c,fa}(\beta,T)=\frac{1}{(\beta-1)^2}\int_{1}^{\beta-1}\int_{1}^{\beta-1}P(k\le T|\lambda_c)dydx. \quad (15)$$

Hence, the system false alarm rate is written as

$$P_{fa}=\frac{(b-4)^2}{(b-2)^2}P_{w,fa}+\frac{4(b-2\beta)(\beta-1)}{(b-2)^2}P_{g,fa}+\frac{(\beta-1)^2}{(b-2)^2}P_{c,fa}. \quad (16)$$

With the growth of λ, $P_{c,fa}$ and $P_{g,fa}$ are converged to a lower bound and $P_{w,fa}$ is close to zero. When λ and b are large, P_{fa} is dominated by the second term of (16). The result indicates that system false alarm rate is independent of the network scale with large λ and b. With the growth of λ, the P_d increases and P_{fa} decreases.

Given λ and b, the system detection rate and false alarm rate are derived in the (12) and (16). The results can be solved by the numeric methods even though (12) and (16) are not in the close form. Moreover, the optimal threshold is derived by the (12) and (16). According to the maximum a-posteriori (MAP), the area of boundary region is assumed to be known, and then the optimal threshold can be derived. Taking the square as an example, the probability of nodes in the boundary region is $P_1=(4b-4)/b^2$ and the probability of nodes in interior region is $P_0=(b-2)^2/b^2$. Besides, we assume the cost of false negative is C_{01}, and the cost of false positive is C_{10}. With the parameters, P_1, P_0, C_{01}, C_{10} , the cost of function is defined as

$$Cost=C_{01}P_1(1-P_d)+C_{10}P_0P_{fa}, \quad (17)$$

Since the P_d and P_{fa} are functions of T, the minimum cost can be obtained by

$$\arg\min_T(C_{01}P_1(1-P_d)+C_{10}P_0P_{fa}). \quad (18)$$

IV. Simulation and Evaluation

In order to evaluate the detection rate and false alarm rate of our algorithm, there are N nodes deployed in a square with the square area 100×100 (m^2) or 200×200 (m^2) by following PPP. The N varies from 1000 to the 3000 with step 500 for b=100 and varies from 4000 to 12000 with step 2000 for b=200. Every node has the same communication range $CR=10$ (m), so the parameter λ varies from 31 to 94. Every node exchanges information with neighboring nodes for calculating the number of nodes in $R_{c,2}$, and then every node makes a binary decision by itself with the specific threshold T. The average detection rate and false alarm rate are calculated by averaging values from 100 different topology simulations.

The receiver operation curves (ROCs) of simulation results are shown in the Fig. 6, and it is noted that the x-axis is in the log scale. For clarity of presentation, the data lines of Fekete's algorithm in 200×200 (m^2) are not plotted in Fig. 6. Our algorithm has high detection and lower false alarm compared with the Fekete's algorithm. Besides, the threshold can be designed to achieve the detection rate greater than 90% and the false alarm rate lower than 10% except $\lambda\approx31$.

The rectangle of Fig. 6 points out the similar detection rates of our algorithm. The result indicates the detection rate is close when λ and b are large, and the result is also consistent with the analysis in the section III. In addition, the false alarm rates are similar, too. Because of the different ratio of the gray and dot regions, the small difference of the false alarm rates between b=200 and b=100 comes from the second and third terms of (16).

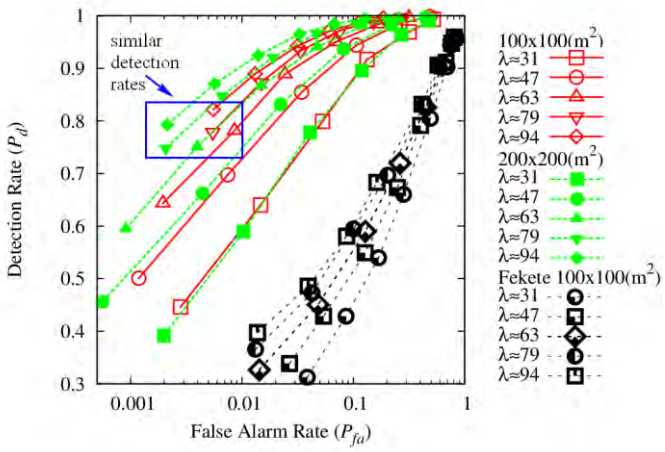

Fig. 6. The ROCs of the square ROI

We evaluate the impact of different ratios of gray, dot and white regions in the topology of a network. The topology of the simulations is shown in Fig. 7, and the square of grid lines with width d is removed in the field. The topology is similar to the form of a character C. The values of d are 40 (m) and 80 (m) in the 100×100 (m^2) and 200×200 (m^2) respectively.

In order to keep the node density in the C shape topology, the N is varied from 840 to 2520 in 100×100 (m^2) and N is varied from 5040 to 10080 in 200×200 (m^2). The ROC curves of the C shape topology are shown in the Fig. 8. The thresholds in the square and the thresholds in the C shape are set as the same. The average detection rate and false alarm rate are calculated by averaging values from 100 different topology simulations, too.

The detection rate in the square field is little higher than the detection rate in the C shape topology, because two different networks have different ratio of the boundary regions. However, the similar detection rates still occur in the C shape topology when λ and b is large. The result is congruent with the analysis section III again. Let $P_{fa,c}(100, T, N)$ denote the false alarm rate with specific parameters in the C shape topology and $P_{fa,s}(100, T, N)$ denote the false alarm rate with specific parameters in the square field. Since the ratio of gray regions of (16) in the C shape topology is greater the ratio in the square field, the difference of $P_{fa,c}(200, T, N)$ and $P_{fa,c}(100, T, N)$ is greater the difference of $P_{fa,s}(200, T, N)$ and $P_{fa,s}(100, T, N)$.

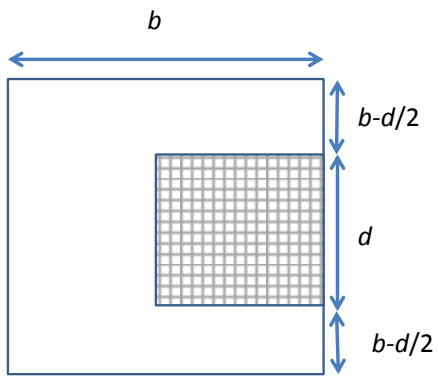

Fig. 7. The C shape topology

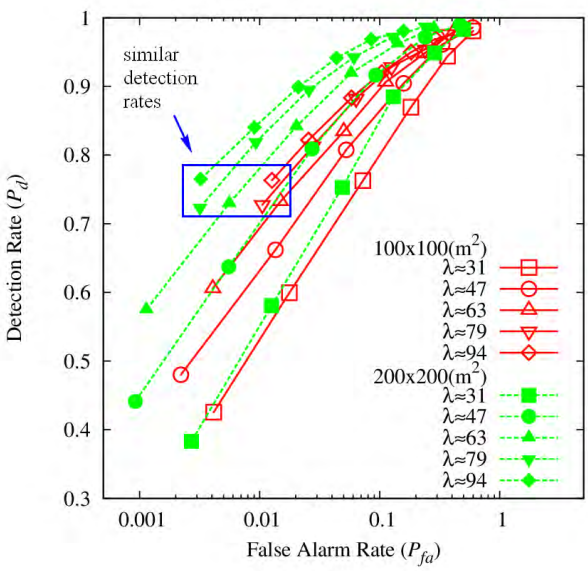

Fig. 8. The ROCs of the C shape topology

The overall results reveal that the detection rate and false alarm rate is relative to the topology. Moreover, the detection rate and false alarm rate are independent to the scale of the networks under different topology when λ and b is sufficiently large.

V. Conclusion

We propose a novel self-detection scheme for the boundary recognition by utilizing the two-hop information. With the aid of two-hop circle, the simulations reveal that the algorithm has high detection rate and low false alarm rate in the lower node density network. In the case where the topology information is available, our analyses enable the end users to choose the optimal threshold to minimize the cost function. Besides, the analysis also reveals the relations among the detection rate, false alarm rate, the topology and the scale of network size.

Our algorithm can be easily extended to utilize multi-hop circle. However, the amount of the information collected by multi-hop nodes needs to be balanced with the energy cost in collecting the multi-hop information.

Reference

[1] Q. Fang, J. Gao, and L. J. Guibas, "Locating and bypassing routing holes in sensor networks," in *Proc. INFOCOM 2004. Twenty-third AnnualJoint Conf. of the IEEE Computer and Communications Societies*, 2004, pp. 2458-2468.

[2] S. Funke and C. Klein, "Hole detection or: "how much geometry hides in connectivity?"," in *Proceedings of the twenty-second annual symposium on Computational geometry*, New York, NY, USA, 2006, pp. 377-385.

[3] I. Khan, H. Mokhtar, and M. Merabti, "A new self-detection scheme for sensor network boundary recognition," in *Proc. IEEE 34th Conf. Local Computer Networks LCN 2009*, 2009, pp. 241-244.

[4] Y. Wang, J. Gao, and J. S. B. Mitchell, "Boundary recognition in sensor networks by topological methods," in *Proceedings of the 12th annual international conference on Mobile computing and networking*, New York, NY, USA, 2006, pp. 122-133.

[5] A. Kroller, S. P. Fekete, D. Pfisterer, and S. Fischer, "Deterministic boundary recognition and topology extraction for large sensor networks," in *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, New York, NY, USA, 2006, pp. 1000-1009.

[6] O. Saukh, R. Sauter, M. Gauger, P. J. Marron, and K. Rothermel, "On Boundary Recognition without Location Information in Wireless Sensor Networks," in *Proc. Int. Conf. Information Processing in Sensor Networks IPSN '08*, 2008, pp. 207-218.

[7] Y. Liao and M. Qi, "Boundary Discovery Based on Overlapping Clustering in Non-Uniform Wireless Sensor Networks," in *Proc. Int Information Science and Applications (ICISA) Conf*, 2011, pp. 1-7.

[8] S. P. a. K. Fekete, Alexander and Pfisterer, Dennis and Fischer, Stefan and Buschmann, Carsten, "Neighborhood-Based Topology Recognition in Sensor Networks," 2004, pp. 123-136.

[9] S. P. Fekete, M. Kaufmann, A. Kröller, and K. A. Lehmann, "A New Approach for Boundary Recognition in Geometric Sensor Networks," in *CCCG*, 2005, pp. 84-87.

[10] K. Bi, K. Tu, N. Gu, W. L. Dong, and X. Liu, "Topological Hole Detection in Sensor Networks with Cooperative Neighbors," in *Proc. Int. Conf. Systems and Networks Communications ICSNC '06*, 2006.

[11] R. Ghrist and A. Muhammad, "Coverage and hole-detection in sensor networks via homology," in *Proc. Fourth Int. Symp. Information Processing in Sensor Networks IPSN 2005*, 2005, pp. 254-260.

[12] G. Destino and G. T. F. de Abreu, "Network boundary recognition via graph-theory," in *Proc. 5th Workshop Positioning, Navigation and Communication WPNC 2008*, 2008, pp. 271-275.

[13] R. Nagpal, H. Shrobe, and J. Bachrach, "Organizing a global coordinate system from local information on an ad hoc sensor network," in *Proceedings of the 2nd international conference on Information processing in sensor networks*, Berlin, Heidelberg, 2003, pp. 333-348.