International Journal of Computational Intelligence Systems, Vol. 4, No. 5 (September, 2011), 749-758

# A Projection Pursuit Based Risk Assessment Method in Mobile Ad hoc Networks

[1]Fu Cai, [1] Liu Ming*, [2]Chen Jing, [3]Zhang Li, [4]Xiao-Yang Liu

[1]*College of Computer Science & Technology, Huazhong University of Science &Technology, Wuhan, 430074,Chinas*
[2]*School of Computer, Wuhan University, WuHan City, China*
[3]*Department of Computer Science, School of Computing, National University of Singapore, Singapore*
[4]*Department of Computer Science and Engineering, Shanghai Jiao Tong University, China*
*E-mail: liuming2008@gmail.com,stand_fucai@126.com*

**Abstract**

Establishing high performance cooperation and estimating nodes' risk level in mobile ad hoc networks (MANETs) are currently fundamental and challenging due to the inherent characteristics of MANETs, such as the highly dynamic topology and the absence of an effective security mechanism. Trust based assessment methods were recently put forward but presumed restrictions to the data samples or presumed weights for node's attributes are required. In this paper, Projection Pursuit based Risk Assessment (PPRA), is proposed to analyze node's creditability. As projection pursuit turns high-dimensional node properties to low-dimension space, all nodes' risk levels could be clustered effectively and accurately. Projection index, the same as judgment index of clustering consequence, is utilized to reveal the behavior of different nodes. By maximizing projection index through Genetic Algorithm (GA), optimal projection direction is obtained, and then the projection values of each node could be calculated. Finally, the results in one-dimension or two-dimension projection space show that our method is more efficient and practical than traditional methods.

*Keywords*: project pursuit, genetic algorithm, risk assessment, projection direction, projection index.

## 1. Introduction

Mobile ad hoc networks (MANETs) are complex distributed systems that can dynamically self-organize into "ad-hoc" net work topologies with arbitrariness and temporality, which allows people and devices seamlessly net worked together in areas with no pre-deployed infrastructure Ref.1. Since nodes may not reside in physically protected places, they may fall under attackers' control. Due to the broadcasting nature of wireless channels, message eavesdropping and injection are possible. Thus, the security issue of MANETs is a difficult problem. However, classic security solutions based on certification authorities and on-line servers are inapplicable because of the absence of infrastructure. Therefore the design of a new and effective security mechanism for MANETs is quite important and necessary.

Currently a variety of assessment models are put forward, most of which reflect the risk level of communication through trust or reputation evaluation. Generally, these methods could be classified into two categories: probability-based and fuzzy theory-based.

Many works relate closely with probability-based models. Beth and his research group put forward an experience-and-probability-based trust model in Ref.2. Authors in Ref.3 adopted the beta distribution probability method to obtain the trust relationship according to a basic principle, namely the posterior probability distribution of the node action is subject to beta distribution. In Ref.4, trust was derived from evidence theory. Ref.5

* Corresponding author

and 6 applied fuzzy theory to conceptualize the subjectivity and uncertainty of trust. Current assessment models have made impressive progress. However, the complex characteristics of MANETs still deserve a lot of concern.

(i) The multiple-attribute and multiple-polarity of the sample data: when considering the trust level of a node, the communication characters (e.g., packet losing rate) and physic attributes (e.g., signal intensity) should also be included. Some attributes are better when the data size grows larger, while others are opposite. We call this multiple-polarity. Attributes change temporally which leads to a high dimension space. Current assessment models lack effective mechanisms to directly unveil the information within this high-dimension data, thus assessing nodes' trust level is challenging.

(ii) Temporal characteristic and the reliability of trust evaluation: the assessment is a dynamic and complex process. We need to process comprehensively regarding each stage, without ignoring the temporal characteristic. The stages of trust establishment, feed back and adjustment should be able to process direct and indirect information. Current evaluation systems deal with the data using probability or experience, which suffer the problem of bad expansibility. Moreover, they cannot 'sense' slight abnormal attributes which leads to unreliability.

(iii) Uncertainty of evaluation subject: this includes the needs, the experience and the knowledge of the subject. E.g., current evaluation system uses the same strategy to deal with different requirements. In general, if the subject lacks experience and knowledge, evaluation methods will lead to uncertain even contradictory results. Therefore, we need more objective methods to reduce the negative impact brought by subjective factors.

(iv) Unpredictability and diversity of attacks: due to its self-organization, dynamic routing and open wireless channels, MANETs are vulnerable to various kinds of attacks, e.g., the black hole attack and relatively covert wormhole attack. Some traditional evaluation mechanisms are based on particular distribution principle and others use functions which lack enough mathematic basis or are limited to some particular attacks. We need to seek for new evaluation methods to provide all-aspect basis.

Besides, the methods above all have some mentioned drawbacks. First, nodes with the dynamic and transitional properties, such as malicious nodes which are always self-protective, deviate the behaviors from the statistical law. Using the probability theory will reduce the accuracy of assessment. Second, a huge sample data is required when presumed distribution is applied; otherwise the result would be unconvincing. Third, the fuzzy theory requires the behavior of the node to satisfy their presumed patterns.

To solve the above problems, PPRA is proposed. It first maps multi-dimensional data into low-dimensional manifolds for visual inspection, and then clusters nodes according to their projection values. Projection index which is subjected to the projection direction, as the judgment of clustering effect, should be maximized. Projection pursuit (PP) algorithm [7-10] searches all possible projection directions in order to find the optimal projection direction that maximizes the projection index.

Due to incomplete information and no presumed pattern of nodes' behavior in the MANETs, many aspects must be considered to reveal the risk level. So a high-dimension matrix, called indicator matrix, is constructed. Through analyzing the sample data of the indicator matrix, Genetic Algorithm (GA) is employed to search for the optimal projection direction by maximizing the projection index. PPRA model is a "data-driven" model and risk level is achieved through analyzing sample data till the convergence. No presumed requirements are needed. Through experiment, PPRA model is proved to be robust, which can endure noisy of sample data to some degree.

The structure of the paper is as follows: Section 2 provides some basic knowledge about Projection Pursuit Theory. Section 3 introduces the detailed procedures of the assessment method based on PP. Section 4 demonstrates experimental analysis and comparisons with other mechanisms. Performance analysis is in Section 5 and Section 6 concludes this paper.

## 2. Projection Pursuit & Genetic Algorithm

Projection Pursuit (PP) is mainly used for analyzing high-dimension data, especially non-normal population. The main idea of PP is: project high-dimension data to low-dimension through the optimum projection direction which could best reflect the data structure and characteristic. While traditional methods treat time coordinate as probability or particular distribution function, projection pursuit unifies different attribute values at different time.

A simple genetic algorithm (SGA) is a numerical search technique used to find the exact or approximate solution for optimization problems. In SGA, a population is the abstract representation of candidate solutions. Besides, it evolves toward better solutions each step and models loosely on the principles of natural selection: employing a population of individuals that undergo selection in the presence of variation-inducing operators such as mutation and crossover.

In the following, we provide some basic concepts and definitions.

**(I) *Overall Data Dispersion***

The high-dimension attribute matrix *Y* is indicated as *Y*= $(y_1, y_2,\ldots,y_n)$. Overall data dispersion *S(a)* indicates the rate of deviation from the average:

$$S_y=\sqrt{\frac{\sum_{i=1}^{n}((Y(i)-E(Y))^2}{n-1}}. \qquad (1)$$

**(II) *Local Data Density***

*D(a)* indicates the distribution of trust evaluation values in one-dimension space.

$$D_y=\sum_{i=1}^{n}\sum_{j=1}^{n}(R-r_{ij})U(R-r_{ij}). \qquad (2)$$

Where, $r_{ij}$ indicates the distance between evaluation values. For one-dimension, $r_{ij}=|y_i-y_j|$, and for two-dimension, $r_{ij}=\sqrt{|Y_i-Y_j|^2+|Y'_i-Y'_j|^2}$. *R* is the density window width, and its selection should secure that the amount of nodes in the window are not too little and they will not grow too high with the increase of n.

**Theorem 1:** The projection density window width *R* satisfies the following equation: $r_{max}<R\leqslant p$. There are *n* projection samples, the dimension is *p* and $r_{max}$ indicates the maximum range between projections.

**Demonstration:** Based on the definition of local data density

$$D_y=\sum_{i=1}^{n}\sum_{j=1}^{n}(R-r_{i,j})U(R-r_{i,j})$$
$$=\sum_{i=1}^{n}[(R-r_{i,1})U(R-r_{i,1})+(R-r_{i,2})U(R-r_{i,2})+\ldots+(R-r_{i,n})U(R-r_{i,n})]$$
$$=(R-r_{1,1})U(R-r_{1,1})+(R-r_{1,2})U(R-r_{1,2})+\ldots+(R-r_{1,n})U(R-r_{1,n})$$
$$+(R-r_{2,1})U(R-r_{2,2})+(R-r_{2,2})U(R-r_{2,2})+\ldots+(R-r_{2,n})U(R-r_{2,n})+$$
$$\ldots+(R-r_{n,1})U(R-r_{n,2})+(R-r_{n,2})U(R-r_{n,2})+\ldots+(R-r_{n,n})U(R-r_{n,n})$$

Since R>0, and $i=j$, we have

$r_{ij}=r_{ji}=0,(R-r_{ij})*U(R-r_{ij})=R$,

then we can get

$$D_y=nR+2\sum_{i=2}^{n}(R-r_{1i})S(R-r_{1i})+2\sum_{i=2}^{n}(R-r_{2i})S(R-r_{2i})$$
$$+2\sum_{i=n-1}^{n}(R-r_{n-2,i})S(R-r_{n-2,i})+2(R-r_{n-1,n})S(R-r_{n-1,n}),$$

assume $r_{\max}=\max(r_{ji})$, only when $R>r_{\max}$, satisfying $S(R-r_{ij})=1, D_y$ can be the maximum, and the greater R, the greater $D_y$.

On the other hand, $\|X\|\le 1,\|a\|\le 1$, based on the actual physical meaning of the clustering projection to one-dimension space, $0\le a\le 1$ $Y=\alpha^T X$, so $0\le Y_i\le p$, and $r_{ij}=|Y_i-Y_j|$, so $r_{ij}\le p$.

So $r_{\max}<R\le p$ comes into existence.

**(III) *Projection Index***

Projection index is used to judge whether a projection direction is a meaningful target function. In our scheme, we use the classic projection index to optimize projection direction. That is: *Q(a)=S(a)*D(a)*, where *S(a)* indicates overall data dispersion, *D(a)* indicates partial data density, *a* indicates the projection direction.

**(IV) *Optimum Projection Direction***

Given the high-dimension node attribute matrix, the projection index function will change with projection direction. We hope to find out the optimal projection direction which can reveal ultimately the attributes' characteristic. The optimal direction can ultimately reflect the high-dimension data structure, and then we analyze the projected one-dimension data. We define projection direction searching as following:

$$Opt\ a:\ \begin{cases}\max: Q(\alpha)=S(\alpha)D(\alpha)\\ s.t.\quad \sum_{i=1}^{p}a_i=1\end{cases}. \qquad (3)$$

Thus, trust evaluation convert to this problem: solving complex nonlinear optimum by optimizing parameters. An appropriate method to search optimal projection direction is the real coded genetic algorithm (RAGA) [11-12]. Given the obtained optimal direction and the projection index, analysis of unknown nodes depends on simple matrix calculation.

**(V) *Visualized Clustering Projection***

In general projection pursuit, we usually project data to one-dimension space to quantize projection results. While in actual analysis, we need to project data to two-dimensional or three-dimensional space. System administrator can directly judge whether a node is abnormal which strengths the discernment. We call this two -

dimensional and three-dimensional projection as visualized clustering projection. Visualized clustering projection owns two impressive characters: information loss is fewer and the result is visualized.

Choose two one-dimension projection vectors, $a$, $a' \in R^p$, calculate projection values respectively, $Y=a^TX$, $Y'=a'^TX$. These two vectors can maximize the amount of information of joint distribution of $Y$ and $Y'$. The purpose of projection is to find the nonlinear structure of original data, so $Y$ and $Y'$ should be irrelevant which requires $a$ and $a'$ to be orthogonal, such that $a^Ta'=0$, $a^Ta=1$, $a'^Ta'=1$. The distance of sample projections is:

$$r'_{ij} = \sqrt{|Y_i - Y_j|^2 + |Y'_i - Y'_j|^2}.$$

In two-dimension projection, the projection density window width $R$ is a variable to be determined, from theorem 1 we have the following inference.

**Corollary 1**: There are $n$ projection samples, the dimension is $p$ and $r_{max}$ indicates the maximum range between projections. The projection density window width $R$ in two-dimension space satisfies the following equation: $r_{\max} < R \le 2\sqrt{2}p$.

**Demonstration:** From theorem 1 we can get that

$$D_y = nR + 2\sum_{i=2}^{n}(R-r_{1i})S(R-r_{1i}) + 2\sum_{i=2}^{n}(R-r_{2i})S(R-r_{2i}) + 2\sum_{i=n-1}^{n}(R-r_{n-2,i})S(R-r_{n-2,i}) + 2(R-r_{n-1,n})S(R-r_{n-1,n}),$$

assume $r_{\max} = \max(r_{ji})$, only when $R > r_{\max}$ can satisfy $S(R-r_{ij}) = 1, D_y$ can be the maximum, and the greater R, the greater $D_y$.

On the other hand, $\|X\| \le 1, \|a\| \le 1, \|a'\| \le 1$, Based on the actual physical meaning of the clustering projection, that $-1 \le a \le 1$, $0 \le a' \le 1$ and from the projection $Y = \alpha^T X$, then $-p \le Y_i \le p, -p \le Y'_i \le p$, that is: $-p \le Y_i - Y'_i \le p$, and since $r_{ij} = \sqrt{|Y_i - Y_j|^2 + |Y'_i - Y'_j|^2}$, then, $r_{ij} \le 2\sqrt{2}p$.

So $r_{\max} < R \le 2\sqrt{2}p$ comes into existence.

## 3. Projection Pursuit Model

The purpose of PPRA Model is to assess every node's risk level. PPRA can add time character which is appreciated because some attack behaviors seem well in most of the time and only through temporal analysis they can be detected. Secondly, the indicator matrix is constructed. Finally, risk level clustering is assessed by calculating every node's projection value. In this paper, one-dimension projection and two-dimension projection are presented.

### 3.1. *Elementary principle of PP model*

#### 3.1.1 *Constructing indicator matrix*

Let $S^t_{ij}$ be the property value for indicator $Y_j$ in node $X_i$ at the time $t$ $(i = 1,2,\dots,n,\ j = 1,2,\dots,m,\ t=1,2,\dots,T)$.

Then the indicator matrix, in symbol $S$, is as follows:

$$S = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \dots \\ S_n \end{bmatrix} = \begin{bmatrix} s^1_{11} & s^1_{12} & \cdots & s^1_{1m} \cdots & s^t_{1j} & \cdots & s^T_{1m} \\ s^1_{21} & s^1_{22} & \cdots & s^1_{2m} \cdots & s^t_{2j} & \cdots & s^T_{2m} \\ s^1_{31} & s^1_{32} & \cdots & s^1_{3m} \cdots & s^t_{3j} & \cdots & s^T_{3m} \\ \vdots & \vdots & & \cdots & & & \vdots \\ s^1_{n1} & s^1_{n2} & \cdots & s^1_{nm} \cdots & s^t_{nj} & \cdots & s^T_{nm} \end{bmatrix}.$$

#### 3.1.2 *Normalizing*

To eliminate this negative impact of indicators' unit, normalization of $S^t_{ij}$ is necessary. If indicator $j$ $(j=1, 2 \dots m)$ is positive-effect, then,

$$s^i_{kj} = \frac{s^i_{kj} - \min_k(s^i_{kj})}{\max_k(s^i_{kj}) - \min_k(s^i_{kj})}. \tag{4}$$

Otherwise,

$$s^i_{kj} = \frac{\max_k(s^i_{kj}) - s^i_{kj}}{\max_k(s^i_{kj}) - \min_k(s^i_{kj})}. \tag{5}$$

After normalization, indicator characteristic matrix $S$ ranges from 0 to 1. Normalization can reduce the influence of different unit of each attribute, and make our comprehensive analysis possible.

#### 3.1.3 *Constructing projection index*

Projection pursuit is to map high-dimension data onto low dimension through linear combination. Let $Z$ be the projection value as follows:

$$Z_i = f(\alpha) = \sum_{j=1}^{m}(a_j * s_{ij}). \tag{6}$$

Where, $a$ is projection direction. In order to objectively reflect character of high dimension, projection pursuit regression adopts the sum of a series of ridge function to approximate regression function. We adopt the super smooth regression as in Ref.8. The convergence condition, namely, projection index is as follows:

$$Q(a) = S(a) * D(a) \tag{7}$$

Where, $S_z$ stands for standard deviation of projection $Z$, $D_z$ stands for local density of projection $Z$. As to two-dimension projection, projection index is as follows:

$$Q(a,b) = S_z(a,b) * D_z(a,b) \qquad (8)$$

3.1.4 *Optimizing projection index by GA*

Different projection directions reflect different data structure characters. Projection index should be maximized to reach the largest difference among all nodes.

PPRA model could be described as follows:

$$\max: Q(\alpha) = S_z * D_z$$

$$\text{s.t.:} \begin{cases} \sum_{i=1}^{40} a_i^2 = 1 \\ 0 \le a_i \le 1 \quad for\ i = 1, 2 \ldots 40 \end{cases} \qquad (9)$$

Where *R* is an experiment-driven parameter, which is decided by experiment result and *r(i,j)=|z(i)-z(j)|*.

As mentioned above, GA algorithm is employed to solve the problem. Through searching possible solutions in the whole solution space, optimal projection direction *a* is obtained.

Similar to above, two-dimension projection could be described as follows:

$$\max: Q(a,b) = S_z(a,b) * D_z(a,b)$$

$$\text{s.t.}: \begin{cases} S_z(a,b) = S_z(a) * S_z(b) \\ \sum_{i=1}^{p} a_i^2 = 1, \sum_{i=1}^{p} b_i^2 = 1 \\ \sum_{i=1}^{p} a_i b_i = 0 \\ 0 \le a_i \le 1 \quad for\ i = 1, 2 \ldots p \end{cases} \qquad (10)$$

3.1.5 *Sorting the candidate nodes*

First, let's show that the relationship between node's projection value and its corresponding risk level.

**Theorem 2:** Through the steps above, the larger the projection value is, the higher the creditability is.

**Demonstration:** Let us assume one virtual node named node *V*. *V* represents the optimal node, i.e. at anytime at any attribute node V get the best attribute value. And after the temporary normalization, obviously node *V*'s attribute value is all 1.

Because

$$Z_i = f(\alpha) = \sum_{j=1}^{m} (a_j * s_{ij}).$$

And all attribute values of V node are 1, thus:

$$Z_V = f(\alpha) = \sum_{j=1}^{m} (a_j) = 1.$$

That means the optimal node's projection value is 1. So if real node's projection value is greater, the difference between it and optimal node is smaller, i.e. this node is better. This proves our conclusion is right.

Therefore, after obtaining the projection value of each node by the steps above, a final evaluation for these candidate nodes can be achieved. The node's risk level decreased with the increase of the projection value; meanwhile there still exist more similarities for specific two nodes when they're more close to each other. And the same conclusions could be inferred for two-dimension projection. The more similar to *(1, 1)*, the higher the creditability is.

### 3.2. *Procedures of PPRA model*

The procedures of PPRA are as following:

***Step1:*** selecting key indicators.

***Step2:*** normalizing ***S***, and obtaining the ***S'***.

***Step3:*** using *GA* algorithm to obtain optimal projection direction.

***Step4:*** sorting the candidate nodes.

Once the optimal projection direction is obtained, every candidate node's projection value will be calculated in optimal projection direction. Then after sorting these projection values, we can get reliable and precise results concerning nodes' risk and their similarities.

## 4. **Risk Assessment Scheme in MANETs**

Theoretical PPRA model is showed above. A real example in MANET is presented below. However, before carrying out the experiment, defining key attributes is also of significance. After fulfilling PPRA model, analysis of final result and comparison between PPRA and other methods are conducted. We conclude that PPRA model is more suitable to various network behaviors, not only for its accuracy but also its convenience as well.

### 4.1. *Defining key attributes*

The selected indicators should properly reflect the behavior characteristics of the nodes in order to make the risk assessment objective and efficient. Establishing basic principle for selection's convenience is necessary.

(I) Accessible. The indicator's value must be obtained easily. As data is sampled from real MANETs, accessibility is basic.

(II) Relative. The chosen indicators should directly reflect node's property.

(III) Complete. Every node's value should be obtained so that each node could be reflected by the corresponding indicators.

For the risk assessment of nodes in MANETs, aspects including network communication features (i.e. loss tolerance), physical attributes (i.e. mobility), wireless signal and position should all be taken into consideration. In this paper four indicators are pitched on: transmission speed (TS), losing rate (LR), signal intensity (SI), and signal changing rate (SCR). The categories of these four indicators are as follows: P represents positive-effect, whereas N represents negative-effect. And before conducting experiment, basic analysis of selected indicators is prepared.

**Table 1. Categories of selected indicators**

| TS LR | SI | SCR |
|---|---|---|
| P N | P N | |

## 4.2. *Risk assessment example*

Our experiment platform is NS 2.28. Six nodes are designed, in which one attack node and one low-efficiency node are included. Node 5 loses part of the packet temporarily and node 6 transm its packet very slow but completely. We obtained ten successive moments' node attributes for experiment. For the sake of brevity, only losing rate data is shown in Fig.1. From Fig.1, the losing rate of node 6 is very high even though it's very stable; and the value of node 5 is small but not stable, more likely to under an attack. Generally speaking, node 1 and node 2 is better than other nodes.

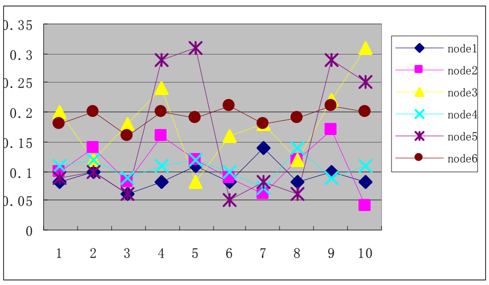


Fig.1. Package losing rate curve

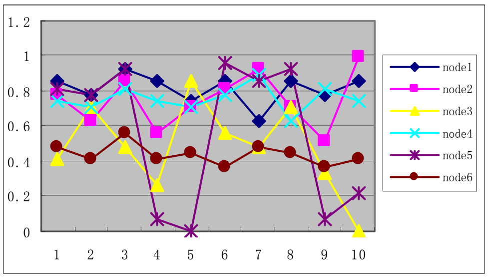


Fig.2. Package losing rate curve after normalization

**(I) Construction of indicator matrix**

From the original sample data, *S* is obtained. Here S is consisted of 6*(10*4) data.

**(II) Normalization**

Using formulas in Section 3.1 for positive and negative attributes to normalize, package losing rate normalization is show in Fig.2.

**(III) Construction of projection index**

One-dimensional projection could be described as formula (9). Two-dimension projection, as formula (10), is more complex than one-dimension projection. However, two-projection remains more information so it's more accurate.

**(IV) GA Searching**

Our platform is MATLAB 7. 9. MATLAB GA toolbox is taken in to solve the problem . To find the optimal projection mechanism, we adopt three methods: MI, project respectively according to each moment then process condensed projection; MII, one-dimension with time characteristics together; MIII, visualized two-dimension projection, shown as following.

## 4.3. *Experiment results*

MI has two processes: project at each moment and then condensed-project for previous values on the time axis. Firstly, use SGA to obtain optimal projection directions of the ten moments in the first stage shown in Fig. 3. The optimal projection result is showed in Fig. 4.

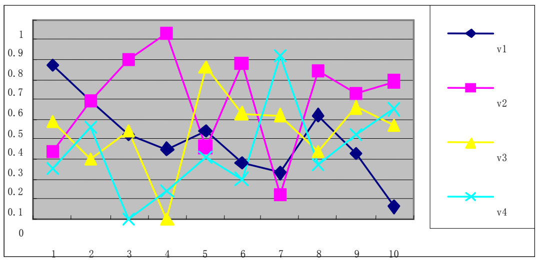


Fig.3. Optimal projection directions of the ten moments

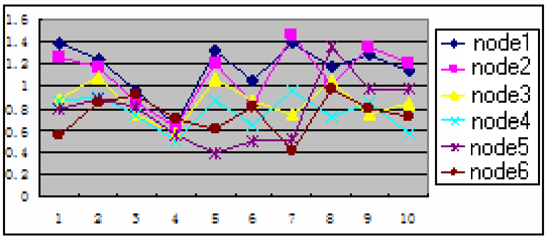


Fig.4. The optimal projection result

Next, project above results on the time axis, the final results is in Fig.5. Nodes 4 and 5 have lower values so they are more risky, PPRA concludes that they are low-efficient or having low trust level. Fig.6 shows that fitness is relatively convergent in this mechanism.

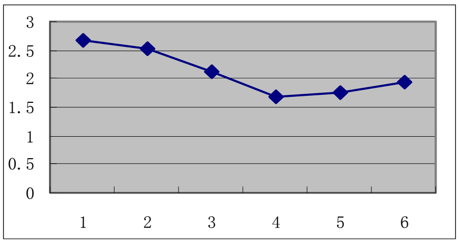

Fig.5. One-dimension projection for MI

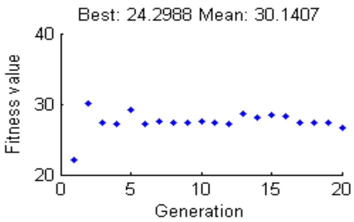


Fig.6. Convergent of one-dimension projection

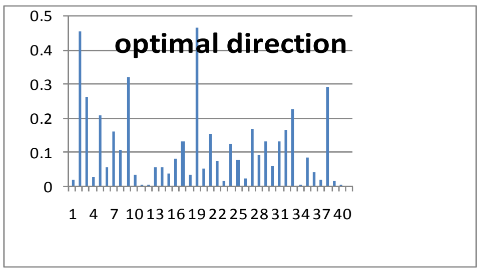


Fig.7. Optimal direction in one-dimension projection

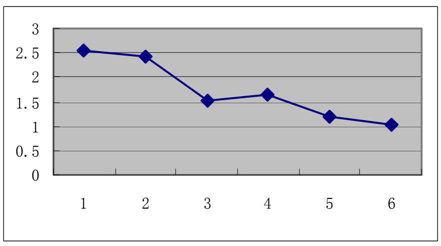

Fig.8. One-dimension projection for MII

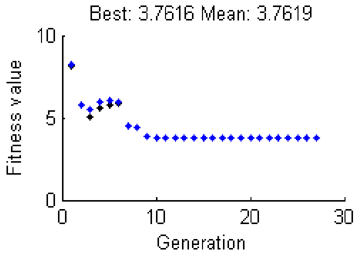


Fig.9. Convergent of one-dimension projection

In MII, with the help of GA, the optimal 40 dimension projection direction is obtained, presented in Fig.7. Given ***Z***, the risk level could be calculated using formula (6), shown in Fig.8. The first and second nodes are the best. Moreover, the distance between them is quite small, so they are thought to have similar properties. And node 5 is vulnerable so its projection value is quite small. Meanwhile node 6 is a low-efficiency node; its projection value is small too. Fig.9 shows that the fitness value is relatively convergent. From the above experiments, we can see the goals of the two mechanisms are consistent. However, MII has better projection result than MI, because MI has two stages, and during the first stage some time characteristics of attack behaviors have been lost already, which expands the gap from the reality.

Even though the result is relatively satisfying, from one-dimension projection, a plenty of information is lost. For example, in one-dimension projection both node 5 and node 6 are bad nodes and it's hard to differentiate them. Moreover, one-dimension projection has weak visibility. So introducing two-dimension projection is quite significant.

As for the two-dimension projection, the experiment result is showed in Fig.10. In Fig.10 there are four categories. Node 1 and node 2 belong to the relative good nodes, while the projection value of node 5 is quite low and so as the node 6. What's more, node 5 is quite different from node 6. The results accurately match our analysis to the original data. Fig.11 shows the convergent of two-dimension projection and it's the certification of our model's correctness.

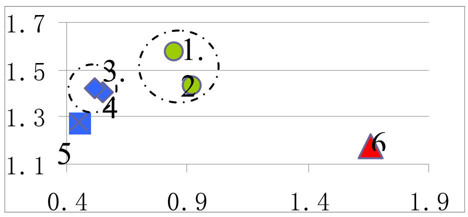

Fig.10. Two-dimension projection value

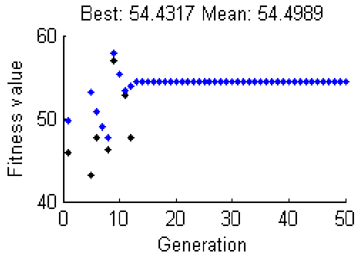


Fig.11. Convergent of two-dimension projection

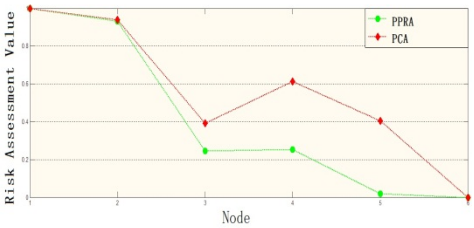


Fig.12. Compare of PP and PCA

## 5. Performance Analysis

### 5.1. Projection effectiveness

We have showed the fitness of convergence. Now we would like to explain the reason. In the analysis of PP effectiveness, we mainly depend on the convergence of SGA algorithm. The most difficult problem in GA is to prove the convergence process. Usually, we adopt the method of analyzing the evolution process of optimal function values, so we list the figures of convergence of optimal function values. The convergent process is obvious in these figures.

### 5.2. Data mining experience-irrelevance

With the nodes in MANET, their attributes are high-dimensional and variable, from this aspect, dynamic trust evaluation based on projection pursuit can be regarded as a kind of data mining based on high-dimensional and variable data. Through the establishment and use of PP model, we can project any multiple-dimension data which represent multiple attributes of nodes in MANET to one-dimension space, so as to obtain the only trust attribute and finish the trust evaluation. The data in the evaluation process comes from multiple-attribute and dynamic values based on nodes behaviors.

The drawback of traditional mechanism involves the massive attributes they need and the large data must fit some particular distributions such as beta and Bayes. While for our model, high-dimension dynamic data can be disposed before projection, so that we do not need massive data for analysis. Even better, a little information can finish the evaluation and we do not need the attributes to accord with any experience probability distribution. Our model is applicable for any high-dimension dynamic data and has pretty good experience-irrelevance. Traditional evaluation need to preprocess the data, while in our model, even without pretreatment, suitable projection index and projection direction can ensure the accuracy of the result. Therefore, RRPA has little requirements for data and owns good robustness.

### 5.3. Accuracy of Risk Assessment

The quantization value from PPRA can appropriately reflects trust information of nodes, especially the cluster status from visualized clustering projection, so that we can judge whether the nodes suffer from the grey hole attack or DoS attack and so on. Our model does not depend on experience probability distribution and has more accuracy. Traditional evaluation mechanisms choose a few principal components as evaluation index or process comprehensive judgment. Previously, the risk value has been defined which actually damage the original amount of information. While our model bases on the original data so there is no impair or influence for the analysis. The result is certainly more accurate.

To illustrate the advantage, we provide an example to compare our model with Principal Component Analysis (PCA)

In view of the above 6 nodes, 4 attributes in 10 different moments of distribution, use PCA to cluster for evaluation then obtain the following 6 evaluation values:

$$Y^{PCA}=[Y_1^{PCA},Y_2^{PCA},Y_3^{PCA},Y_4^{PCA},Y_5^{PCA},Y_6^{PCA}]$$
$$=[1.0000,0.9404,0.3916,0.6129,0.4041,0]$$

Compare this to Fig.12, we find that Node 4 is more risky than Node 5 in PCA, while in actual original data, Node 5 , as a suspicious node, its risk value should near to Node 6'. This result indicates that there exists erroneous judgment of PCA algorithm. PCA is based on mathematic model and its expandability is very limited, especially in dealing with the high-dimension data. Comparably, projection pursuit can search the optimal direction to reduce dimensions. This kind of method considers the inner connection among data so the information loss is very little. We can conclude from the figures that the evaluation value from PPRA is much better than that from PCA. PCA may even reduce some effective information and rely only on a few principal components.

Our model ultimately saves the original information. From above results, we can find out that PCA ignores the dangerous node 5 which conflicts the fact.

### 5.4. Time complexity of PPRA

The analysis of time complexity should be considered from two aspects: one is without optimal direction; the other is with the optimal direction.

If we do not have optimal direction, GA is necessary. The time complexity of PPRA mainly includes *Q(a)*.

For one-dimension and two-dimension projection, the total dispersion complexity of projection data is *O(nm),* n indicates the amount of nodes and m indicates the dimensions of vector space

$$O(\mathrm{s(a)})=\begin{cases} O(n*m) \quad one-\dim ension \\ O(n*(O(d(y_a)+d(y_b))))=O(n*m)\, two-\dim ension \end{cases}.$$

The calculation method of local density is:

$O(\mathrm{D(a)})=\mathrm{n}*\mathrm{n}*O(\mathrm{r}_{ik})=O(n^2m)$.

Moreover, the complexity of GA is proportional to the amount of projection directions of initial subgroup (k), so that the time complexity of GA is:

$O(\mathrm{Q(a)})=O(\mathrm{S(a)}+\mathrm{D(a)})*\mathrm{k}=O(n^2mk)$.

About the space complexity, because a few subgroups are produced which include projection directions, the complexity is: *O(wmk)*, *w* indicates the objective projection dimensions (in this paper, *w=1 or w=2*). And the space of original data is *O(nmT),T* is the time point for sampling (here, *T=10*).

If we have obtained the projection index and optimal projection direction for a special kind of attack through training with massive sample data, the evaluation just needs some simple matrix calculations to obtain the nodes' trust level. Under such circumstances, time complexity is *O(n)*. E.g., for the grey hole attack, we define projection index as *Q(a)=Sz*Dz* and the optimal projection direction *a*. We adopt the obtained optimal direction to calculate the projection values of any node: *a(node 1), a(node 2), …, a(node n),* by analyzing the trust level, so we can decide whether or not to include some nodes in the routing of self-organize networks.

### 5.5. Analysis for realization mechanism

According to the data from stimulation under the gray hole attack, we obtained the sample of 6 nodes and 4 relevant attributes in 10 different moments. If the projection index is given, we have different projection methods to get the optimal projection direction for evaluating node risk.

(a) Firstly reduce the dimensions of the four attributes of each node in every time point. For each time point, we establish a 6*4 matrix, including all attributes of the six nodes. After projection, four attributes become a single comprehensive value. Apply this method to each node, and then we obtain six comprehensive values. Synthesizing ten moments, use the reducing-dimensions process again, attributes in ten moments are projected to one-dimensional space. This can produce an unique trust value for each node.

(b) There are forty relevant data for each node, reduce dimensions at the same time, projecting these forty data to one-dimensional space. It deals with a 6*40 matrix directly. Applying the same projection index, we obtain the projection index through PP, which reflects the characteristics of all attributes in the whole time domain.

(c) Project high-dimension data to two-dimensional space, as formula (10). MI adopts one-dimension projection and projects ten moments respectively. This can lead to the loss of information that expends the gap with fact. MII adopts direct one-dimension projection and combined time characteristics with samples. The result is relatively accurate after optimizing projection direction. MIII not only considers the time characteristics but also reduces the loss of information. So two-dimension projection is more accurate than one-dimension, but the calculation is more complex. Therefore, when calculating the risk values, we usually adopt MII, however, if the data is extraordinarily suspicious, we can combine with MIII to process visualized risk evaluation in order to dig out abnormal nodes which may exert attacks.

### 5.6. Comparison with other methods

PPRA model has advantages in node risk assessment for MANETs, as following:

First, compared with reputation-based and probability-based models, PPRA model is "data-driven" without presumed requirement to sample data. Even the sample data is noisy, PPRA model could be accurate as both the projection index and direction are well chosen. While in Ref.3, the sample data must obey the beta distribution, so in the case that the distribution cannot be satisfied, the assessment result won't be accurate.

Second, PPRA is a visual risk assessment method. Risk level is not simply calculated, but clustered and visualized. We turn the risk calculation into risk level clustering and the final results prove our idea is practical and reliable.

Third, PPRA model could be applied to node behavior's analysis. Only a little calculation would be required if optimal projection index and direction under special condition are found. As we known, the behavior of node in MANETs is quite complex, but behind the complexity there is simplicity. Although the data is huge, corresponding projection index and direction aiming at certain node's specific behavior can be found by supervised learning; and analysis of its behaviors will be proper after quite easy calculation. For example, for the gray hole attack, if the behavior's projection direction is obtained, judgment concerning whether the node is under the gray hole attack or not can be made on the basis of corresponding projection values. Moreover, only simple matrix calculation is needed.

## 6. Conclusion

In this paper, a new model is proposed to deal with the following problems for nodes' risk assessment:

(1) Indicators are high dimensional and multi-polar;

(2) Node's behavior is complex and caprice. Using a comprehensive method to analyze is challenging.

The merits of our proposed scheme could be summarized as follows:

(1) There is no restriction for sample data.

(2) PPRA model is simple and does not need complex calculation. How to define more suitable projection index and projection direction for node behavior's analysis remains to be solved in the future.

(3) Our risk assessment is more accurate than the traditional methods for the reason that the optimal projection direction in PPRA is a comprehensive and effective way to measure nodes' multiple-attributes.

The final result shows the PPRA model is effective and suitable for risk assessment in mobile ad hoc network.

## 7. Further Work

Our model is elementary but useful. In our future work specific network behaviors will be analyzed, such as gray hole attack, how to construct corresponding projection index and what's the optimum direction to discriminate nodes under various conditions is of significance.


## Acknowledgment

This paper is supported by the Provincial Natural Science Foundation of China Hubei (2008CDB352) and the National Natural Science Foundation of China (60903175 and 60903196).


## References


1. I. Chlamtac, M. Conti, and J.J.N. Liu. Mobile ad hoc networking: imperatives and challenges. Ad Hoc Network, vol .1, pp. 13–64, 2003.
2. Beth T., Borcherding M., Klein B. Valuation of trust in open networks. In Proceeding of European Symposium on Research in Security, pp.3–18, 1994.
3. S. Ganeriwal, L.K. Balzano, M.B. Srivastava. Reputation-based framework for high integrity sensor networks. ACM Transactions on Sensor Networks, vol.4 , pp.1–5, 2008.
4. Ye AYong, Ma JianFeng, A trust valuation model in MANET. Journal of Computer Research and Development (Chinese), 45(5)(2008), pp.765–770.
5. Z. Qin, Z. Jia , X. Chen. Fuzzy dynamic programming based trusted routing decision in mobile Ad Hoc Networks. Embedded Computing, SEC '08 . Fifth IEEE International Symposium, pp.180–185.
6. Tang W., Chen Z. Research of subjective trust management model based on the fuzzy set theory. Chinese Journal of Software, pp.1401–1408, 2003.
7. J.H. Friedman and J.W.Tukey. A projection pursuit algorithm for exploratory data analysis. IEEE Transcaction on Computers, September, 1973.
8. Liu M, Gong H. A distributed energy-efficient data gathering and aggregation protocol for wireless sensor networks. Chinese Journal of Software, vol.16, 2106–2116, 2005.
9. Heinzelman W.R. An application-specific protocol architecture for wireless microsensor networks. IEEE Transc. on Wireless Communications, vol.1, pp.660–670, 2002.
10. J. Hwang, S. Lay, M. Maechler, R.D. Martin. Regression modeling in back-propagation and projection pursuit learning. IEEE Transaction on Neural Networks, 1994.
11. F. Qiang, Zhao X. Project pursuit theory and application. Chinese Science Press, pp.6:12–28, 2006.
12. Devinder Kaur , Praneeth Nelapati. Performance enhancement of data classification using selectively cloned genetic algorithm for neural network. International Journal of Computational Intelligence Systems,3(6)(2010), pp. 723-732.