SCIENCE CHINA

Information Sciences

Bayesian cognitive trust model based self-clustering algorithm for MANETs

Wei WANG and GuoSun ZENG

Citation: SCIENCE CHINA Information Sciences **53**, 494 (2010); doi: 10.1007/s11432-010-0050-8

View online: http://engine.scichina.com/doi/10.1007/s11432-010-0050-8

View Table of Contents: http://engine.scichina.com/publisher/scp/journal/SCIS/53/3

Published by the Science China Press

SCIENCE CHINA
Information Sciences

• RESEARCH PAPERS •

March 2010 Vol. 53 No. 3: 494–505
doi: 10.1007/s11432-010-0050-8

Bayesian cognitive trust model based self-clustering algorithm for MANETs

WANG Wei[1,2,3]* & ZENG GuoSun[1,2,3]

[1]*Department of Computer Science & Technology, Tongji University, Shanghai* 200092*, China;*
[2]*Tongji Branch, National Engineering & Technology Center of High Performance Computer,*
Shanghai 200092*, China*
[3]*Key Loboratory of Embedded System and Service Computing, Ministry of Education, Shanghai* 200092*, China*

Received July 7, 2009; accepted January 9, 2010; published online March 1, 2010

Abstract With the introduction of mobile Ad hoc networks (MANETs), nodes are able to participate in a dynamic network which lacks an underlying infrastructure. Before two nodes agree to interact, they must trust that each will satisfy the security and privacy requirements of the other. In this paper, using the cognition inspired method from the brain informatics (BI), we present a novel approach to improving the search efficiency and scalability of MANETs by clustering nodes based on cognitive trust mechanism. The trust relationship is formed by evaluating the level of trust using Bayesian statistic analysis, and clusters can be formed and maintained autonomously by nodes with only partial knowledge. Simulation experiments show that each node can form and join proper clusters, which improve the interaction performance of the entire network. The essence of the underlying reason is analyzed through the theory of complex networks, revealing great scalability of this method.

Keywords trust model, MANET, self-clustering, Bayesian method, cognitive mobile

Citation Wang W, Zeng G S. Bayesian cognitive trust model based self-clustering algorithm for MANETs. Sci China Inf Sci, 2010, 53: 494–505, doi: 10.1007/s11432-010-0050-8

1 Introduction

A mobile Ad hoc network (MANET) is a self-configuring network in which nodes rely on other nodes for communications. Security poses some fundamental challenges in such networks as they are not conductive to centralized trusted authorities. Although several secure routing protocols [1, 2] have been proposed, they are vulnerable to new and dynamically changing attacks. This has led to the development of several trust models [3, 4], in which mobile nodes capture evidence of trustworthiness of other nodes to quantify and represent their behaviors, and then to establish trust relationships between them.

However, there are two crucial issues of current trust based systems. One is the inefficient query routing, which makes these systems not highly scalable. For example, in [3], queries will be flooded throughout the whole network, and the search traffic increases dramatically with the increasing network size. The other is how to select appropriate nodes to cooperate with, since the candidate nodes are autonomous and may be unreliable or dishonest. These raise the question as to how much credence should be given

*Corresponding author (email: willtongji@gmail.com) to each mobile resource. Without a good solution, MANET systems are not likely to be deployed for serious applications.

On the other hand, brain informatics (BI) has recently emerged as an interdisciplinary research field that focuses on the mechanisms underlying the human information processing system (HIPS). It investigates the essential functions of the brain, ranging from perception to thinking, and encompassing such areas as multi-perception, attention, memory, language, computation, heuristic search, reasoning, planning, decision-making, problem-solving, learning, discovery and creativity [5].

Inspired by Bayesian cognitive model, we propose a novel Bayesian method based cognitive trust model. Here we show how it facilitates mobile nodes to represent and manage behavior evidence in their trust relationships with other nodes. In the proposed method, each node has the ability to form such a cluster and reconstruct the topology of mobile networks. The proposed mechanism has the following features:

- Clusters can be formed only by nodes' local knowledge, which makes our algorithm suitable for completely distributed MANETs.
- The cost of cluster calculation and maintenance is very low, making it suitable for the highly dynamic characteristic of MANETs.
- A comprehensive Bayesian trust model is proposed to evaluate the trust degree of nodes.
- A mobile node is more likely to interact with other trustworthy ones, thus minimizing the impact of malicious nodes on the performance of MANETs.

The rest of this paper is organized as follows. We review some related work in section 2. Section 3 introduces the proposed clustering algorithm. The evaluation of our approach by simulations is given in section 4 and finally we conclude this paper in section 5.

2 Related work

2.1 Bayesian models of cognition

Bayesian models are becoming increasingly prominent across a broad spectrum of the cognitive sciences. Just in the last few years, using Bayesian models people have addressed such problems as animal learning, human inductive learning and generalization, visual scene perception, motor control, semantic memory, language processing and acquisition, symbolic reasoning, causal learning and inference, and social cognition, among other topics [6, 7]. Behind these different research programs is the most compelling computational question that we can ask about the human mind. Bayesian models give us ways to approach deep questions of human cognition that have not been previously amenable to rigorous formal study. The Bayesian framework for probabilistic inference provides a general approach to understanding how problems of induction can be solved in principle, and perhaps how they might be solved in the human mind.

2.2 Trust in MANETs

On the other hand, trust management systems [3] enable a trustor to reduce uncertainty in its future interactions with a trustee, who is beyond the control of trustor. In other words, trust is a subjective probability which enables the trustor to take a binary decision by balancing between the known risks and the opinion held for trustee. This is possible as nodes are allowed to share their opinions in the MANET network, and typically it is achieved by disseminating recommendations.

Liu et al. [8] proposed a trust model by monitoring the behavior of neighbors and recommendations received from them. Pirzada et al. [9] proposed a similar approach for establishing trusted routes in dynamic source routing (DSR) protocol [10]. Jiang et al. [11] proposed an ant based evidence distribution approach which uses a swarm intelligence mechanism. Virendra et al. [12] proposed a trust model to establish group keys in MANET using trust relationships that exist among nodes.

There is a trend to complement the basic MANET overlay, irrespective of its structure, with additional connections to friends in network [13]. These additional friendly links can be used to increase the degree of the system availability. However, in order not to endanger the scalability of MANETs, it is crucial to cluster nodes with limited and local knowledge of the system on each mobile node. Moreover, we still need to address the major issue of how to decide the trust degree of each node. We cannot expect each node to know the trustworthiness of the others.

3 Self-clustering of MANET

3.1 Basic ideas

Our model concentrates on enhancing the security of network layer. In this paper, we address only the reactive routing protocols because of their ability to discover routes on demand. We choose the Ad-hoc on-demand distance vector (AODV) protocol [14] to present the details of our model.

Trust is the core of relationships in social networks, which is the evaluation of certain entities' reliable behaviors. The trust degree of a certain entity is always decided by others' recommendations. So the entity can evaluate the copartner through its behavior. Entities can also exchange and transmit evaluation massages in order to obtain the trust of target entity and guide its cooperation decision. In this paper, we define the "trust" in MANETs environment as: the evaluation of the target mobile node's ability of providing service (resource) through the reliability shown by its behavior in certain context, including the observation of its former behaviors and the recommendations from other nodes [15].

Kleinbergs [16] demonstrated that people with only local knowledge of the network were quite successful at constructing acquaintance chains of short length, leading to "small world" networks.

Based on this idea, the community structure of overlay networks can be described as follows: groups of nodes (called "cluster" in this paper) have a high density of connections within them but with a lower density of connections between them.

To form clusters, we construct a trust relation graph $G = (N, A)$, where N is the set of nodes, and edge $\{(i, j) \in A | i, j \in N\}$. If successful transaction has occurred more than one time between nodes i and j, we say that there is a trust relationship between i and j. The weight of edge is the level of trust between them. This trust relation graph is a directed graph, which means if i trusts j, j may not trust i. In Figure 1, node S interacts with other nodes, such as a_3, b_1, c_2, d_1. Based on the result of the transaction, node S can evaluate each node's level of trust, and then construct trust-based shortcut between them.

3.2 Clustering by max-flow min-cut method

The proposed method is based on graph theory. The trust relation graph consists of nodes and edges that represent their trust-based connections. We regard node communities as the dense part in the node graph, but the graph structures of the dense subgraphs are different from each other.

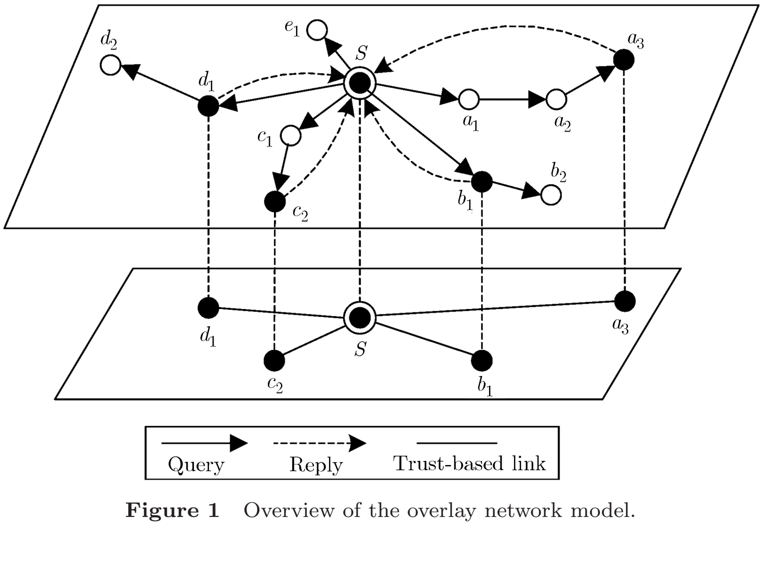

Figure 1 Overview of the overlay network model.

http://engine.scichina.com/doi/10.1007/s11432-010-0050-8

The maximum flow algorithm was first used to extract web community from the web. Kumar et al. [17] estimated the number of web communities in the web by counting the complete bipartite graphs. This kind of algorithms is promising in MANETs due to the properties of the overlay mentioned above. In addition, as far as we know, this is the first work of using max-flow min-cut theorem in clustering nodes in distributed MANETs. Other similar algorithms are HITS [18] and PageRank [19] which are used widely in information retrieval.

A node community graph can be obtained by separating a subgraph from the overlay network using maximum flow algorithm. The basic concept is that a community is recognized as a collection of nodes such that each member node has more links within the community than without the community. The maximum flow algorithm was first designed to solve the $s-t$ maximum flow problem defined by Ford and Fulkerson [20]. In light of this, we apply the max-flow min-cut theorem to cluster MANETs. We set a threshold value k of link degree at each node. When the threshold is reached, the node will begin to run the clustering algorithm to create new clusters.

3.3 Clustering based on Bayesian cognitive trust model

It is crucial to evaluate the trust degree of the nodes in order to make this method work well. In this paper, we present a Bayesian cognitive trust model. We try to find an important feature of trust within MANETs, that is, the successful cooperation probability between two nodes, and try to estimate it, as the Bayesian method supports statistical evidence for trust analysis. The main innovations are listed below:

- The problems of evaluating direct recommendation trust degree based on Bayesian method are theoretically analyzed.
- Factors, such as time and degree of belief of recommendation information, are comprehensively considered.
- The confidence level of trust degree is evaluated by using interval estimation.
- A unified paradigm of trust evaluation is proposed based on trust relationship between nodes.

For simplicity, we only consider a system within the same context during a period of time. The successful cooperation probability between two nodes x and y is denoted by θ. There may be direct interactions between them, and there may also be other intermediate nodes and each of them has direct experiences with x and y. On one hand, if there are direct interactions between x and y, we can obtain direct probability of successful cooperation, which is called direct trust degree, and denoted by θ_{dt}. On the other hand, if there is an intermediate node z between x and y, and there are interactions between x and z, z and y, then, we can also obtain an indirect probability of successful cooperation between x and y, which is called recommendation trust degree, and denoted by θ_{rt}. So, there are two kinds of probabilities of successful cooperation, which can be integrated into a global successful cooperation probability as follows:

$$\widehat{\theta} = f(\widehat{\theta}_{dt}, \widehat{\theta}_{rt}), \tag{1}$$

where $f(\cdot)$ is trust degree combination function, satisfying the property of convex function. Let $S \subset \mathbb{R}^n$ be a nonempty convex set, and let f be a function defined on S. f is a convex function on S if for every θ_{dt}, $\theta_{rt} \in S$, $\lambda \in (0, 1)$, we have

$$f(\lambda \cdot \theta_{dt} + (1-\lambda) \cdot \theta_{rt}) \leqslant \lambda f(\theta_{dt}) + (1-\lambda) f(\theta_{rt}), \tag{2}$$

$f\ (\cdot)$ is decided by the subject factors of x, such as personality and emotion. For example, a common trust degree combination function is $\hat{\theta} = \lambda\theta_{dt} + (1-\lambda)\theta_{rt}$, $\lambda \in (0, 1)$, and a node will choose $\lambda > 0.5$ if it trusts more its direct experiences than others' recommendations.

In light of this, we analyze how to obtain these two kinds of trust degree by Bayesian method.

3.3.1 *Direct trust degree*

Proposition 1. Let x and y be two nodes in MANETs, and let their interaction results be described by binomial events. When there are n times interactions between them, u times successful cooperation,

http://engine.scichina.com/doi/10.1007/s11432-010-0050-8

and v times failure cooperation, we define $\hat{\theta}_{dt}$ as the probability successful cooperation at n+1 times. Then, the posterior distribution of successful cooperation between x and y is a Beta distribution with the density function:

$$Beta(\theta|u,v)=\frac{\Gamma(u+v+2)}{\Gamma(u+1)\Gamma(v+1)}\theta^{u}(1-\theta)^{v}, \tag{3}$$

and

$$\widehat{\theta}_{dt}=E(Beta(\theta|u+1,v+1))=\frac{u+1}{u+v+2}, \tag{4}$$

where, $0<\theta<1$, and u, $v>0$.

According to Proposition 1, direct trust degree is related with the probability of successful service provider of the target node and the number of total interactions. It reflects the ability of reliable service a target node provides in the network.

Although formula (4) gives the method of computing direct trust degree, there are still two problems. First, a node may not have interacted with other nodes before, so we cannot measure the trust degree of it. Second, a node may have few interactions with the target nodes, which is not enough for us to perform trust evaluation. Under both situations, due to the lack of evidence (observations), it is not feasible to use $\hat{\theta}_{dt}$ as the trustworthiness of nodes. We need to estimate the confidence value of $\hat{\theta}_{dt}$. In fact, the measure of "reliability" about these intermediates is required. We evaluate the confidence level of trust degree by interval estimation. Let $(\hat{\theta}_{dt}-\varepsilon,\hat{\theta}_{dt}+\ \varepsilon)$ be the confidence interval with degree γ of $\hat{\theta}_{dt}$, where ε is the error level [21]. Confidence degree of $\hat{\theta}_{dt}$ can be modeled as

$$\gamma=P(\widehat{\theta}_{dt}-\varepsilon<\theta_{dt}<\widehat{\theta}_{dt}+\varepsilon)=\frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}\int_{\widehat{\theta}_{dt}+\varepsilon}^{\widehat{\theta}_{dt}-\varepsilon}\theta^{u-1}(1-\theta)^{v-1}\mathrm{d}\theta. \tag{5}$$

The confidence degree and accuracy of interval estimation are two tradeoff factors. When the number of interactions is fixed, they cannot be improved simultaneously. So, according to the rules of Neyman proposed in [21]: consider confidence degree first, and then improve the accuracy as much as possible in this condition. We can select a threshold of confidence level γ_0, and then improve the accuracy by increasing the number of interactions. When the accuracy is at an acceptable level, that is, $\gamma\geqslant\gamma_0$, the trust degree can be evaluated with the samples (evidence) at this time. Increasing the number of samples is collecting target node's interactions with other nodes. This kind of trust evaluation is the recommendation trust degree mentioned above. Based on [21], the relationship between number of samples n_0 and ε and γ_0 can be modeled as follows:

$$n_0\geqslant-\frac{1}{2\varepsilon^2}\ln\left(\frac{1-\gamma_0}{2}\right). \tag{6}$$

3.3.2 *Recommendation trust degree*

We also use the approach above to evaluate recommendation trust, as the recommendation is formed by several direct interactions. The selection of recommended nodes can also be decided by their trust degree.

Proposition 2. Let the interactions between x and y, z and y be independent of each other, and let the number of interactions between them be n_1 and n_2 respectively, with successful cooperations u_1 and u_2, and failure cooperations v_1 and v_2. Then the trust degree of x to y by z can be modeled as

$$\widehat{\theta}_{rt}=E(Beta(\theta|u_1+u_2+1,v_1+v_2+1))=\frac{u_1+u_2+1}{n_1+n_2+2}. \tag{7}$$

When there are several recommendation nodes, it is easy to extend formula (7), and combine it with the accuracy analysis above. We can get

$$\widehat{\theta}_{rt}=\frac{\sum_{\gamma\geqslant\gamma_0}u+1}{\sum_{\gamma\geqslant\gamma_0}(u+v)+2}. \tag{8}$$

http://engine.scichina.com/doi/10.1007/s11432-010-0050-8

In formula (8), the following assumption is given: a node can always get the interaction history of other nodes by searching the whole network, and increase the number of samples. When the condition of $\gamma \geqslant \gamma_0$ is satisfied, the searching can be stopped and the trust degree is evaluated by formula (8). Considering the confident level of trust degree, we can define the confidence of the recommendation y to x as the real number of interactions to the total required numbers between them.

$$w_{xy} = \begin{cases} \dfrac{n_{xy}}{n_0}, & \text{if } n_{xy} < n_0, \\ 1, & \text{otherwise.} \end{cases} \tag{9}$$

Considering that the global trust degree is affected by both positive and negative feedbacks, the value of $\hat{\theta}_{rt}$ can be mapped onto $[-1, 1]$. So formula (8) can be modified into

$$\widehat{\theta}_{rt} = \frac{\sum w \cdot (u - v)}{\sum w \cdot (u + v) + 2}. \tag{10}$$

In formula (10), it is not necessary to search the whole network to get the number of interactions n_0. For example, a node can only query the related information by asking its neighbor nodes. This is promising in deducing the communication efficiency of the network.

3.3.3 *Effect of time factor on trust evaluation*

Besides, like [24], we also consider the factor of time in our model. As the trust degree is also affected by time, the impact of time is varying to the trust degree. The more recent the history information is, the more impact the factor has. We introduce a decay factor to reflect the importance of the history information, which decreases with time. When it decreases to a certain level, it should be discarded. The concept of segment of time is used here, which can be a minute, an hour, a day, a month, or even a year. In practical applications, using day as the unit is reasonable. It can not only reflect the change of trust degree with time, but also enable the computation to perform efficiently. The interaction of nodes is composed of a series of time sequences. Given a certain sequence i, the number of the successful and failure interactions are denoted by u_i and v_i respectively. Then we have

$$u(n) = \sum_{i=1}^{n} u_i \cdot \eta^{(n-i)}, \qquad v(n) = \sum_{i=1}^{n} v_i \cdot \eta^{(n-i)}, \tag{11}$$

where $u(n)$ and $v(n)$ are the numbers of successful and failure interactions after nth sequence, and $0 \leqslant \eta \leqslant 1$. When η =1, nothing is affected by history interactions, and the whole record is aggregated; when η= 0, the latest history record is considered. A problem concerning formula (11) is that the whole history interactions have to be recorded. We can solve it by proposing the following recursive algorithm:

$$u(i) = u(i-1) \cdot \eta + u_i, \qquad v(i) = v(i-1) \cdot \eta + v_i, \tag{12}$$

where $u(i)$ and $v(i)$ are the numbers of successful and failure interactions at the ith sequence. The direct and recommended trust degree at time i can be evaluated by formulas (4) and (8).

3.3.4 *Analysis of trust relationship between nodes*

The relationships between two nodes, x and y, can be classified into four categories according to whether there are direct interactions or recommendations between them. Let dt= 1 (or 0) indicate whether there are interactions between x and y or not, and let rt= 1 (or 0) indicate whether there are intermediate nodes between them or not. Then, the four kinds of relationships can be described as $TR(dt, rt)$. We analyze the evaluation of trust degree in their relationships one by one.

1) $TR(dt, rt) = (0, 0)$. It means there is neither recommendation nor interaction between x and y. So, we should select uniform distribution, the no-information prior distribution, to be prior distribution. So, the estimator of total trust value is $\hat{\theta} = 1/2$.

2) $TR(dt, rt) = (1, 0)$. It means there is only direct interaction between x and y, and given the threshold γ_0, if $\gamma \geqslant \gamma_0$, the estimator of successful cooperation probability can be evaluated according to formula (4), or trust value $\hat{\theta}$=1/2.

Table 1 Evaluation of the trust degree in four categories

$TR(dt, rt)$	γ	$\hat{\theta}_{dt}$	$\hat{\theta}_{rt}$	$\hat{\theta}$
(0, 0)	–	1/2	0	1/2
(1, 0)	$\gamma \geqslant \gamma_0$	$\frac{u+1}{u+v+2}$	0	$\hat{\theta}_{dt}$
	$\gamma < \gamma_0$	1/2	0	1/2
(0, 1)	–	1/2	$\frac{\sum w\cdot(u-v)}{\sum w\cdot(u+v)+2}$	$f(\widehat{\theta}_{dt}, \widehat{\theta}_{rt})$
(1, 1)	$\gamma \geqslant \gamma_0$	$\frac{u+1}{u+v+2}$	$\frac{\sum w\cdot(u-v)}{\sum w\cdot(u+v)+2}$	$f(\widehat{\theta}_{dt}, \widehat{\theta}_{rt})$
	$\gamma < \gamma_0$	1/2	$\frac{\sum w\cdot(u-v)}{\sum w\cdot(u+v)+2}$	$f(\widehat{\theta}_{dt}, \widehat{\theta}_{rt})$

3) $TR(dt, rt) = (0, 1)$. It means there are only recommendation nodes between x and y. So direct trust value can still be 1/2, and recommendation trust value can be computed by formula (8). With formula (1), the total trust value can be computed.

4) $TR(dt, rt) = (1, 1)$. It means there are both recommendation and interaction between x and y. When $\gamma < \gamma_0$, the direct experience is not reliable. This situation degrades to 3), and more interaction records have to be collected. If $\gamma \geqslant \gamma_0$, the total trust value can be evaluated by formula (1), in which $\hat{\theta}_{dt}$ and $\hat{\theta}_{rt}$ can be computed by formulas (4) and (8), respectively. We sum up the above discussion in Table 1.

3.3.5 *Dynamic clustering algorithm*

Then, we propose the following dynamic clustering algorithm for depicting the change in the networks dynamically.

Let $\hat{\theta}$ donate the trust value ranging from x to y, N_x be the neighbor agents of x, and M_x be agents set which x trusts. After x updates the trust value of a target agent y, if $y \in M_x - N_x$, and z has the minimum trust value in N_x with $\hat{\theta} > \hat{\theta}_z$, then x exchanges the link to y with z; if $y \in N_x$, and z has the maximum trust value in $M_x - N_x$ with $\hat{\theta} < \hat{\theta}_z$, then x exchanges the link to z with y. The whole algorithm can be described in Figure 2.

This algorithm is very simple yet powerful, and it can enable parts of the nodes to provide better service. The phenomenon of "small world" [22] and "power law" [23] emerges in our proposed trust overlay networks. It has the following advantages: decreased average path length of the overlay, and increased convergence degree of the centre nodes at the same time.

4 Simulation results and discussions

In this section, we evaluate the proposed clustering algorithms by simulations.

```
Algorithm 1:
updateTrust ($y$) {
  if $|N| < k$ then //$k$ is the max number of node's neighbor connectTo ($y$);
    else
    if $y \in (M_x - N_x)$ then
        $z$ = chooseAgentWithMinTrustFrom ($N_x$);
        if ($\hat{\theta} > \hat{\theta}_z$) then
              linkReplaceWith ($z$, $y$);
    else
        $z$= chooseAgentWithMaxTrustFrom ($M_x - N_x$);
              if ($\hat{\theta} < \hat{\theta}_z$,) then
              linkReplaceWith ($y$, $z$);
}
```

Figure 2 Algorithm of connection update based on trust.

Table 2 Parameters for simulation

Simulation parameters	Parameter value
Total simulation time	1000 s
Max velocity ($V_{\max}$)	15 m/s
Max pause time	10 s
Simulation area	1000×1000 m^2
Packet rate	5 packets/s

4.1 Simulation setup

NS2 is invoked for our simulations. It uses random waypoint model for mobility, and fixes the transmission range for a node at 250 m. Constant bit rate (CBR) traffic with a data rate of 1.5 Mbps and packet size of 512 bytes is selected. Remaining parameters are summarized in Table 2.

4.2 The effect of trust model

We choose the trustworthy combination function as a simple linear function: $\hat{\theta} = \lambda\theta_{dt} + (1 - \lambda)\theta_{rt}$, $\lambda \in (0, 1)$, and discuss the effect of trustworthy evaluation by λ. We set the initial direct trustworthy of node x at 0.5, and then re-evaluate its trustworthiness by using feedbacks (recommendation) of other nodes. λ is set at 0.0, 0.5 or 1.0. Figure 3 shows that when λ= 1.0, the average ratio of successful query execution approaches to 1 quickly, which fully reflects the effect of recommendation information; when λ= 0.0, the recommendation has no effect on the trustworthy evaluation of node x. Therefore the average ratio of successful execution is always equal to 0.5.

Then, we consider the effect of decay factor η on task execution. We set λ at 1.0 and divide time into 20 sequences. In the first ten sequences, we recommend target nodes positive evaluation every time, that is, $(u, v) = (1, 0)$; in the last ten sequences, we recommend target nodes negative evaluation, that is, $(u, v) = (0, 1)$. η is set at 0.0, 0.5 or 1.0. The result is shown in Figure 4, from which we can learn that when $n \leqslant 10$, the average ratio of successful execution declines with different degree. The smaller the decay factor is, the more quickly the trust degree reaches a stable level. This indicates that the proposed trust model is adaptable and can objectively reflect the behavior of the nodes.

Next, based on the experiments above, we analyze the performance of the MANETs with this model. Here, λ and η are both set at 0.8.

4.3 Performance of the MANET overlay

First, we examine the clustering effect of the proposed method. We set the network size at 100 nodes.

Figure 5 shows the cover rate which is the percentage of nodes belong to at least one cluster, and Figure 6 shows the number of clusters to which each node belongs after issuing 5000 queries, which is considered as a stable state of the MANET. Most nodes only belong to a few clusters while a minority of nodes belongs to a large number of clusters. This means that most nodes are more likely to connect with most trusted ones. We can see that most nodes can join a cluster after sufficient queries have been processed in the MANET and many nodes can find a cluster within 3 or 4 queries, which proves the efficiency of our clustering algorithm.

4.4 Property of complex networks

The phenomenon of "small world" [22] and "power law" [23] emerges in our proposed trust MANETs overlay. Average path length and clustering coefficient are the two key characteristics of the complex networks. The reason why power law phenomenon emerges is that there are some preferences ("trust" in this paper) between nodes when they cooperate with each other. In our opinion, the clustering algorithm is based on trust mechanism, so the phenomenon of large in-degree nodes can be avoided, which shows the ability of self-adaptation.

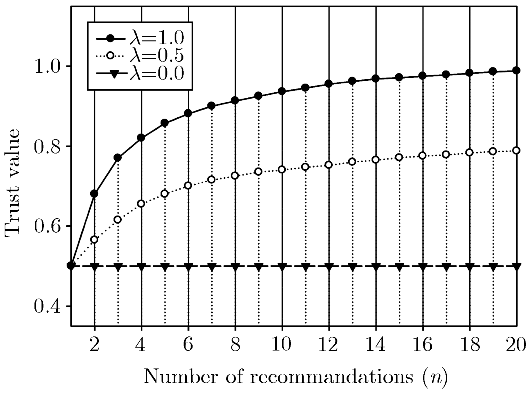

Figure 3 The effect of varying λ on trustworthiness.

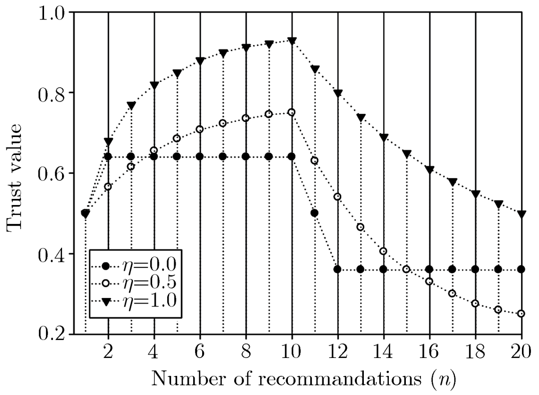

Figure 4 The effect of varying η on trustworthiness.

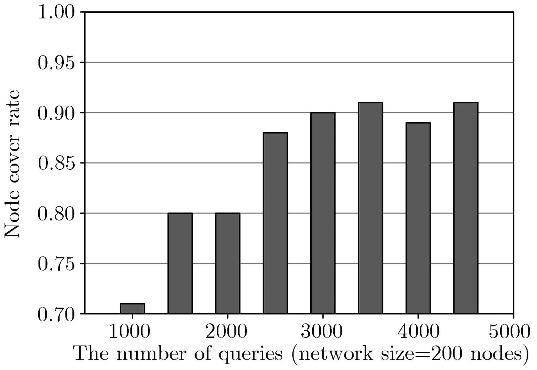

Figure 5 Node cover rate of clusters.

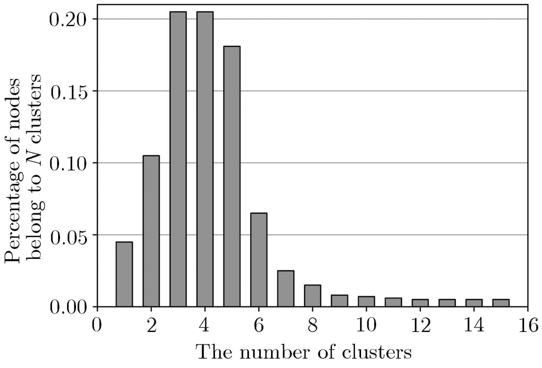

Figure 6 Distribution of the number of nodes in different number of clusters.

Besides the small world phenomenon, we also observed that the proposed trust overlay has the middle clustering coefficient, which is different from the high clustering coefficient in random networks and low clustering coefficient in classic complex networks. And we compared the clustering coefficient values in three conditions: random connection between two nodes with the probability $p = 0.5$, trust-based connections without nodes' ability constraint and trust-based connection with node's ability constraint. We first give the following definitions:

Definition 1. In graph G, a node's neighbors can be defined as $N_u = \{v|d(u, v) = 1\}$, where v refers to arbitrary nodes in G except u.

Definition 2. The clustering coefficient of node u is $\phi_u = \frac{|E(N_u)|}{C_k^2}$, where $|E(N_u)|$ is the number of the edges in subgraph formed by u's neighbors.

Definition 3. The clustering coefficient of graph G is $\phi_u = \frac{\sum_{u \in G} \phi_u}{|G|}$, where $|G|$ is the number of the nodes.

The clustering coefficient value varies according to the change of network size (200 to 1000). The results are shown in Table 3.

We found that the clustering coefficient under the schema of trust-based connection without constraint is higher. But in trust-based connection with nodes' constraint, the clustering coefficient decreased dramatically. So, based on this algorithm, the in-degree of a trusted node is sure to decrease, reaching at a stable level of the whole system.

The main goal of the next experiment is to examine the result of successful service providing rate when malicious nodes exist. A malicious node tries to assign positive ratings to any other malicious node in the network.

Table 3 Comparison of clustering coefficients

Network size (n)	Clustering coefficient		
	Random connection	Trust-based connection (with constraint)	Trust-based connection (without constraint)
200	0.584	0.755	0.894
400	0.578	0.765	0.904
600	0.590	0.778	0.912
800	0.608	0.768	0.894
1000	0.611	0.770	0.901

Table 4 Successful service providing rate with different malicious node numbers

Number of interaction	Successful service providing rate (%)		
	No malicious nodes	40% malicious nodes	80% malicious nodes
50	71.8	71.5	28.3
100	85.6	73.1	15.6
150	83.5	78.8	19.9
200	85.9	79.4	18.1
250	82.0	78.9	20.9

Table 4 shows that even when the malicious nodes ratio is about 40%, the successful service provide rate is still 80%. But when there are 80% malicious nodes, the successful rate decreases dramatically. These experiments show that our proposed clustering algorithm is suitable for large scale and dynamic MANETs with trust assurance and it is promising to use it in the real world.

5 Conclusions

We have proposed a new self-clustering algorithm to improve the security performance and scalability of MANETs with trust assurance. By the proposed algorithm, each node is capable of forming clusters by only local knowledge, which makes our algorithm suitable for distributed autonomous MANETs. The simulation results show that each node can form and join proper clusters based on their trust degree, and the cluster-based search with trust assurance performance outperforms those in current popular trust models.

Acknowledgements

This work was supported by the National High-Tech Research & Development Program of China (Grant Nos. 2007AA01Z425, 2009AA012201), the National Basic Research Program of China (Grant No. 2007CB316502), the National Natural Science Foundation of China (Grant Nos. 90718015, 70903051), and the joint of NSFC and Microsoft Asia Research (Grant No. 60970155), the Ph.D. Programs Foundation of Ministry of Education (Grant No. 20090072110035), the Program of Shanghai Subject Chief Scientist (Grant No. 10XD1404400), the State Key Laboratory of High-end Server & Storage Technology (Grant No. 2009HSSA06), Program for Young Excellent Talents in Tongji University (Grant Nos. 0800219105, 2009KJ030). We are grateful to all reviewers for their comments. The first author would like to thank Wang Li for her good advice and insights that shaped this paper.

References

1 Hu Y C, Perrig A, Johnson D B. Ariadne: a secure on-demand routing protocol for Ad hoc networks. In: International Conference on Mobile Computing and Networking. Atlanta, USA, 2002. 12–23

2 Papadimitratos P, Haas Z J. Secure routing for mobile Ad hoc networks. In: SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS 2002), San Antonio, TX, USA, January 27-31, 2002. 193–204

3 Bansal S, Baker M. Observation-based cooperation enforcement in Ad hoc networks. Technical Report. Stanford University, 2003

4 Eschenauer L, Gligor V D, Baras J. On trust establishment in mobile Ad-hoc networks. In: 10th International Security Protocols Workshop. Cambridge, UK, 2004. 47–66

5 Griffiths T L, Kemp C, Tenenbaum J B. Bayesian models of cognition. In: The Cambridge Handbook of Computational Cognitive Modeling. Cambridge: Cambridge University Press, 2008

6 Courville A C, Daw N D, Touretzky D S. Bayesian theories of conditioning in a changing world. Trends Cogn Sci, 2006, (10): 294–300

7 Tenenbaum J B, Griffiths T L, Kemp C. Theory-based Bayesian models of inductive learning and reasoning. Trends Cogn Sci, 2006, (10): 309–318

8 Liu Z, Joy A W, Thompson R A. A dynamic trust model for mobile Ad hoc networks. In: 10th IEEE International Workshop on Future Trends of Distributed Computing Systems (FTDCS 2004). Suzhou, China, 2004. 80–85

9 Pirzada A A, Datta A, McDonald C. Propagating trust in Ad-hoc networks for reliable routing. In: IEEE International Workshop on Wireless Ad-Hoc Networks. Oulu, Finland, 2004. 58–62

10 Johnson D B, Maltz D A, Broch J. DSR: The dynamic source routing protocol for multihop wireless Ad hoc networks. In: Perkins C E, ed. Ad hoc Networking. Boston: Addison-Wesley Longman Publishing Co., Inc., 2001. 139–172

11 Jiang T, Baras J S. Ant-based adaptive trust evidence distribution in MANET. In: 24th International Conference on Distributed Computing Systems Workshops (ICDCSW 2004). Tokyo, Japan, 2004. 588–593

12 Virendra M, Jadliwala M, Chandrasekaran M, et al. Quantifying trust in mobile Ad-hoc networks. In: International Conference Integration of Knowledge Intensive Multi-Agent Systems (KIMAS 2005): Modeling, Evolution and Engineering. Massachusetts, USA, 2005. 65–70

13 Naor M, Wieder U. Know my neighbor's: better routing for skip-graphs and small worlds. In: Proceedings of the Third International Workshop on Peer-to-Peer Systems, IPTPS 2004, San Diego, CA, USA, 2004. 269–277

14 Perkins C E, Royer E M, Das S R, et al. Performance comparison of two on-demand routing protocols for Ad hoc networks. IEEE Pers Commun, 2001, 8: 16–28

15 Tempich C, Staab S, Wranik A. REMINDIN': Semantic query routing in peer-to-peer networks based on social metaphors. In: 13th WWW Conference. New York: ACM, 2004. 640–649

16 Kleinberg J. Navigation in a small world. Nature, 2000, 406: 845

17 Kumar R, Raghavan P, Rajagopalan S, et al. Trawling the web for emerging cyber-communities, In: Philip H E, ed. Proceedings of the 7th Internation Cenference on World Wide Web. New York: Elsevier Horth-Holland, Inc., 1999. 1481–1493

18 Agosti M, Pretto L. A theoretical study of a generalized version of Kleinberg's HITS algorithm. Inform Retrieval, 2005, 8: 219–243

19 Page L, Brin S. The PageRank citation ranking: bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998

20 Ford J, Fulkerson D R. Maximal flow through a network. Canadian J Math, 1956, 8: 399–404

21 Thomas L, John S J. Bayesian Methods: an Analysis for Statisticians and Interdisciplinary. Cambridge: Cambridge University Press, 1999

22 Duncan W, Steven S. Collective dynamics of 'small world' networks. Nature, 1998, 393: 440–442

23 Michalis F, Petros F, Christos F. On power-law relationship of the internet topology. Comput Commun Rev, 1999, 29: 251–262

24 Wang W, Zeng G S. Trusted dynamic level scheduling based on Bayes trust model. Sci China Ser F-Inf Sci, 2007, 50: 456–469

Appendix

A Proof of Proposition 1

Proof. Let $p = P\ (S)$ denote the probability of successful cooperation in one interaction. The prior probability of p can be a random variable in (0, 1). Given no more information about p and according to Bayesian theorem, p can be assumed to be a uniform distribution $U(0,1)$ with the prior distribution $\pi\ (p)$. When there are n times interactions, which is new information, let event $A =$ "u times successful results in n interaction". Then the result of the interaction is a binomial event, which is $P\ (A|p = p) = p^u(1-p)^{n-u}$. According to the continuous form of Bayesian theorem,

$$P(A) = \int_0^1 P(A|p = p)\pi(p)dp = \frac{(n-u)!u!}{(n+1)!}.$$

The posterior distribution function $f\ (p|A)$ reflects the information updating of event A. According to Bayesian theorem,

$$f_{p|A}(p|A) = \frac{(n+1)!}{(n-u)!u!}p^u(1-p)^{n-u} = \frac{\Gamma(u+v+2)}{\Gamma(u+1)\Gamma(v+1)}p^u(1-p)^v.$$

The posterior distribution function is not the uniform but *Beta* distribution $Beta(u+1, v+1)$.

We can use this distribution function to predict the probability in the future. Let event B = "know u times successful cooperation in n times interaction, and the n+1 time is also successful". Then

$$P(B) = \int_0^1 P(B|p=p)f(p|A)dp = \int_0^1 p\frac{(n+1)!}{(n-u)!u!}p^u(1-p)^{n-u}dp = \frac{u+1}{n+2}.$$

According to the properties of *Beta* distribution, the expect value of this distribution E $(Beta(\theta|u_1, v_1))$ is $(u+1)/(n+2)$.

B Proof of Proposition 2

Proof. n_1 and n_2 are independent with the same distribution. According to Proposition 1, the prior distribution of n_1 is the *Beta* distribution. When x observes the interaction results between y and z, it can update its prior information by Bayesian theorem. According to the properties of the *Beta* function, the posterior of it is still *Beta* distribution with the expected value $E(Beta(\theta|u_1+u_2+1,\ v_1+v_2+1))$. The proof detail can be found in [21]. So, formula (8) is the trust evaluation x to z.