available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Security threats scenarios in trust and reputation models for distributed systems

*Félix Gómez Mármol**, *Gregorio Martínez Pérez*

Departamento de Ingeniería de la Información y las Comunicaciones, University of Murcia, 30.071 Murcia, Spain

ARTICLE INFO

Article history:
Received 17 December 2008
Received in revised form
3 April 2009
Accepted 1 May 2009

Keywords:
Trust
Reputation
Security threats
Distributed environments
Threats taxonomy

ABSTRACT

Trust and reputation management over distributed systems has been proposed in the last few years as a novel and accurate way of dealing with some security deficiencies which are inherent to those environments. Thus, many models and theories have been developed in order to effective and accurately manage trust and reputation in those communities. Nevertheless, very few of them take into consideration all the possible security threats that can compromise the system. In this paper, we present some of the most important and critical security threats that could be applied in a trust and reputation scheme. We will describe and analyze each of those threats and propose some recommendations to face them when developing a new trust and reputation mechanism. We will also study how some trust and reputation models solve them. This work expects to be a reference guide when designing secure trust and reputation models.

1. Introduction

Trust and reputation models have been recently proposed by many researches as an innovative solution for guaranteeing a minimum level of security between two entities belonging to a distributed system that want to have a transaction or interaction.

Thus, many studies, works and models have been designed, carried out and developed in this direction, leading to a current solid research field on which both academia and industry are focusing their attention.

Many methods, technologies and mechanisms like fuzzy logic (Tajeddine et al., 2006), bayesian networks (Wang et al., 2006b) or even bio-inspired algorithms (Gómez Mármol, 2008) have been proposed in order to manage and model trust and reputation in systems such as P2P networks (Almenárez et al., 2004), ad-hoc ones (Moloney and Weber, 2005), wireless sensor networks (Boukerche et al., 2007) (WSN) or even multi-agent systems (Sabater and Sierra, 2001).

Analyzing and studying some of these models (Josang et al., 2007; Sabater and Sierra, 2005) we realized that there are some security threats directly related to this specific kind of models, which are common and applicable to most of these approaches.

Nevertheless, we also noticed that each author proposed his/her own threats when testing their developed models, revealing the lack of a commonly agreed process of checking the robustness of a trust and reputation model against the mentioned risks.

This paper presents the most important security threat scenarios that can be found in the area of trust and reputation in a distributed system where some entities request some services and other ones provide those services. As far as we know, this is one of the first research works making such a thorough analysis.

We will describe each threat and propose a possible solution for tackling it. We will additionally study how some of the most representative models deal with those threats and analyze their proposed solutions.

* *Corresponding author*. Tel.: +34 868 887866.
E-mail addresses: felixgm@um.es (F.G. Mármol), gregorio@um.es (G.M. Pérez).
0167-4048/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2009.05.005

Every accurate and robust trust and reputation model should have some mechanisms to effectively overcome all the threats that could be applied to it. Therefore, this work is intended to serve as a reference guide for developing secure trust and reputation models.

The rest of the paper is organized as follows: Section 2 makes a brief description of trust and reputation management and discusses the importance of dealing with the security threats exposed in Section 3, where some solutions are also proposed. A taxonomy of these threats is described in Section 4 and how some trust and reputation models face them is shown in Section 5. Finally, some conclusions and future work are depicted in Section 6.

2. Trust and reputation management

Trust and reputation management has recently become a very useful and powerful tool in some specific environments where a lack of previous knowledge about the system can lead participants to undesired situations, specifically in virtual communities where users do not know each other at all or, at least, do not know everyone.

It is in those cases where the application of trust and reputation mechanisms is more effective, helping a peer to find out which is the most trustworthy or reputable participant to have an interaction with, preventing thus the selection of a fraudulent or malicious one.

We have noticed that most of the current trust and reputation models in the literature follow these four general steps (Marti and Garcia-Molina, 2006) (as shown in Fig. 1):

1. Collecting information about a certain participant in the community by asking other users their opinions or recommendations about that peer.
2. Aggregating all the received information properly and somehow computing a score for every peer in the network.
3. Selecting the most trustworthy or reputable entity in the community providing a certain service and effectively having an interaction with it, assessing a posteriori the satisfaction of the user with the received service.
4. According to the satisfaction obtained, a last step of punishing or rewarding is carried out, adjusting consequently the global trust (or reputation) deposited in the selected service provider.

Additionally, each model manages concepts such as trust or reputation in many different ways. For instance, some models like PTM (Almenárez et al., 2004; Almenárez et al., 2006) or AFRAS (Carbó et al., 2003) make use of fuzzy logic in order to deal with those topics.

On the other hand, bayesian networks are used by authors of MTrust (Songsiri, 2006) and BNBTM (Wang et al., 2006b). And even bio-inspired algorithms are used in AntRep (Wang et al., 2006a) or TACS (Gómez Mármol et al., 2008). Other models like Eigentrust (Kamvar et al., 2003) or PeerTrust (Xiong and Liu, 2004) just give some analytic expressions.

However, we also realized that not all the models address all the possible threats that could be found and applied in those scenarios. In fact, some of them do not even deal with these risks at all.

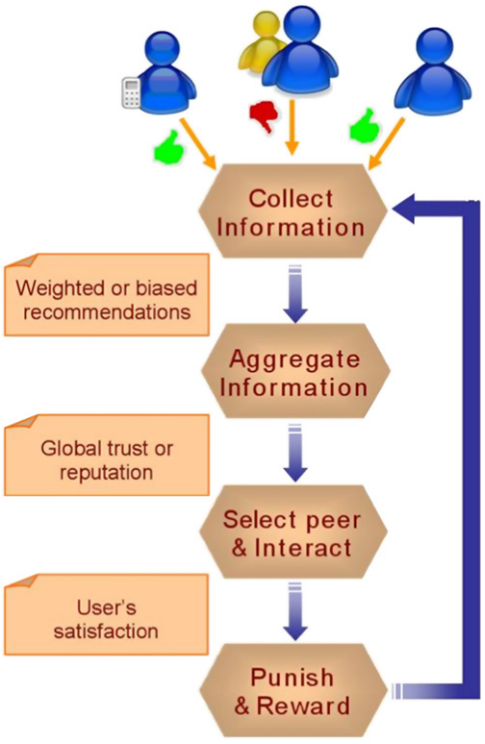

Fig. 1 – Trust and reputation models' steps.

In our opinion, this is an issue that should not be underestimated when designing and developing a new trust and reputation model over distributed and heterogeneous systems, since an inaccurate management of these threats could result in important security deficiencies and weaknesses.

It is also worth mentioning that the distinction between a trust and a reputation model is not always clear. However, in our opinion, those models making an explicit use of other participants' recommendations could be categorized as reputation models while the rest could be considered just as trust models.

Finally, some scenarios where a trust and reputation model may prove useful could be, among many others, a P2P file sharing system, an ad-hoc routing protocol or a streaming service in case of accident over a WSN.

3. Security threats

In this section we will present and describe the most common security threats applicable in the field of trust and reputation management over distributed environments. Moreover, an approach aimed to tackle and solve each of those threats will be also proposed.

It is important to note that, although all of these threats can be applied to some trust and reputation models, not all of them can be applied to any model, since some threats are specific of one or another type of trust and reputation model.

Without loss of generality we will consider a scenario where several participants (entities, nodes, peers, agents, users, …) belong to a virtual community (P2P network, WSN, ad-hoc network, multi-agent system, …) where a certain set of services is offered.

When a specific participant is requested to provide one of the services it offers it can effectively provide the offered service and act, therefore, in a benevolent way or, on the other hand, it can provide a worse service, acting thus fraudulent or maliciously.

3.1. *Individual malicious peers*

Description. Malicious peers always provide bad services when selected as service providers (Fig. 2).

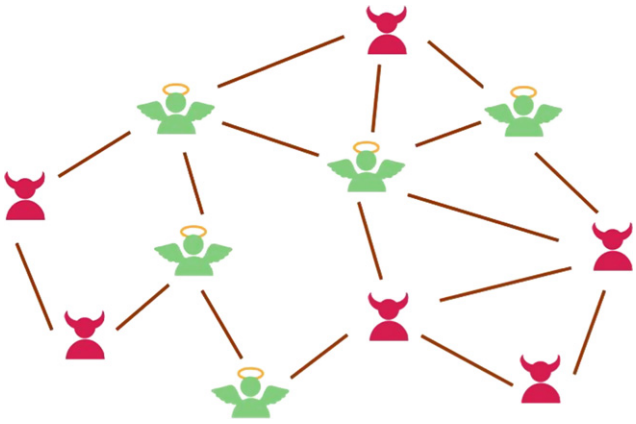

Fig. 2 – Individual malicious peers.

Discussion. This is the simplest threat that can be found in a trust and reputation system. Every trust and reputation model deals with this kind of attack.

Solution. The way of preventing such a misbehavior is by decreasing the level of trust or reputation of those participants who always provide bad services, categorizing them, therefore, as malicious peers.

3.2. *Malicious collectives*

Description. Malicious peers always provide bad services when selected as service providers. Malicious peers form a malicious collective by assigning the maximum trust value to other malicious peers in the network (Fig. 3).

Discussion. Not many trust and reputation models treat the problem arisen from the constitution of a collusion among malicious peers, having thus an important security deficiency.

Solution. The first thing needed to be able to overcome this threat is to somehow manage, not only the goodness of every user when supplying services, but also their reliability when giving recommendations about other peers. Thus, a user who provides unfair ratings will be also discarded as a service provider.

3.3. *Malicious collectives with camouflage*

Description. Malicious peers provide bad services in p% of all cases when selected as service providers. Malicious peers

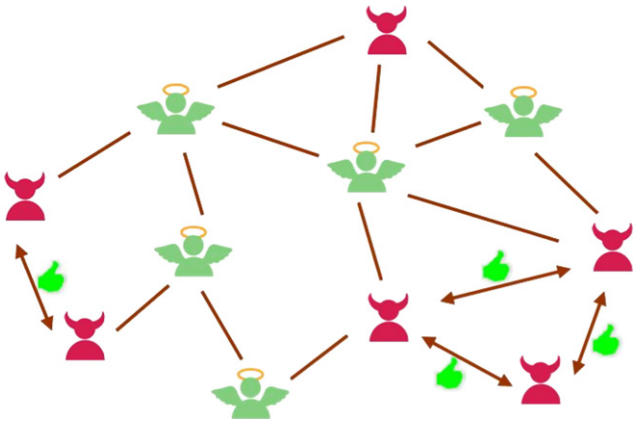

Fig. 3 – Malicious collectives.

form a malicious collective by assigning the maximum trust value to other malicious peers in the network (Fig. 4).

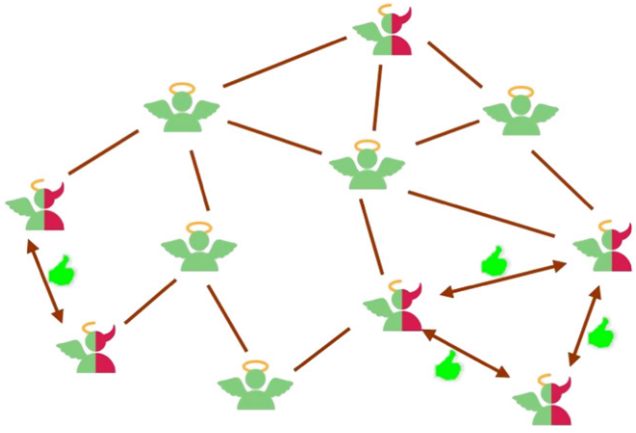

Fig. 4 – Malicious collectives with camouflage.

Discussion. This is, in many cases, a threat which is not always easy to tackle, since its resilience will mostly depend on the behavioral pattern followed by malicious peers. That is, it is not equal to battle against an oscillating pattern (being fully benevolent for a period of time, and fully fraudulent for the next period, and so on, as shown in Fig. 5(a)), for instance, than against an increasing and decreasing one (Fig. 5(b)), or even a random pattern (Fig. 5(c)).

Furthermore, the variable behavior is not even considered as a threat in many models in the sense that they do not punish that kind of behavior, but they just try to adjust the trust and reputation given to a peer to its real and current goodness. Other models (Kamvar et al., 2003), however, demonstrate the uselessness for malicious peers to behave in this way.

Solution. The first topic to address is to somehow distinguish the confidence deposited in a peer as a recommender and the trust deposited in the same peer as a service provider. This mechanism can be very helpful when trying to avoid unfair ratings from malicious entities. Additionally, the variable behavior of a peer, when detected, could be punished and avoided.

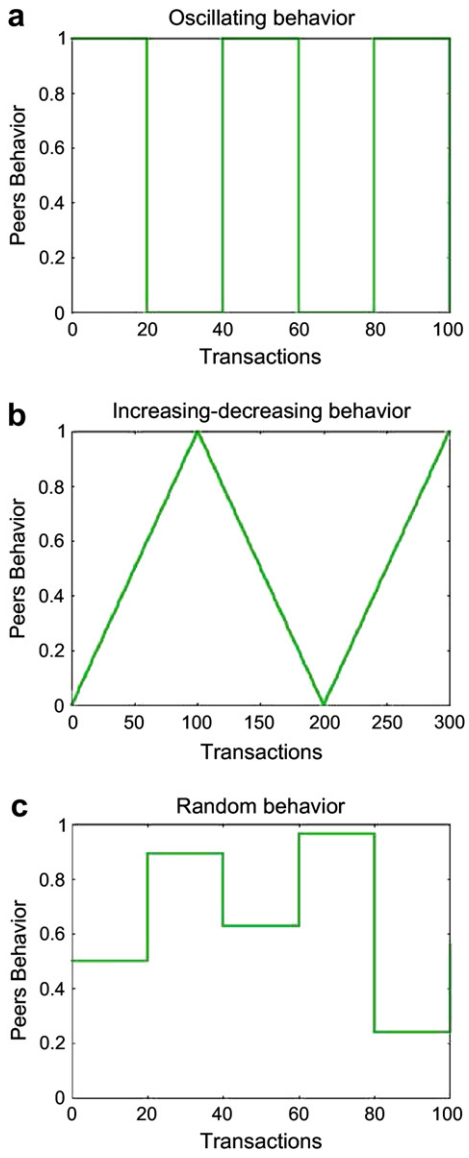

Fig. 5 – Variable behavior.

3.4. *Malicious spies*

Description. Some malicious peers always provide bad services when selected as service providers. Those malicious peers form a malicious collective by assigning the maximum trust value to other malicious peers in the network. Other distinct malicious peers, known as malicious spies, always provide good services when selected as service providers, but they also give the maximum rating values to those malicious peers who always provide bad services (Fig. 6).

Discussion. In this threat, the malicious spies may gain a high level of trust and reputation, since they always provide good services, being able then to easily subvert the trust and reputation mechanism applied in the system. Most of the times, this kind of attack has not a trivial or easy way of being effectively tackled.

Solution. Like in previous threats, an accurate management of the reliability of the peers, not only as service providers, but also as recommendation providers may effectively help to prevent this kind of abuse, although it will probably take longer (more effort and more resources needed, therefore) in order to be able to identify both the malicious peers and the malicious spies.

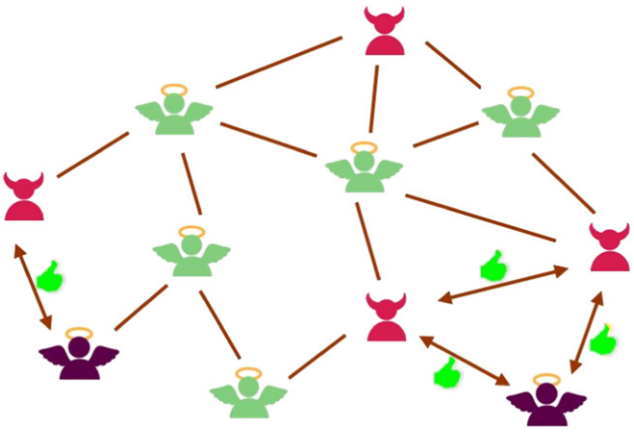

Fig. 6 – Malicious spies.

3.5. *Sybil attack*

Description. An adversary initiates a disproportionate number of malicious peers in the network. Each time one of the peers is selected as a service provider, it provides a bad service, after which it is disconnected and replaced with a new peer identity (Fig. 7) (Douceur and Donath, 2002).

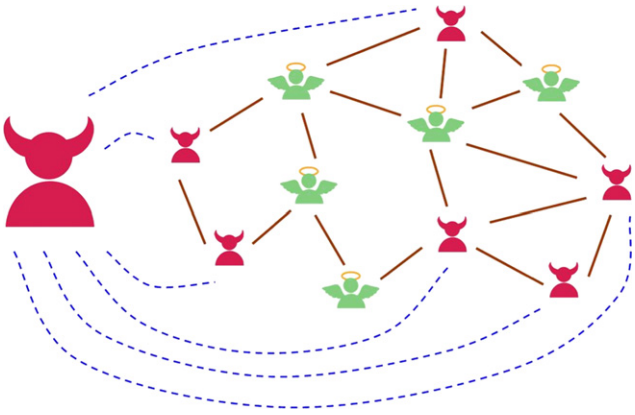

Fig. 7 – Sybil attack.

Discussion. This kind of attack might prove quite problematic since it could prevent good peers from being able to gain a good reputation, since they might not be selected most of the times.

Again, not many trust and reputation models deal with such an important and potentially dangerous threat like the Sybil attack leading thus to an underestimated but great risk.

Solution. One of the most common solutions proposed in the literature for this kind of threat consists of associating a cost to the generation of new identities in the community.

This cost is not necessarily economic, but it can also be a cost in terms of time or resources, for instance.

Another suggested way of dealing with this problem (Girao et al., 2006) makes use of a central entity managing (virtual) identities in the system, or even a set of identity providers ensuring that every participant in the community has a unique and immutable identity.

3.6. *Man in the middle attack*

Description. A malicious peer can intercept the messages from a benevolent service provider peer to the requestor and rewrite them with bad services, making therefore the reputation of the benevolent peer to decrease. That participant could even maliciously modify the recommendations given by an honest peer, in order to benefit his/her own interests (Fig. 8).

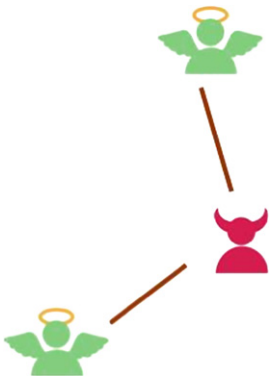

Fig. 8 – Man in the middle attack.

Discussion. One more time, this is a threat which has not been associated with trust and reputation systems traditionally. Most of the authors consider or assume the authenticity of the peer providing either a service or a recommendation. Nevertheless, as explained before, this attack can cause a great damage and effect in the system if its application is possible.

Solution. A simple way of avoiding this risk could be by the use of cryptography schemes in order to authenticate each user in the system (maybe with a digital signature or any similar mechanism). However, and unfortunately, it is not always feasible to apply such a solution, above all in highly distributed environments like wireless sensor networks.

3.7. *Driving down the reputation of a reliable peer*

Description. Malicious peers always provide bad services when selected as service providers. Malicious peers form a malicious collective by assigning the maximum trust value to other malicious peers in the network. Additionally, they give the worst rating to those benevolent peers, who indeed provide good services (Fig. 9).

Discussion. This kind of attack can be even worse than the ones named malicious collectives and malicious spies, since in this case benevolent peers also receive unfair critics from

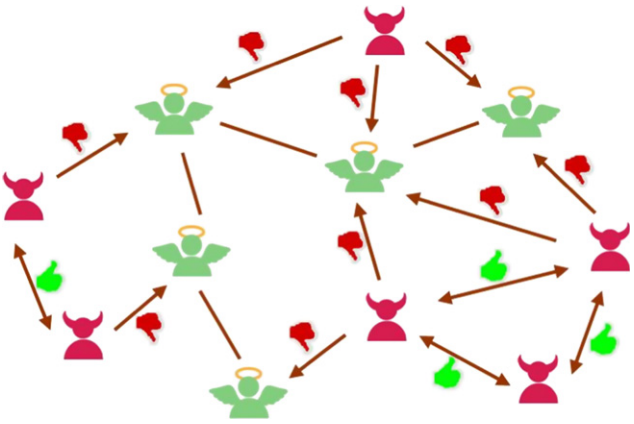

Fig. 9 – Driving down the reputation of a reliable peer.

malicious pees. In such a situation if an interaction with a malicious peer is carried out it can be identified as malicious, but if an interaction has never been performed with a peer which is actually benevolent but whose reputation has been driven down by malicious participants, then that peer will not probably be chosen as the peer to have an interaction with.

Solution. The differentiated management of the trust given to a participant when supplying services and the reliability of his/her recommendations can be very useful in this scenario as well. However, there are some trust and reputation models (Gómez Mármol, 2008) where this distinction is not explicitly done but, due to their dependency on the topology of the network, are able to find the most trustworthy path leading to the most reputable peer offering a certain service.

3.8. *Partially malicious collectives*

Description. Malicious peers always provide bad services when selected as certain service providers. However, they always provide good services when selected as other different service providers (Fig. 10).

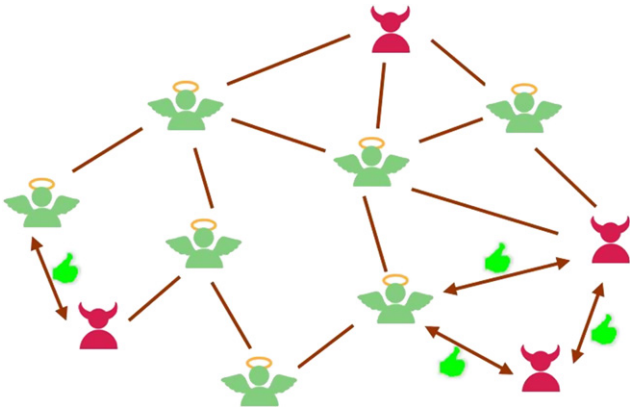

Fig. 10 – Partially malicious collectives.

That is, for certain services they behave properly, while for other specific services, they act maliciously. Malicious peers

form a malicious collective by assigning the maximum trust value to other malicious peers in the network.

Discussion. There are some trust and reputation models which are not resilient to this kind of attack since they just perform a global computation of the trust and/or reputation of a peer, regardless the service they are providing. In such a situation some distortion can emerge, considering a peer as fully or quite benevolent (malicious) although it can also provide some fraudulent (good) services.

Solution. By just considering a different score for every service offered by a peer, this threat is mitigated most of the times. However, it is not always possible to make this distinction since in some environments (for instance, those with a great amount of services offered) it could lead to some scalability problems.

3.9. *Malicious pre-trusted peers*

Description. Some or all the pre-trusted benevolent peers become malicious ones, maybe by always providing bad services when selected as service providers or by rating with maximum trust value other malicious peers who always provide bad services when selected as service providers.

Discussion. First it is worth mentioning that it is not always feasible to find a set of peers that can be trusted before any transaction is carried out in the system (Fig. 11).

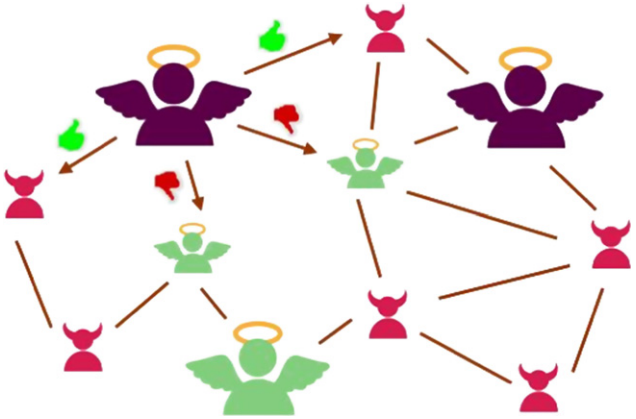

Fig. 11 – Malicious pre-trusted peers.

Some models (like Eigentrust (Kamvar et al., 2003), for instance) base their strategy on this kind of participants. However, and maybe in a paranoic way of thinking, every user in a virtual community can behave inappropriately at some point. If such a thing occurred with a pre-trusted peer, those models mentioned before would be in a risk.

Solution. Our suggestion for such situations would be to be able to decide at any time which peers belong to the set of pre-trusted ones, depending on their behavior.

4. Security threats taxonomy

This section will describe several properties or dimensions related to a generic security threat for trust and reputation systems (Lam and Riedl, 2004). These dimensions will help us to create a taxonomy of the previously exposed threats, analyzing and categorizing each one of them according to these properties. In fact, a summary table (Table 1) has been designed and included showing that classification.

- *Attack intent*

An adversary may have several different goals when trying to subvert a trust and reputation system. Two straightforward intents are to fraudulently praise an entity in order to increase her reputation in the system and, conversely, to drive down the reputation of a reliable entity.

A third possible goal could be just to damage the reputation system as a whole, so users may decrease their trust in it and, eventually, stop using it.

Thus for instance, malicious collectives, collectives with camouflage and malicious spies attacks will try to unfairly praise and increase the reputation of some entities which actually do not deserve it. The rest of threats will just try to subvert the whole system in one or another way.

- *Targets*

Some security threats focus their efforts on a subset of users or entities belonging to the system, whereas other threats center on specific individual users. There are even some threats which do not distinguish and are applicable to the whole community.

In this sense, individual malicious peers and man in the middle attacks can be classified as individual attacks, while driving down the reputation of a reliable peer affects all the members of the community. Other threats' targets are composed by a subset of the entities belonging to the system.

- *Required knowledge*

The amount of information needed to be gathered or collected from the system in order to effectively perform an attack is another important issue in these scenarios. Thus, some threats will require a comprehensive knowledge about the whole system or about some particular entities, while some other threats will work properly with a small knowledge about the trust and reputation system (its users, the trust and reputation model applied, ratings distribution, etc.).

Regarding this point, creating a collusion, for instance, will need more information about the system (each member of the collusion needs to know the rest of them) than an individual attack such as individual malicious peers or Sybil attack. If they also need to know, for example, the goodness of each member for every given provided service, then the amount of required knowledge in order to perform the attack is higher.

- *Cost*

The less expensive an attack is, the more beneficial is its application. Once again, the cost of running an attack is not necessarily economic, but it can be also measured in terms of resources or time requirements, for instance.

Table 1 – Security threats taxonomy.

Security threats	Attacks dimensions					
	Attack intent	Target	Required knowledge	Cost	Algorithm dependence	Detectability
Individual malicious peers	Whole	Individual	Low	Low	Generic	High
Malicious collectives	Praise	Subset	Medium	Medium	Generic	Medium
Malicious collectives with camouflage	Praise	Subset	Medium	Medium	Generic	Low
Malicious spies	Praise	Subset	High	High	Generic	Low
Sybil attack	Whole	Subset	Low	Medium	Generic	Low
Man in the middle attack	Whole	Individual	Medium	Medium	Generic	Medium
Driving down the reputation of a reliable peer	Whole	All	High	High	Generic	Low
Partially malicious collectives	Whole	Subset	High	High	Generic	Low
Malicious pre-trusted peers	Whole	Subset	High	High	Specific	Low

Thus, some threats will have a higher associated cost and will be therefore more difficult to be performed, while others will be easily applicable, since their corresponding cost will make them worthy.

As it can be observed in Table 1 the cost of applying an attack is directly related to its associated amount of required knowledge. The only case where both dimensions do not match is for the Sybil attack, because although it needs (nearly) no knowledge about the system, it is not usually so easy to create a disproportionate number of entities enough to cause a really important damage to the community.

- *Algorithm dependence*

Some security threats take advantage of a specific trust and reputation algorithm or model vulnerability and exploits it in order to create a great damage to the system. On the other hand, other attacks are more generic and, consequently, applicable in a wider set of scenarios or environments.

Most of the described security threats for trust and reputation system could be applied in almost any scenario or environment. Malicious pre-trusted peers, however, is an specific attack related and, therefore, only applicable to those trust and reputation algorithms or models which actually make use of pre-trusted peers, as we will see later in the case of EigenTrust (Kamvar et al., 2003).

- *Detectability*

Finally, an attack over trust and reputation systems is desired to be as less detectable as possible. Later an attack is detected, the higher might be the damage caused. That is the reason why most of the threats act trying not to induce suspicion as much as possible, i.e., they do not cause drastic changes in the system, but they rather make slight ones.

In some way, the detectability of an attack or threat is a measurement of its resilience and effectiveness. Thus, the easiest threat of the previously presented ones to be detected would be the individual malicious peers. As the collaboration between attackers and their gathered knowledge about the system increases, those attacks become more and more undetectable. That is the reason why all the threats based on a collusion are, generally, more difficult to tackle.

5. Dealing with main security threats in major trust and reputation models

This section will present some of the most representative trust and reputation models for distributed systems and will show how each of them face the threats exposed in Section 3.

Some experimental results taken from the reference papers highlight how each model is reacting against certain attacks they are covering.

5.1. *EigenTrust*

5.1.1. *Brief introduction*

The first trust model we will describe is called EigenTrust (Kamvar et al., 2003), and it is one of the most known and cited ones in this field. It is characterized by the assignment of a unique global trust value to each peer in a P2P file sharing system, based on the peer's history of contributions.

Thus, authors define s_{ij} as the local trust of peer i about peer j, in the following way:

$$s_{ij} = \mathrm{sat}(i,j) - \mathrm{unsat}(i,j)$$

i.e., the difference between the satisfactory and unsatisfactory interactions of peers i and j. Moreover, they also define a normalized local trust value $c_{ij} \in [0, 1]$ as:

$$c_{ij} = \frac{\max(s_{ij}, 0)}{\sum_j \max(s_{ij}, 0)}$$

Peer i's global reputation is given by the local trust values given to it by other peers, weighted by the global reputation of the assigning peers. Let C be the matrix $[c_{ij}]$ and $\vec{c}_i$ a vector defined as follows:

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1j} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2j} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{i1} & c_{i2} & \cdots & c_{ij} & \cdots & c_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nj} & \cdots & c_{nn} \end{pmatrix}, \quad \vec{c}_i = \begin{pmatrix} c_{i1} \\ c_{i2} \\ \vdots \\ c_{ij} \\ \vdots \\ c_{in} \end{pmatrix}$$

Having this, t_{ik} represents the trust that peer i places in peer k based on asking his friends, and defined as:

$$\vec{t}_i = C^T \vec{c}_i = \left(\sum_{j=1}^{n} c_{ij} c_{j1}, \ldots, \sum_{j=1}^{n} c_{ij} c_{jk}, \ldots, \sum_{j=1}^{n} c_{ij} c_{jn} \right)$$

By querying his friends' friends, peer i gets a wider view of peer's k reputation, that is:

$$\vec{t}_i = (C^T)^2 \vec{c}_i$$

Going on in this way, after a large enough number m of queries, peer i will get the same eigenvector $\vec{t}_i = (C^T)^m \vec{c}_i$, as every other peer in the system.

Additionally, authors propose more sophisticated ways of computing this eigenvector based on pre-trusted peers. They also consider that a peer who is honest providing a service (in their case sharing a file) is also likely to be honest in reporting its local trust values, which, as we have seen before, has not to be necessarily always like this.

5.1.2. *Security analysis*

Regarding the security threats this model covers, when a set of individual malicious peers is present in the system, those peers receive high local trust values only from other malicious peers, since they are the only ones who value the supply of malicious services (i.e., they compute $s_{ij} = \text{unsat}(i, j) - sat(i, j)$). And even that only occasionally, since malicious peers have to meet each other through an interaction. Because of their low trust values, malicious peers are rarely chosen as service providers (around 10% of the times).

Forming a malicious collective does not increase the global trust values of malicious peers enough in order for them to have impact on the network due to the presence of pre-trusted peers. A user will always have the opportunity to perform a transaction with one of those pre-trusted peers and if an interaction is performed with a malicious peer (which occurs again around 10% of the times), it will be identified as malicious by the whole system.

However, it is worthy to mention that authors do not consider a collusion exactly in the same way we defined it previously, since in their scenario, every peer belonging to the collusion gives the maximum rate to the "next" peer in the collusion (and the minimum to everybody else), forming thus a ring or chain.

Moreover, the optimum scenario for a malicious collective with camouflage in this model consists of providing 50% of the times a fraudulent service (in that case 28% of the transactions correspond to a malicious service). Kamvar et al. (2003) demonstrate the unworthiness of such behavior for malicious peers relying on the cost those peers have in order to sometimes provide a service properly.

Finally, authors also deal with the problem of Sybil attack by imposing some kind of cost to the generation of new identities, but they also show the vulnerability of their model against malicious spies, since their opinions and recommendations will be taken into account (even when rating malicious peers) due to their proper behavior when supplying services.

5.2. *PeerTrust*

5.2.1. *Brief introduction*

PeerTrust (Xiong and Liu, 2004) is a trust and reputation model that combines several important aspects related to the management of trust and reputation in distributed systems, such as: the feedback a peer receives from other peers, the total number of transactions of a peer, the credibility of the recommendations given by a peer, the transaction context factor and the community context factor.

This accurate aggregation is performed through the following expression, representing the trust value of peer u, $T(u)$:

$$T(u) = \alpha \sum_{i=1}^{I(u)} S(u,i) \text{CR}(p(u,i)) \text{TF}(u,i) + \beta \times \text{CF}(u)$$

where $I(u)$ denotes the total number of transactions performed by peer u with all other peers, $p(u, i)$ denotes the other participating peer in peer u's ith transaction, $S(u, i)$ denotes the normalized amount of satisfaction peer u receives from $p(u, i)$ in its ith transaction, $\text{CR}(v)$ denotes the credibility of the feedback submitted by v, $\text{TF}(u, i)$ denotes the adaptive transaction context factor for peer u's ith transaction, and $\text{CF}(u)$ denotes the adaptive community context factor for peer u.

On the other hand, the credibility of v from w's point of view, is computed as:

$$\text{Cr}(p(u,i)) = \frac{\text{Sim}(p(u,i), w)}{\sum_{j=1}^{I(u)} \text{Sim}(p(u,j), w)}$$

where

$$\text{Sim}(v,w) = 1 - \sqrt{\sum_{x \in \text{IJS}(v,w)} \left(\frac{\sum_{i=1}^{I(x,v)} S(x,i)}{I(x,v)} - \frac{\sum_{i=1}^{I(x,w)} S(x,i)}{I(x,w)} \right)^2 \Big/ |\text{IJS}(v,w)|}$$

being $I(u, v)$ the total number of transactions performed by peer u with peer v, $\text{IS}(v)$ the set of peers that have interacted with peer v and $\text{IJS}(v, w)$ the common set of peers that have interacted with both peer v and w, computed as $\text{IS}(v) \cap \text{IS}(w)$.

Additionally this model introduces a trust-based peer selection scheme, according to the third step described in Section 2 and depicted in Fig. 1. A simple rule for peer w to decide whether to have an interaction with peer u or not could be $T(u) > T_{\text{threshold}}(w)$, where the value of $T_{\text{threshold}}(w)$ depends on several factors such as the importance of the transaction, or the disposition of w to trust unknown peers or not, among many others.

5.2.2. *Security analysis*

The accurate management of the credibility of a peer as a recommender, as well as the context factor or the community one allows PeerTrust model to effectively overcome many of the security threats described previously.

Thus, malicious individual peers, malicious collectives, malicious collectives with camouflage and driving down the reputation of a reliable peer are some of the threats that are solved by PeerTrust.

This ability to deal with those threats is due to, among other factors, the definition of credibility in terms of the

similarity between two peers, which allows the model to accurately detect and identify in the community malicious service providers as well as malicious recommenders.

Additionally it stimulates the community to supply recommendations by building incentives or rewards to those peers who provide feedbacks to others. And this is done through the context factor, with the following definition:

$$CF(u) = \frac{F(u)}{I(u)}$$

where $F(u)$ represents the total number of feedback peer u gives to others. This stimulation also helps and is very useful to avoid (almost any kind of) malicious peers to gain a high reputation in the system and therefore, to be selected many times as service providers.

When a threat of the type malicious individual peers, malicious collectives or driving down the reputation of a reliable peer occurs, PeerTrust achieves the selection of fraudulent peers to remain less than a 10% of the times, being the worst case that one where 50% of the peers are malicious.

Regarding malicious collectives with camouflage, authors test the oscillating scenario described before (Fig. 5(a)) obtaining reasonably good outcomes due to the use of a time windows-based metric that discounts the old feedbacks of peers.

Finally, PeerTrust can also overcome the threats of partially malicious collectives (since it introduces a context factor to measure the importance of each transaction) and the man in the middle attack.

The latter is tackled making use of cryptographic mechanisms. Specifically, authors propose that every identity is established by a public key corresponding to a unique private key, avoiding thus the spoofing of an identity without the knowledge of such private key. Additionally, any content properly signed will not have its integrity or origin compromised.

5.3. BTRM-WSN

5.3.1. Brief introduction

BTRM-WSN (Gómez Mármol, 2008) is a novel trust model for wireless sensor networks (WSN) based on the bio-inspired algorithm of ant colony system (ACS, Dorigo et al., 2006). It allows to find the most trustworthy path leading to the most reputable service provider in a network. Its intrinsic nature makes it to be easily adaptable to sudden changes in the topology of the network as well as in the behavior of its participants.

In this model, a set of ants (artificial agents) is launched through the WSN. While they are searching for the most reputable service provider, they leave some pheromone traces in every link connecting two nodes. That pheromone between sensors a and b, denoted as τ_{ab}, is identified with the confidence sensor a has on finding the most trustworthy path through sensor b.

At each node, every ant has to decide which next sensor to move towards. In order to carry out this decision, a probability is given to each arc not visited yet by that ant as follows:

$$p_k(r,s) = \begin{cases} \dfrac{[\tau_{rs}]^{\alpha}[\eta_{rs}]^{\beta}}{\sum_{u \in J_k(r)} [\tau_{ru}]^{\alpha}[\eta_{ru}]^{\beta}}, & \text{if } s \in J_k(r) \\ 0, & \text{otherwise} \end{cases}$$

being $p_k(r,s)$ the probability of ant k to move from sensor r to s, η_{rs} the heuristic associated with the link joining r and s, identified with the distance that separate both sensors, $J_k(r)$ the set of neighbors of node r not visited yet by ant k, and α and β, two parameters to balance the pheromone and the heuristic.

Every time an ant crosses a link, it modifies its pheromone trace in the following way:

$$\tau_{s_1 s_2} = (1-\varphi)\tau_{s_1 s_2} + \varphi\Omega$$

where $\Omega = (1 + (1-\varphi)(1 - \tau_{s_1 s_2}\eta_{s_1 s_2}))\tau_{s_1 s_2}$ is the convergence value of $\tau_{s_1 s_2}$ and ϕ is a parameter controlling the amount of pheromone left by an ant.

In the same way, the best path found by all ants receives an additional updating, as follows:

$$\tau_{rs} = (1-\rho)\tau_{rs} + \rho(1 + \tau_{rs}\eta_{rs}Q(S_{Global\_Best}))\tau_{rs}$$

being $Q(S_{Global\_Best})$ the quality of such path. The quality of a path S_k is measured in terms of the average pheromone of the edges belonging to that path, $\overline{\tau}_k$, the percentage of ants that have selected that precise path as the most trustworthy, $\%A_k$, and its length, as it can be observed next:

$$Q(S_k) = \frac{\overline{\tau}_k}{\sqrt{\text{Length}(S_k)}}\ \%A_k$$

Furthermore, when ant k finds a peer offering the desired service, it has to decide whether to stop and return that found service provider, or to travel ahead trying to find a better (more reputable) one. In order to make that decision, the average pheromone trace of the edges composing the current path is computed, $\overline{\tau}_k$.

If $\overline{\tau}_k$ is greater than a given threshold, then ant k stops and returns current solution with a probability equal to $\overline{\tau}_k$ (which means that better paths have more probabilities to be chosen). Otherwise, if $\overline{\tau}_k$ is less than or equal to that certain threshold, ant k considers current service provider not enough trustworthy and keeps trying a better one.

As we indicated in Section 2, the last general step of every trust and reputation model consists of punishing or rewarding the selected service provider, according to the user's satisfaction. In BTRM-WSN this step is explicitly performed in terms of pheromone evaporation (punishment) or reinforcement (reward) of the path leading to the selected peer, as shown next:

$$\tau_{rs} = (\tau_{rs} - \varphi \times df_{rs})\frac{Sat}{df_{rs}}$$

where ϕ is the same parameter used in the local pheromone updating, Sat is the user's satisfaction and df_{rs} represents a distance factor of the link joining sensors r and s, which is defined as follows:

$$df_{rs} = \sqrt{\frac{d_{rs}}{L(S_k)(L(S_k) - d_{rs} + 1)}}$$

being d_{rs} the distance of link joining sensors r and s from the client and $L(S_k)$ the length of the solution found by ant k, S_k.

5.3.2. Security analysis

Regarding the performance of BTRM-WSN against certain threats, it has been demonstrated its accuracy in situations of malicious individual peers, malicious collectives, malicious

collectives with camouflage and driving down the reputation of a reliable peer.

When a peer is selected as a service provider and it supplies a worse service than the one it initially offered, not only the path leading to that server is punished (by means of pheromone evaporation), but also all the links or edges falling into that node, hindering this way other ants to choose that peer as the next hop in their route.

Malicious individual peers are, in this way, accurately identified in the community. Less than 10% of the times they are wrongly selected when the 90% of the nodes are individual malicious peers, in a WSN composed by of 100 sensors.

Due to the definition of the algorithm, where every peer only stores the pheromone traces of its neighbors, if a malicious peer forms a collusion and gives unfair ratings (in terms of pheromone traces) to its neighbors, ants are able to overcome this situation and find alternative paths (if they exist) leading to the most reputable nodes. This definition allows the resilience against a man in the middle attack, as well.

BTRM-WSN is therefore resilient in the presence of malicious collectives. In this case the selection percentage of malicious service providers (also called the error of the model) remains under the 10% regardless the size of the wireless sensor network, when the percentage of malicious peers forming a collusion is below the 60%.

Actually, the collusion threat model implemented in BTRM-WSN corresponds to the threat we called here driving down the reputation of a reliable peer, which is in fact a particular case of a collusion. So this model has been demonstrated to be able to overcome both threats.

Once again the oscillating scenario of Fig. 5(a) has been chosen in order to test the model against malicious collectives with camouflage. In this case, since there are some benevolent peers not belonging to this collusion, they gain a high trust level in the system and are, therefore, selected most of the times as service providers, obtaining similar outcomes than in the case of malicious collectives (less than 10% of error when the percentage of malicious peers is under 60%).

Partially malicious collectives are also avoided since BTRM-WSN uses different and independent pheromone traces for each service offered by the WSN.

5.4. *PowerTrust*

5.4.1. *Brief introduction*

PowerTrust (Zhou and Hwang, 2007) is a robust and scalable P2P reputation system which leverages the power-law feedback characteristics found applicable in dynamically growing P2P networks, either structured or unstructured.

Authors made several comprehensive experiments over a data set extracted from eBay transactions and concluded that the feedback numbers in eBay follow a power-law distribution. Even more, they demonstrate that power-law feedback distribution is applicable to every P2P reputation system in general.

The power-law distribution implies that the node with a few feedbacks is common, whereas the node with a large number of feedbacks is extremely rare. Therefore, only a few nodes have much higher degree than others, and specifically those nodes are dynamically selected as power nodes and considered as most reputable in the system.

Nevertheless, power nodes can be dynamically replaced if they become less active or demonstrate unacceptable behavior. Actually, the m most reputable nodes are selected using a distributed ranking mechanism which in turn applies a locality preserving hashing in order to sort all nodes with respect to their global reputation scores.

To do so, PowerTrust builds a trust overlay network (TON) on top of all nodes in a P2P system where every peer evaluates each other whenever a transaction takes place between a pair of them. Therefore, all nodes have local trust scores and the system aggregates those scores in order to calculate the global reputation score of each participating peer. All global scores form a reputation vector $V=\{v_1, v_2, \ldots, v_n\}$ fulfilling that $\sum v_i = 1$.

In order to compute vector V, consider the trust matrix $R=(r_{ij})$ defined over an n-node TON, where $r_{ij} \in [0, 1]$ is the normalized local trust score defined by $r_{ij} = s_{ij}/\sum_j s_{ij}$ (with $\sum_j r_{ij} = 1$), and s_{ij} is the most recent feedback score that node i rates node j. Next an initial reputation vector $V_{(0)}$ is set assuming, for instance, $v_i = 1/n$. And while $|V_{(t)} - V_{(t-1)}| > \epsilon$ the successive reputation vectors are recursively computed as:

$$V_{(t+1)} = R^T \times V_{(t)}$$

After a sufficient number of k iterations, the global reputation vector will converge to the eigenvector of the trust matrix R. Finally, this global reputation scores updating is carried out by power nodes.

5.4.2. *Security analysis*

The use of reliable power peers as global reputation scores updaters makes PowerTrust a resilient model against a wide variety of security threats. Specifically, authors demonstrate the robustness and accuracy of their approach through a set of developed experiments.

Thus, PowerTrust has been proved to be resistant against an individual malicious peers attack, achieving good outcomes in presence of this type of adversaries (less than a 35% of error).

Even more, since authors consider that a node providing corrupted services is highly likely to issue dishonest scores, PowerTrust is also resilient (with experiments supporting this fact) against malicious collectives, malicious collectives with camouflage and driving down the reputation of a reliable peer.

Nevertheless, it is vulnerable to a malicious pre-trusted peers threat, because in this model, power nodes are considered as fully reliable peers (as pre-trusted peers are in EigenTrust). So if those power peers become malicious, they can cause a great damage in the system.

5.5. *Tackling summary*

In this section, we present a summary table (Table 2) indicating for each one of the described trust and reputation models which threats can be overcome, which not and which are just not applicable. In order to make a more complete table, we have also included some models (ATSN (Chen et al., 2007) and DWTrust (Huang et al., 2006)) not described in this paper.

As it can be observed, individual malicious peers, malicious collectives and malicious collectives with camouflage

Table 2 – Tackling summary.

Security Threats	Trust and reputation models					
	EigenTrust	PeerTrust	BTRM-WSN	PowerTrust	ATSN	DWTrust
Individual malicious peers	✓	✓	✓	✓	✓	✓
Malicious collectives	✓	✓	✓	✓	✓	✓
Malicious collectives with camouflage	✓	✓	✓	✓	✓	✓
Malicious spies	†	†	†	†	†	†
Sybil attack	✓	†	†	†	†	†
Man in the middle attack	†	✓	✓	†	†	†
Driving down the reputation of a reliable peer	†	✓	✓	✓	†	†
Partially malicious collectives	†	✓	✓	†	†	✓
Malicious pre-trusted peers	†	✠	✠	†	✠	✠

✓, Resilient; †, vulnerable; ✠, not applicable.

are the most common tackled threats, while malicious spies and Sybil attack are not overcome by any or nearly any trust and reputation model.

Additionally, we have only found two models where the last threat (malicious pre-trusted peers) can be applied, which are EigenTrust and PowerTrust.

It is important to note that none of the presented models can absolutely prevent all the threats and that the proposed solutions given in Section 3 are just some helpful guides that aim to decrease the impact of each one of the associated threats, but they cannot (and the do not pretend to) completely overcome them.

6. Conclusions and future work

Trust and reputation management over distributed and heterogeneous systems has emerged in the last few years as a novel and accurate way of dealing with some security risks related to these environments.

Nevertheless, the application of such mechanisms involves the arising of new specific and related threats that should not be underestimated. As far as we know, this is one of the first works mainly focused on describing such threats and proposing solutions to overcome them.

In this paper, we have analyzed the main security threats that can be applied in most of trust and reputation schemes. Moreover, we have discussed them and suggested a possible way of tackling each one of those risks in the design phase.

A complete taxonomy of those threats or attacks has been developed as well, describing several possible dimensions of an attack over trust and reputation systems, and categorizing the exposed threats according to these dimensions or properties.

Additionally, we have presented some representative trust and reputation models and shown how they deal with those threats that can be applied to them, revealing that not all the threats are paid the same attention and none of them is categorically solved.

As for future work, we consider that an implementation and comparison of several of the most representative trust and reputation models, in terms of their response against some of the threats presented in this paper could be an interesting research line. In that way, we will focus on the development of a validation tool allowing researchers to perform such tests.

Finally, we hope this work helps to the development of this research field by constituting a guide for new trust and reputation model designers.

Acknowledgements

This work has been supported by a Séneca Foundation grant within the Human Resources Researching Training Program 2007. Thanks also to the Funding Program for Research Groups of Excellence granted as well by the Séneca Foundation with code 04552/GERM/06.

REFERENCES

Almenárez F, Marín A, Campo C, García C. PTM: a pervasive trust management model for dynamic open environments. In: Privacy and trust. First workshop on pervasive security and trust, Boston, USA; Aug 2004.

Almenárez F, Marín A, Díaz D, Sánchez J. Developing a model for trust management in pervasive devices. In: PERCOMW '06: proceedings of the 4th annual IEEE international conference on pervasive computing and communications workshops. Washington, DC, USA: IEEE Computer Society; 2006. p. 267.

Boukerche A, Xu L, El-Khatib K. Trust-based security for wireless ad hoc and sensor networks. Computer Communications 2007;30(11–12):2413–27.

Carbó J, Molina J, Dávila J. Trust management through fuzzy reputation. International Journal of Cooperative Information Systems March 2003;12:135–55.

Chen H, Wu H, Zhou X, Gao C. Agent-based trust model in wireless sensor networks. In: Eighth ACIS international conference on software engineering, artificial intelligence, networking, and parallel/distributed computing, SNPD 03; 2007. p. 119–24.

Dorigo M, Gambardella L, Birattari M, Martinoli A, Poli R, Stützle T. Ant colony optimization and swarm intelligence. In: Fifth international workshop, ANTS 2006. LNCS, vol. 4150. Brussels, Belgium: Springer; 2006.

Douceur JR, Donath JS. The Sybil attack. In: Proceedings for the 1st international workshop on peer-to-peer systems (IPTPS '02); 2002. p. 251–60.
Girao J, Sarma A, Aguiar R. Virtual identities – a cross layer approach to identity and identity management. In: Proceedings for the 17th wireless world reseach forum, Heidelberg, Germany; Nov 2006.
Gómez Mármol F, Martínez Pérez, G. Providing trust in wireless sensor networks using a bio-inspired technique. In: Proceedings of the networking and electronic commerce research conference, NAEC'08. Lake Garda, Italy; Sep 2008.
Gómez Mármol F, Martínez Pérez G, Gómez Skarmeta AF, TACS, a trust model for P2P networks. Wireless personal communications, special issue on ''information security and data protection in future generation communication and networking''; 2008.
Huang C, Hu H, Wang Z. A dynamic trust model based on feedback control mechanism for P2P applications. In: Autonomic and trusted computing, Third international conference, ATC. LNCS, vol. 4158. Wuhan, China: Springer; 2006. p. 312–21.
Josang A, Ismail R, Boyd C. A survey of trust and reputation systems for online service provision. Decision Support Systems 2007;43(2):618–44.
Kamvar S, Schlosser M, Garcia-Molina H. The EigenTrust algorithm for reputation management in P2P networks, Budapest, Hungary; May 2003.
Lam SK, Riedl J, Shilling recommender systems for fun and profit. In: WWW '04: proceedings of the 13th international conference on World Wide Web; 2004.
Marti S, Garcia-Molina H. Taxonomy of trust: categorizing P2P reputation systems. Computer Networks 2006;50(4):472–84.
Moloney M, Weber S, A context-aware trust-based security system for ad hoc networks. In: Workshop of the 1st international conference on security and privacy for emerging areas in communication networks, Athens, Greece; Sep 2005. p. 153–60.
Sabater J, Sierra C. REGRET: reputation in gregarious societies. In: Müller JP, Andre E, Sen S, Frasson C, editors. Proceedings of the fifth international conference on autonomous agents. Montreal, Canada: ACM Press; 2001. p. 194–5.
Sabater J, Sierra C. Review on computational trust and reputation models. Artificial Intelligence Review 2005;24(1): 33–60.
Songsiri S. MTrust: a reputation-based trust model for a mobile agent system. In: Autonomic and trusted computing, Third international conference, ATC. LNCS, vol. 4158. Wuhan, China: Springer; 2006. p. 374–85.
Tajeddine A, Kayssi A, Chehab A, Artail H. PATROL-F – a comprehensive reputation-based trust model with fuzzy subsystems. In: Autonomic and trusted computing, Third international conference, ATC. LNCS, vol. 4158. Wuhan, China: Springer; 2006. p. 205–17.
Wang W, Zeng G, Yuan L. Ant-based reputation evidence distribution in P2P networks. In: GCC: fifth international conference on grid and cooperative computing. Changsha, Hunan, China: IEEE Computer Society; 2006a. p. 129–32.
Wang Y, Cahill V, Gray E, Harris C, Liao L. Bayesian network based trust management. In: Autonomic and trusted computing, Third international conference, ATC. LNCS, vol. 4158. Wuhan, China: Springer; 2006. p. 246–57.
Xiong L, Liu L. PeerTrust: supporting reputation-based trust in peer-to-peer communities. IEEE Transactions on Knowledge and Data Engineering 2004;16(7):843–57.
Zhou R, Hwang K. PowerTrust: a robust and scalable reputation system for trusted peer-to-peer computing. Transactions on Parallel and Distributed Systems 2007.

Félix Gómez Mármol is a PhD student in the Department of Information and Communications Engineering of the University of Murcia. His research interests include authorization, authentication and trust management in distributed and heterogeneous systems, security management in mobile devices and design and implementation of security solutions for mobile and heterogeneous environments. He received an MSc in computer engineering from the University of Murcia. Contact him at felixgm@um.es.

Gregorio Martínez Pérez is an associate professor in the Department of Information and Communications Engineering of the University of Murcia. His research interests include security and management of IPv4/IPv6 communication networks. He received an MSc and PhD in computer engineering from the University of Murcia. Contact him at gregorio@um.es.