Trust Estimation in autonomic networks: a statistical mechanics approach

Stefano Ermon
Department of Computer Science
Cornell University
Ithaca, New York 14850
Email: ermonste@cs.cornell.edu

Luca Schenato
Department of Information Engineering
Universita di Padova
Padova
Email: schenato@dei.unipd.it

Sandro Zampieri
Department of Information Engineering
Universita di Padova
Padova
Email: zampi@dei.unipd.it

***Abstract*— Trust management, broadly intended as the ability to maintain belief relationship among entities, is recognized as a fundamental security challenge for autonomous and self-organizing networks.**

In this work, we focus on the evaluation process of trust evidence in distributed networks, where no pre-established infrastructure can be assumed. After casting the problem into the framework of Estimation Theory, a distributed Maximum Likelihood trust estimation algorithm is proposed. Strong parallels with Spin Glasses Theory are shown, providing key insights about the algorithm performance and limitations, as well as useful formulas for parameters tuning.

This work presents a mathematically rigorous analytical approach to the problem, and proposes the use of statistical physics methods not only to understand the complex dynamics that arise from the interactions of peers in decentralized networks but also to design robust protocols and algorithms whose performance can be rigorously evaluated.

I. Introduction

A major effort of the networking community is currently devoted at introducing security services into decentralized autonomous networks, such as MANETs and Wireless Sensor Networks (WSN). In fact security represents a major issue in these kind of networks, both for the lack of infrastructure and for the use of wireless communications that can be easily eavesdropped. Moreover the physical control on the access to the network, that is usually guaranteed by cables, is often completely lost. Even if for different reasons, similar issues also affect peer to peer networks [2], [15] and electronic commerce communities [24], [20], [22], since they all share the lack of a centralized control on the network and they usually provide open access, even to malicious users.

In the outlined setting, the most basic problem that needs to be addressed is that of being able to assess the trustworthiness of the nodes in the network [15], [4], [1]. In our context, we will broadly interpret trust as a belief relationship, where an entity is confident that another peer will operate fairly, or as it is designed [17].

This kind of information is in fact essential to be able to predict the future behavior of the peers, hence empowering the decision-making process of the protocols underlying the network (such as routing in MANETS and WSN, as pointed out in [17], [13]). In this way, potential damage caused by malicious users will be reduced, for example because most entities will avoid interacting with them, or at least they will do it in a conscious way. Moreover, trustworthiness quantification increases the degree of cooperation between peers by forcing them to act responsibly, since selfish actions are spotted and recorded. In other words, the ability to evaluate and keep record of trust relationships represents the main driving force of cooperation in most settings [13], [5].

The lack of infrastructure and of any authoritative entity in the network enforces the use of reputation-based systems, where trust is established by protocols that try to evaluate only the previous behavior of the entities, the only information available in a setting where peers have no prior knowledge of each other.

Unfortunately, despite the growing importance of this problem, state of the art design of trust management systems is still mostly at an empirical level [15], [21], [18], [1], [13], [22], [5]. As it is pointed out in [17] and in [10], most of the work on trust management in the literature is prevalently based on heuristics and simulation as evaluation method. The validation of the proposed systems is often an overlooked aspect, where not all solutions are actually verified and almost none are implemented and tested in a real environment. In this context, theoretical analysis is extremely rare and the comparison between different methods is therefore extremely difficult to accomplish, mainly because of the great simulative effort that would be required. Solutions are often hard to compare even on a simulative basis, since they often rely on different hypothesis and are aimed at different application scenarios.

In this context, a rare exception is the interesting analytical study of the problem presented in [8], a work that considerably inspired this study. Given the importance of a theoretical framework, the aim of this work is to provide a deeper understanding of the problem through a more rigorous approach that makes use of powerful tools and ideas arisen in statistical physics.

In particular we tackle the problem from a mathematically formal point of view by casting it into the framework of Estimation Theory. A distributed maximum likelihood trust estimation algorithm is then proposed and strong parallelisms with the dynamics of disordered magnetic systems are shown. Properties of Spin glasses discovered by physicists are used to develop an intuitive understanding of the algorithm behavior and to predict its performance.

Spin glasses models such as the famous Sherrington-Kirckpatrick model [16] have already proved to be extremely versatile and valuable to understand the global behavior of complex systems such as neural networks, whose dynamics can be modeled by the statistical mechanics of infinite-range Ising spin glasses [3]. In this work we show how they can also be used to understand the complex dynamics that arise from the local interactions of peers in decentralized networks, and we also demonstrate how to exploit the emergence of collective behavior in large families of correlated random variables in the design of distributed protocols to achieve a desired global behavior.

II. The Trust Estimation problem formulation

In our model, we consider a network consisting of N nodes, represented by a directed graph $G = (V, E)$ in which each entity can communicate with a certain subset of other nodes according to an adjacency matrix A.

We represent the real trustworthiness status of each node i with a bit variable $T_i \in \{-1, 1\}$, so that we collectively describe the trust status of the network with a *real trust vector* $T \in \{-1, 1\}^N$, adopting the convention

$$T_i = \begin{cases} 1 & \text{if node } i \text{ is trustworthy} \\ -1 & \text{otherwise} \end{cases}$$

In a real setting the complete *real trust vector* T is unknown to the peers, nonetheless, as it is outlined in the introduction, many protocols need to somehow estimate it to be able to operate correctly.

Even if T is unknown, nodes are usually able to judge their neighbors on the base of the history of their previous interactions, that we assume to be statistically correlated with the real trustworthiness status of the nodes. In particular we assume that T does not change over time and it is related to an *opinion matrix* $C \in \mathcal{R}^{N \times N}$ by the following equation

$$C = f(T, \omega), \qquad \omega \in \Omega \tag{1}$$

where Ω is a sample space and $f(\cdot)$ represents the way in which opinions are formed. In this setting an element c_{ij} of the opinion matrix C is the opinion that node i has on node j, and we assume that it is significant only if i and j are neighbors, since it is based on the history of their previous interactions. The outlined frameworks is very general and, with slight modifications, can be easily applied also to the case of P2P networks and e-commerce communities.

Within this model, the role of a trust management algorithm is that of estimating T from C, assuming that C and the form of $f(\cdot)$ in equation (1) are known. In the following sections we will show how to design such an algorithm in a distributed way, so that in each iteration only local opinions are used but still obtaining the solution that maximizes the likelihood given the entire matrix C.

A. The Gaussian case

In the following work we will mainly consider a special case of equation (1) in which the opinions between two entities i and j are modeled in the following way:

$$c_{ij} = \begin{cases} T_i T_j + w_{ij} & \text{if } A_{ij} = 1 \\ 0 & \text{if } A_{ij} = 0 \end{cases} \tag{2}$$

where $w_{ij} \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian random variable that models the uncertainty that affects the opinion-forming process.

With this choice, untrustworthy nodes are not selfish and isolated but they rather try to act as a group, since they tend to have good opinions of nodes that should not be trusted.

The role of the trust estimation algorithm is to find a trust configuration $\hat{T} \in \{-1, 1\}^N$ that represents a good estimate of the real trust vector T. The most natural approach here is to search for the configuration that is more likely to have generated a certain observed opinion matrix $\overline{C}$, or in other words the trust configuration with the highest a posteriori probability, given $\overline{C}$.

Under the assumptions made so far, we can compute the likelihood $LH(S; \overline{C})$ of any configuration S given an opinion matrix $\overline{C}$ in the following way:

$$LH(T; \overline{C}) := p(T|\overline{C})$$

where $p(T|\overline{C})$ is the probability of T conditioned that $C = \overline{C}$. In this way the maximum likelihood estimator $g : \mathcal{R}^{N \times N} \to \{-1, 1\}^N$ is

$$g(\overline{C}) := \arg\max_{\hat{T}} LH(\hat{T}; \overline{C})$$

Observe that the Bayes rule yields

$$p(T|C) = \frac{p(C|T)p(T)}{p(C)}$$

where $p(T)$ is the a priori probability of the discrete random variable $T \in \{-1, 1\}^N$ while $p(C)$ and $p(C|T)$ are the density and conditional density of the continuous random variable $C \in \mathcal{R}^{N \times N}$. This shows that

$$g(\overline{C}) = \arg\max_{\hat{T}} p(\overline{C}|\hat{T})p(\hat{T})$$

For the Gaussian model described in (2), assuming independence, we have that $p(C|T) = 0$ if C has a nonzero entry in position $(i, j) \notin E$. If instead C has nonzero entries only in $(i, j) \in E$, then

$$\begin{aligned} p(C|T) &= \prod_{(i,j) \in E} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(c_{ij} - T_i T_j)^2}{2\sigma^2}} \\ &= q(C) e^{\frac{1}{\sigma^2} \sum_{(i,j) \in E} T_i T_j c_{ij}} \end{aligned}$$

where $q(C)$ is a normalization constant independent of T.

Therefore maximizing $p(C|T)$ is equivalent to maximize $U(T; C) := \sum_{(i,j) \in E} T_i T_j c_{ij}$, so that

$$\arg\max_{\hat{T}} p(\overline{C}|\hat{T}) = \arg\max_{\hat{T}} U(\hat{T}; \overline{C})$$

It is easy to see that $U(T; C) = U(-T; C)$, a straightforward consequence of the symmetrical behavior of trustworthy and untrustworthy nodes. As a consequence, if also $p(T)$ is symmetrical, the resulting complete likelihood function

$LH(T;C)$ becomes symmetrical in T, a situation in which effectively distinguishing between the two kinds of nodes would be clearly impossible. Therefore we will concentrate on a priori distributions of the real trust vector T that are unbalanced, that is they privilege the presence of one kind of nodes. In a practical setting, such a requirement is not restrictive, because it is likely that more nodes are trustworthy rather than not.

Suppose that the a priori probability distribution is Bernoulli-distributed with parameter p, namely

$$p := P(T_i = 1)$$

Then, assuming independence, if we define

$$w(T) = |\{i|T_i = 1\}|$$

then we have

$$p(T) \quad = \quad p^{w(T)}(1-p)^{N-w(T)} = (1-p)^N \left(\frac{p}{1-p}\right)^{w(T)}$$

Since

$$w(T) = \frac{N + \sum_i T_i}{2}$$

we also have

$$\begin{aligned} p(T) &= (1-p)^N \left(\frac{p}{1-p}\right)^{\frac{1}{2}(N+\sum_i T_i)} \\ &= [(1-p)p]^{N/2} e^{\frac{1}{2}\log\left(\frac{p}{1-p}\right)\sum_i T_i} \end{aligned}$$

In this way we obtained that

$$p(T) = \gamma e^{-\lambda \sum_i T_i} \tag{3}$$

where γ normalizes to a probability distribution and

$$\lambda = -\frac{1}{2}\log\left(\frac{p}{1-p}\right)$$

Clearly the sign of λ determines if the a priori distribution is biased toward trustworthy nodes or the opposite, while for $\lambda = 0$ (or equivalently $p = 0.5$) we have the symmetrical case in which we cannot expect good results from the estimation.

Putting all together we obtain

$$\begin{aligned} LH(T;C) &= q(C) e^{\frac{1}{\sigma^2}\sum_{(i,j)\in E} T_i T_j c_{ij}} \gamma e^{-\lambda \sum_i T_i} \\ &= \gamma q(C) e^{\frac{1}{\sigma^2}(\sum_{(i,j)\in E} T_i T_j c_{ij} - \lambda\sigma^2 \sum_i T_i)} \end{aligned}$$

We conclude that the following proposition holds.

Proposition 1: The likelihood $LH(T;C)$ of a configuration T is proportional to a monotonic increasing function of

$$H(T) := \sum_{(i,j)\in E} T_i T_j c_{ij} - \eta \sum_i T_i \tag{4}$$

where $\eta = \lambda\sigma^2$.

We can therefore compute a maximum likelihood estimate of the real trust vector by setting

$$\eta = \lambda\sigma^2 = -\frac{\sigma^2}{2}\log\left(\frac{p}{1-p}\right)$$

and by maximizing (4) over all possible configurations T. Equation (4) is very important because it represents the energy or Hamiltonian of a configuration S in an Ising Model [16] in the presence of an external magnetic field of strength η that breaks the symmetry of the system. Again the physical interpretation confirms that in the case in which the a priori distribution of T is symmetrical, that is $p = 0.5$, the magnetic field disappears and the system becomes completely symmetrical.

The statistical physics interpretation ensures an intuitive understanding of the dynamics of the system and enables us to take advantage of the rich literature in the field to study our problem. In particular we are referring to systems known as Spin Glasses [16], that exhibit randomly distributed ferromagnetic and anti ferromagnetic interactions between spins. They represent the first studied class of systems with frustrated behavior, where the presence of conflicting interactions forbids simultaneous minimization of the interaction energies and hence the existence of a trivial global ground state.

So far we have shown how the original problem of maximum likelihood estimation can be reduced to the problem of finding the maxima of (4), or in other words the global minima of $-H(S)$, configurations known in physics as *ground states* of the system.

Unfortunately this optimization problem has been proved to be NP-Complete for generic graphs in [6], and hence an exhaustive search for global minima is widely believed to be computationally intractable. However a natural approach to tackle the problem seems to be a local search method based on Simulated Annealing. In fact this powerful method has been introduced in [9] to solve NP-complete optimization problems by searching for the ground states of a system described by a proper Hamiltonian function, that is exactly the same problem we need to solve.

In a nutshell, the general algorithm works as follows. Starting with any configuration S, we randomly choose one node i. If by flipping its trustworthiness value S_i we obtain a higher energy (we want to maximize), we accept the change. Otherwise, in case of a downhill move, the flip is accepted with a certain probability. This probability is chosen to be exponentially decreasing with the energy loss and depends upon the global parameter t, representing the temperature of the corresponding physical system that is gradually reduced by the algorithm.

An apparently similar approach has been previously proposed in [8], but using a model in which the energy was unrelated to statistical properties of the estimate. Moreover they did not solve the optimization problem, since they only used a Metropolis-like algorithm to generate a suitable Markov Chain with a Boltzmann-distributed steady state probability distribution, without being able to provide any

guarantee on the quality of the solution sampled from this distribution.

III. The Trust Estimation algorithm

The simulated annealing algorithm sketched out in the previous section can be seen as an iterative application of a *voting rule*, in which each node is repeatedly evaluated by its neighbors. In particular they express their opinions with a vote on its trustworthiness, and the *voting rule* takes them into consideration together with the current estimated trustworthiness status of the participants to the vote. To emulate the Metropolis-Hastings [11] algorithm we introduce stochasticity into the rule so that we obtain the desired Markov Chain structure with the proper steady state probability distribution.

The remarkable fact is that, despite the two levels of randomness present in the system, both in the local interactions between peers and in the way the opinions are generated, we can show the emergence of a collective behavior of the system that guarantees an efficient trust estimation.

Precisely, as mentioned before, at each time step a node node i is chosen randomly. The trustworthiness $S_j(k+1)$ of nodes j different form i are kept constant while, as in [8], the node i adopt the following voting rule for computing $S_i(k+1)$

$$P[S_i(k+1)|m_i(k)] = \frac{e^{\frac{S_i(k+1)(m_i(k)-\eta)}{t(k)}}}{e^{\frac{(m_i(k)-\eta)}{t(k)}} + e^{-\frac{(m_i(k)-\eta)}{t(k)}}} \quad (5)$$

where $m_i(k)$ is defined to be

$$m_i(k) = \sum_{j\in\mathcal{N}_i} (c_{ij} + c_{ji}) S_j(k), \quad (6)$$

$\mathcal{N}_i$ is the set of neighbors of i (we assume that i does not belong to $\mathcal{N}_i$) and $t(k)$ is the temperature parameter at iteration k.

In this way we obtain a Markov chain with state space $\{-1,1\}^N$ and with transition probability $p_{S,R} := P[S(k+1) = R|S(k) = S]$ which is equal to zero if the Hamming distance of S and R is greater than 1, while, if if the Hamming distance of S and R is less than or equal to 1, we have that

$$p_{S,R} = \frac{1}{N} \frac{e^{\frac{R_i(m_i(S)-\eta)}{t(k)}}}{e^{\frac{(m_i(S)-\eta)}{t(k)}} + e^{-\frac{(m_i(S)-\eta)}{t(k)}}}$$

where i is the index such that $S_j = R_j$ for all $j \neq i$ and

$$m_i(S) := \sum_{j\in\mathcal{N}_i} (c_{ij} + c_{ji}) S_j,$$

It can be shown([14]) that with a logarithmic temperature scheduling

$$t(k) = \frac{t_0}{\log(2+k)}$$

and with an initial temperature t_0 large enough, the probability of finding a global minimum converges to 1 as $k \to \infty$.

If instead we choose to fix the temperature parameter t over time, the voting rule defined by equation (5) is simply a modified version of the classical Metropolis-Hastings algorithm, where we introduce a Markov Chain with different transition probabilities but with the same steady state probability distribution. The graph associated with the Markov Chain is strongly connected and consists of a finite number (2^N) of states, each one with a self-loop. Therefore the transition matrix is primitive and the resulting chain is *regular* so that by Perron-Frobenius Theorem we conclude that there exists a unique steady state probability distribution that is reached from any initial probability distribution.

We will show in the following proposition that that with the proposed voting rule (5) we obtain a *reversible Markov Chain*, that is the following *detailed balance equation* is satisfied by π

$$\pi_S \; p_{S,R} = \pi_R \; p_{R,S} \; \forall \; R, S \quad (7)$$

where R, S are generic states and $p_{S,R}$ is the transition probability from state S to R.

Proposition 2: If $t = \frac{1}{\beta}$ is fixed, then the voting rule defines a Markov Chain whose steady state probability distribution π is Boltzmann-distributed

$$\pi_T = \frac{e^{-\beta H(T)}}{Z} \quad (8)$$

where

$$Z = \sum_S e^{-\beta H(T)}$$

plays the same physical role of a *partition function*.

Proof: We will show that (7) holds true. Notice that, if the Hamming distance between S and R is greater than 1, then $p_{S,R} = p_{R,S} = 0$ and so (7) holds true. If the Hamming distance between S and R is zero, then $S = R$ and so (7) holds true. Assume now that the Hamming distance between S and R is 1, and let i be the index such that $S_j = R_j$ for all $j \neq i$ and $S_i \neq R_i$. Observe now that $m_i(S) = m_i(R)$, and so denote these number with the symbol m_i. Then

$$\frac{p_{S,R}}{p_{R,S}} = \frac{e^{\beta R_i(m_i(S)-\eta)}}{e^{\beta S_i(m_i(S)-\eta)}} = e^{-2\beta S_i(m_i-\eta)}$$

On the other hand notice that

$$H(R) - H(S) = \\ 2S_i\eta - 2S_i \sum_{j|(i,j)\in E} S_j c_{ij} - 2S_i \sum_{j|(j,i)\in E} S_j c_{ji}$$

where the sums are over all outgoing and ingoing edges from i. By substituting (6)

$$H(R) - H(S) = -2S_i(m_i - \eta)$$

Therefore

$$\frac{p_{S,R}}{p_{R,S}} = e^{\beta(H(R)-H(S))}$$

and hence (7) holds true with $\pi_S = e^{\beta H(S)}$. Finally notice that

$$\sum_S \pi_S p_{S,R} = \sum_S \pi_R p_{R,S} = \pi_R \sum_S p_{R,S} = \pi_R$$

which shows that π_S is the steady state probability distribution. ■

The choice of the *voting rule* (5) however is not fundamental, because the steady state probability distribution (8) could clearly be also obtained using the standard Metropolis Algorithm, that is by choosing the following transition probability

$$p_{S,R} = \begin{cases} 1 & \text{if } \Delta U > 0 \\ e^{\beta(\Delta U)} & \text{otherwise} \end{cases} \quad (9)$$

where $\Delta U = 2S_i(\eta - m_i(S))$, and where S and R are states with Hamming distance equal to 1 and i is such that $S_j = R_j$ for all $j \neq i$ and $S_i \neq R_i$.

Both (9) and (5) are valid choices in the sense that the associated Markov Chain is guaranteed to converge to the steady state probability distribution (8). However the convergence rate depends on the eigenvalues of the transition matrix and hence on the voting rule used. By Perron-Frobenius theorem we know that in both cases 1 is a dominant eigenvalue, so that the remaining eigenvalues are located strictly into the unit circle, but it is their exact position that determines the convergence rate.

The method described so far is an application of the general idea arising from statistical physics to transform a combinatorial optimization problem, where the largest value of a target function is searched for over all possible configurations, into the sampling of all its large values through a distribution that assigns them the appropriate probabilities or weights. In our case the weight function is given by the parameter β, that represents the inverse of the temperature of the system. Its role is to tune the degree of randomness introduced by thermal fluctuations, as opposed to the one *quenched* into the system and described by equations (2) and (3). If we consider equation (8), when the value of β is large, the probability is concentrated on the configuration that minimizes $H(S)$, while for smaller values of β its relative weight decreases in favor of configurations with lower values of $H(S)$.

The role of the parameter β can therefore be read as a coefficient that enables us to tune the difficulty of the problem, that ranges from a trivial case ($\beta = 0$, high temperature) and a very difficult one (as $\beta \to \infty$, and the system freezes concentrating all the weight on the *ground state*). In fact even if we can use algorithms such as (9) or (5) to sample such a probability distribution for any value of $\beta > 0$, their convergence speed (determined by the eigenvalues of the transition matrix as previously discussed) to the desired steady state probability distribution decreases as the temperature is lowered. Intuitively, this happens because the Markov Chain is indeed ergodic for every value of β, but the degree of ergodicity decreases as β increases because downhill moves become more unlikely and therefore it is easier to get stuck in local minima of $H(S)$. On the other hand, for small values of β there is a faster convergence, but the resulting steady state probability distribution is noisy because of the weight distribution at high temperatures.

In this context Simulated Annealing can be seen as an attempt to gradually increase the difficulty of the problem by decreasing the temperature, while taking advantage of the solution found for a somewhat easier problem.

According to the previously described trust management system, each iteration of the algorithm consists in a local vote, where the results are decided according to equation (5). The most remarkable result is that the iterations are local, that is they involve only the opinions of the neighbors of a node being voted. In this way the opinions data do not have to travel all over the network, as it happens for example with a consensus-based system, but yet it achieves an estimate as good as it would be the one obtained by a centralized server that knows the entire opinion matrix C. This fact is particularly important in a decentralized setting, where it would be just too expensive to propagate all the information through the network.

Another important aspect of the dynamics that can be inferred from the statistical physics literature is that the existence of a global behavior essentially descends from the concentration of measure phenomenon. This fundamental and ubiquitous concept refers to the phenomenon of a function of a large number of random variables that, under certain conditions of regularity and on the statistical dependence between the random variables, tends to concentrate its values in a set of relatively small measure. The famous Laws of Large Numbers and the Central Limit Theorem are just some examples of this general principle applied to sums of independent variables. In our case, even if the global outcome is the result of a complicated local interaction between various level of randomness, the result not only appears to be almost deterministic, but it is also extremely robust. In fact the concentration of measure phenomenon ensures that the global properties are obtained to some extent regardless of the particular probability distribution of the noise that affects the opinions modeled in equation (1).

A mathematically rigorous approach for the Sherrington-Kirkpatrick model con be found in [7], where it is shown that the Spin glass qualitative behavior relies on weak hypothesis on the distribution of the couplings c_{ij} in equation (4) (that we assumed to be Gaussian). This ensures a great degree of robustness to the algorithm proposed, as it is confirmed in the simulative analysis, where it is tested in settings significantly different from the ideal Gaussian framework for which the algorithm has been originally conceived.

IV. Analysis

Thanks to a much stronger mathematical understanding of the problem, we have been able to design a distributed algorithm that finds a reasonable solution in the form of a maximum likelihood estimate. In this section we address the problem of understanding the average performance of the algorithm, both from a theoretical point of view and by the means of Monte Carlo simulations.

From a qualitative point of view, we can start the analysis by noticing that we cannot expect any topology-independent result. For example, in a network made by isolated vertices,

we cannot do any better than just using the a priori knowledge. We will therefore need to fix a topology to be able to show some meaningful results.

A. Case study: complete graphs

Even if it not representative of the topologies of any real world network, we will focus our attention on the case of a complete communication graph, mainly because most analytical results from Spin glasses theory are derived for this topology. However the intuition developed in the previous sections through physical interpretations suggests that the qualitative behavior should not change for non-singular topologies, and we expect to be able to take advantage in a near future of the recent analytical results obtained for other topologies such as Bethe lattices [12].

In the the case of a complete communication graph with N nodes, equation (2) becomes

$$C = TT' + W$$

where each element of the matrix W is $w_{ij} \sim \mathcal{N}(0, \sigma^2)$. Let

$$\hat{T} := \text{argmax}_{S \in \{1,-1\}^N} H(S)$$

Let moreover

$$h(\hat{T}) := |\{i : \hat{T}_i \neq T_i\}|$$

namely the number of correct estimates given by $\hat{T}$. We have the following result.

Proposition 3: If $\eta \neq 0$

$$\lim_{N\to\infty} \mathbb{E}\left[\frac{h(\hat{T})}{N}\right] = 0$$

Proof: Observe that

$$\begin{aligned}
\mathbb{E}[H(T)] &= \mathbb{E}[T'CT - \eta \sum_i T_i] = \\
&= \mathbb{E}[T'TT'T] + \mathbb{E}[T'WT] - \eta \sum_i \mathbb{E}[T_i] = \\
&= N^2 + \sum_{ij} \mathbb{E}[w_{ij}]\mathbb{E}[T_iT_j] - \eta N(2p-1) = \\
&= N^2 - \eta N(2p-1)
\end{aligned}$$

On the other hand we have that

$$\mathbb{E}[H(\hat{T})] = \mathbb{E}[(\hat{T}'T)^2] + \mathbb{E}[\hat{T}'W\hat{T}] - \mathbb{E}[\eta \sum_i \hat{T}_i]$$

Notice now that $\hat{T}'T = N - 2h(\hat{T})$ and that $-\eta \sum_i \hat{T}_i \leq |\eta|N$. Moreover from spin glass theory we know [19], [7] that the sequence $N^{-\frac{3}{2}}\mathbb{E}[\max_S S'WS]$ converges as N tends to infinity and so there exists a constant α such that $\mathbb{E}[\hat{T}'W\hat{T}] \leq \alpha N^{3/2}$ for all N. These facts imply that

$$\mathbb{E}[H(\hat{T})] \leq \mathbb{E}[(N - 2h(\hat{T}))^2] + \alpha N^{3/2} + |\eta|N$$

Since we always have that $H(\hat{T}) \geq H(T)$, then $\mathbb{E}[H(\hat{T})] \geq \mathbb{E}[H(T)]$ which implies that

$$N^2 - \eta N(2p-1) \leq \mathbb{E}[(N - 2h(\hat{T}))^2] + \alpha N^{3/2} + |\eta|N$$

If we denote $h(\hat{T})/N$ with the symbol x_N, then the previous inequality together with $0 \leq x_N \leq 1$ proves that

$$\mathbb{E}[x_N - x_N^2] \longrightarrow 0$$

as N tends to infinity. We need to show that this implies that $\mathbb{E}[x_N] \longrightarrow 0$.

In the remaining part of the proof we will restrict ourselves to the case $\eta < 0$ for the ease of explanation. However a totally symmetric argument can be developed for the case $\eta > 0$.

Remind that the symbol $w(T)$ means the number components in T equal to $+1$. Now notice that $w(\hat{T}) \geq N/2$. Indeed this follows from the fact that $H(\hat{T}) \geq H(-\hat{T})$ which implies that

$$H(\hat{T}) = \hat{T}'C\hat{T} - \eta \sum_i \hat{T}_i \geq \hat{T}'C\hat{T} + \eta \sum_i \hat{T}_i = H(-\hat{T})$$

and so $-2\eta \sum_i \hat{T}_i \geq 0$. Since $\eta < 0$ then we must have $\sum_i \hat{T}_i \geq 0$ and so $w(\hat{T}) \geq N/2$. Consider now the three sets $\mathcal{A}_1 = \{i | T_i = \hat{T}_i\}$, $\mathcal{A}_2 = \{i | T_i = -1\}$, and $\mathcal{A}_3 = \{i | \hat{T}_i = -1\}$. Clearly $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 = \{1, \ldots, N\}$ from which follows that their cardinality satisfy $|\mathcal{A}_|+|\mathcal{A}_2|+|\mathcal{A}_3| \geq |\{1, \ldots, N\}|$, or equivalently $(N-h(\hat{T}))+(N-w(\hat{T}))+(N-w(\hat{T})) \geq N$, which implies:

$$h(\hat{T}) \leq 2N - w(T) - w(\hat{T})$$

Using this inequality together with $w(\hat{T}) \geq N/2$, we can argue that $h(\hat{T}) \leq (3/2)N - w(T)$. Observe now that $w(T)$ is a binomial random variable, namely

$$P[w(T) = k] = \binom{N}{k} p^k (1-p)^{N-k}$$

We want to use this fact in order to estimate $P[x_N \geq 1 - \delta]$ where δ is such that $0 < \delta < p - 1/2$. Since $h(\hat{T}) \leq (3/2)N - w(T)$, then $h(\hat{T})/N \geq 1 - \delta$ implies that $w(T)/N \leq 1/2 + \delta$ and so

$$P[x_N \geq 1 - \delta] \leq P[w(T) \leq (1/2 + \delta)N]$$

Since $(1/2 + \delta)N \leq Np = \mathbb{E}[w(T)]$ we are in a position to apply the Chernoff bound which ensures that

$$P[w(T) \leq (1/2 + \delta)N] \leq e^{-\nu N}$$

where

$$\nu := \frac{(p - 1/2 - \delta)^2}{2p}$$

We can argue therefore that $P[x_N \geq 1 - \delta] \leq e^{-\nu N}$. We want to use this inequality in order to estimate $\mathbb{E}[x_N^2]$. Indeed, observe that

$$\begin{aligned}
\mathbb{E}[x_N^2] &= \frac{1}{N^2} \sum_{k=0}^{N} k^2 P[h(\hat{T}) = k] = \\
&= \frac{1}{N^2} \sum_{k \leq (1-\delta)N} k^2 P[h(\hat{T}) = k] + \\
&\quad + \frac{1}{N^2} \sum_{k > (1-\delta)N} k^2 P[h(\hat{T}) = k]
\end{aligned}$$

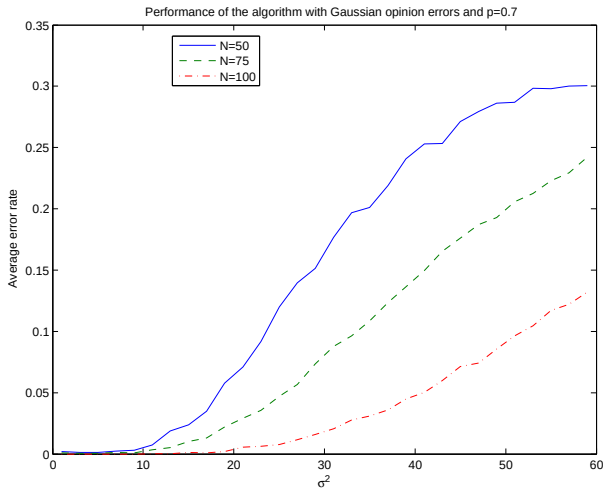

Fig. 1. Performance of the algorithm with a complete communication graph of N nodes for several values of N. The a priori probability p that a node is trustworthy is 0.7.

Observe that $k^2 \leq N^2$ and that, when $k \leq (1-\delta)N$, then $k^2 \leq (1-\delta)Nk$. Using these inequalities we obtain that

$$\begin{aligned}\mathbb{E}[x_N^2] &\leq \frac{(1-\delta)}{N}\sum_{k\leq(1-\delta)N} kP[h(\hat{T})=k] + \\ &\quad + \sum_{k>(1-\delta)N} P[h(\hat{T})=k] \leq \\ &\leq \frac{(1-\delta)}{N}\sum_{k=1}^{N} kP[h(\hat{T})=k] + \\ &\quad + P[h(\hat{T}) \geq (1-\delta)N] \leq \\ &\leq (1-\delta)\mathbb{E}[x_N] + e^{-\nu N}\end{aligned}$$

Observe finally that

$$\begin{aligned}\delta\mathbb{E}[x_N] &= \mathbb{E}[x_N - x_N^2] + \mathbb{E}[x_N^2] \\ &\quad -(1-\delta)\mathbb{E}[x_N] \leq \mathbb{E}[x_N - x_N^2] + e^{-\nu N}\end{aligned}$$

Since both term in the sum tends to zero, also $\delta\mathbb{E}[x_N]$ tends to zero.

In the case $\eta > 0$ one should consider $r(T) = N - w(T) = |\{i|T_i = -1\}|$ in place of $w(T)$ and repeat an analogous argument.

■

B. Simulative Results

From a simulative point of view, we are interested in measuring what is the fraction of nodes that the Simulated Annealing-based algorithm is not able to correctly identify, in expectation. If S^* is the maximum likelihood configuration returned by the algorithm, we are interested in the average error rate

$$\mathbb{E}\left[\frac{||S-T||_1}{2N}\right] = \mathbb{E}\left[\frac{h(S^*)}{N}\right]$$

where the expectation is taken over all levels of randomness. The first experiment is performed by simulating the environment described by the Gaussian model presented in section II-A, for various values of N and σ^2. The estimation algorithm uses the simulated annealing approach, with an exponential temperature cooling $t(k+1) = \alpha\ t(k)$ of parameter $\alpha = 0.91$ starting from an initial temperature of $10N^2$. However the choice of these parameters is not very important and does not affect significantly the performance.

As we can note in figure 1 the performance of the algorithm decreases as does the quality of the a posteriori information (measured by a larger variance on the opinions). However it is remarkable that the algorithm is never outperformed by the optimal estimator that is based solely on the a priori information S^*_{ap}:

$$S^*_{ap} = \begin{cases} [1,\ldots,1] & \text{if } p > \frac{1}{2} \\ -[1,\ldots,1] & \text{if } p \leq \frac{1}{2} \end{cases},$$

that clearly shows an average error rate of $(1-p)$.

Moreover we can see that proposition 3 is confirmed by the data, where the error rate decreases as N grows.

To show the robustness of the algorithm proposed we consider another reasonable model for (1), where the errors are Bernoulli distributed. In particular we assume that if $A_{ij} = 1$ then

$$c_{ij} = \begin{cases} T_iT_j & \text{with probability } 1-p_e \\ -T_iT_j & \text{with probability } p_e \end{cases} \tag{10}$$

This means that if a node is trustworthy ($T_i = 1$), then $c_{ij} = T_j$ with probability $1-p_e$, while the contrary holds when $T_i = -1$. Thus the parameter p_e represents the probability for a trustworthy node of misjudging a neighbor.

The results obtained with various error probabilities p_e and various networks sizes are shown in figure 2. The trust estimation algorithm uses a value of

$$\sigma^2 = \mathbb{E}[(c_{ij} - T_iT_j)^2] = 4p_e \tag{11}$$

and it shows an exceptional performance at least until p_e approaches 0.5. The results are comparable with those obtained with model (2), when the variance of the error on the opinions is the same according to equation (11). However when $p_e > 0.5$, on average there are more wrong opinions than correct, and the algorithm is outperformed by the one based solely on the a priori information. The average error rate shows a sharp phase transition phenomenon around $p_e = 0.45$, that is typical of spin glasses systems.

CONCLUSIONS

In this work we present a mathematically sound framework for trust evaluation in decentralized autonomic networks by casting it as an estimation problem. A maximum likelihood estimation algorithm is developed, with the fundamental property that it is completely based on local interactions between nodes without any need for central coordination.

These local interactions are characterized by several levels of randomness, both unavoidable because residing in the uncertainty in the opinions that the nodes have on their neighbors and artificially introduced by the algorithm in the voting rule. Despite that, an almost deterministic global behavior is obtained, as predicted by Spin glass theory models in the

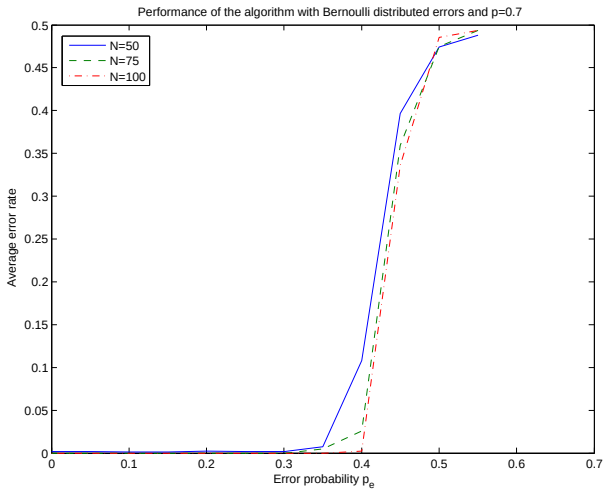

Fig. 2. Performance of the algorithm with a complete communication graph of N nodes for several values of N and opinions generated according to model (10). The a priori probability p that a node is trustworthy is 0.7.

more general framework of the concentration of measure phenomenon. In our opinion this concept might play a fundamental role in the design of protocols for decentralized settings where little is known or can be assumed on the behavior of singular nodes, but it is necessary to obtain a desired behavior of the network as a whole.

In this perspective statistical physic tools and more generally theories about disordered systems have already been successfully applied to the study of collective animal behavior and flocking. This case study on trust management represents a first attempt to lift the use of these tools to a designing perspective, from an engineering point of view.

As outlined in section IV, the resulting algorithm exhibits an excellent performance, as it was predicted by the theoretical analysis carried out. Another positive aspect is that the corresponding physical model enables us to understand at least at an intuitive level what will be the effect of parameter tuning on the dynamics of the system.

The great degree of robustness of the algorithm makes it suitable for a variety of settings in distributed networks, even where equation (2) is not a good model of the way in which opinions are generated, as the simulative analysis suggests. This property is particularly important because it ensures that to some extent the system is resilient to malfunctioning or even malicious users that try to jeopardize the system.

Even if there is evidence from spin glasses theory [7] that the qualitative behavior of the system relies on really weak hypothesis on the distributions of c_{ij} in the Hamiltonian (4) (a finite fourth moment is sufficient), a promising research direction is certainly to quantify the robustness of the method from a quantitative point of view, both from a theoretical and simulative perspective.

Immediate future work will also include the study of the performance of the algorithm on more realistic network topologies, such as the one generated with the Watts-Strogatz model [23] that exhibit small-world properties, including short average path lengths and high clustering.

ACKNOWLEDGMENT

The authors would like to thank Nicola Chesini and Giampietro Marcolin for their help on our initial investigations on the problem.

REFERENCES

[1] A. Abdul-Rahman and S. Hailes. A distributed trust model. In *Proceedings of the 1997 workshop on New security paradigms*, pages 48–60. ACM New York, NY, USA, 1998.

[2] K. Aberer and Z. Despotovic. Managing trust in a peer-2-peer information system. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 310–317. ACM New York, NY, USA, 2001.

[3] D. Amit, H. Gutfreund, and H. Sompolinsky. Spin-glass models of neural networks. *Math. Biosci Phys Rev A*, 32:1007, 1974.

[4] M. Blaze, J. Feigenbaum, and J. Lacy. Decentralized trust management. In *1996 IEEE Symposium on Security and Privacy, 1996. Proceedings.*, pages 164–173, 1996.

[5] S. Buchegger. Performance analysis of the CONFIDANT protocol. In *Proceedings of the 3rd ACM Int. Symp. on Mobile ad-hoc networking & computing*, pages 226–236. ACM New York, NY, USA, 2002.

[6] B.A. Cipra. The Ising model is NP-complete. *SIAM News*, 33(6), 2000.

[7] F. Guerra and F.L. Toninelli. The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics*, 230(1):71–79, 2002.

[8] T. Jiang and J.S. Baras. Trust evaluation in anarchy: A case study on autonomous networks. In *Proceedings of IEEE Infocom06*, 2006.

[9] S. Kirkpatrick, CD Gelatt, and MP Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[10] M. Langheinrich. When trust does not compute-the role of trust in ubiquitous computing. In *Workshop on Privacy at UBICOMP*, 2003.

[11] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087, 1953.

[12] M. Mézard and G. Parisi. The cavity method at zero temperature. *Journal of Statistical Physics*, 111(1):1–34, 2003.

[13] P. Michiardi and R. Molva. Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. In *Advanced Communications and Multimedia Security: Sixth Joint Working Conference on Communications and Multimedia Security, 2002, Portorož, Slovenia*, page 107. Kluwer Academic Publishers, 2002.

[14] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli. Convergence and finite-time behavior of simulated annealing. In *Decision and Control, 1985 24th IEEE Conference on*, volume 24, 1985.

[15] AA Selcuk, E. Uzun, and MR Pariente. A reputation-based trust management system for P2P networks. In *IEEE Int. Symp. on Cluster Computing and the Grid, CCGrid'04*, pages 251–258, 2004.

[16] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792–1796, 1975.

[17] Y. Sun, Z. Han, W. Yu, and K.J.R. Liu. A trust evaluation framework in distributed networks: Vulnerability analysis and defense against attacks. In *Proc. of IEEE Infocom*, 2006.

[18] YL Sun and Y. Yang. Trust establishment in distributed networks: Analysis and modeling. In *IEEE International Conference on Communications, 2007. ICC'07*, pages 1266–1273, 2007.

[19] M. Talagrand. Probability theory and spin glasses. *Unpublished manuscript*, 2000.

[20] Y.H. Tan. Toward a generic model of trust for electronic commerce. *International Journal of Electronic Commerce*, 5(2):61–74, 2000.

[21] G. Theodorakopoulos and J.S. Baras. Trust evaluation in ad-hoc networks. In *Proceedings of the 3rd ACM workshop on Wireless security*, pages 1–10. ACM New York, NY, USA, 2004.

[22] M. Venkatraman, B. Yu, and MP Singh. Trust and reputation management in a small-world network. In *Proceedings of Fourth Int. Conf. on MultiAgent Systems*, pages 449–450, 2000.

[23] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[24] L. Xiong and L. Liu. A reputation-based trust model for peer-to-peer e-commerce communities. In *IEEE International Conference on E-Commerce, 2003. CEC 2003*, pages 275–284, 2003.