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, self-managing and self-configuring networks, broadly referred to as autonomic networks. In fact the lack of a predefined and fixed infrastructure and their highly dynamic nature pose a number of new challenges ahead, especially from the security point of view.

As pointed out in [2] and [1], one of the most important challenges for autonomic networks is that of developing protocols to establish the trustworthiness status of the nodes in the network. Following [10], we will broadly interpret trust as a belief relationship, where an entity is confident that another one will operate fairly, or as it is designed.

In a setting where the global performance depends critically on the collaborations that take place among self-managing entities, maintaining belief relationships plays a key role. In particular it is essential to predict the future behavior of other entities, that is a fundamental aspect in the decision-making process of many protocols underlying these network architectures (such as routing in MANETs and WSN, as pointed out in [10] and [7]). In general potential damage caused by malfunctioning or even malicious and selfish entities can be greatly reduced by the employment of a trust management system, mainly because entities will generally avoid interacting with nodes that should not be trusted.

The lack of centralized coordination and of authoritative entities in general enforces the use of reputation-based systems, where trust relationships are established and maintained by protocols that evaluate the history of previous interactions with other entities.

Despite the growing role of autonomic networks, most trust management systems in the literature are still mostly at an empirical level. As it is pointed out in [10] and in [6], theoretical analysis is extremely rare and most state of the art systems are prevalently based on heuristics and on simulation as evaluation method. 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.

A rare exception is the interesting analytical study of the problem presented in [5], a work that considerably inspired this study.

The aim of this work is to provide a deeper theoretical understanding of the problem through a more mathematically sound approach that makes use of some powerful tools and ideas arisen in statistical physics. In particular for the sake of tractability we will focus on a non adversarial setting where all entities cooperate in the identification of faulty nodes. A practical example is a WSN where malfunctioning sensors need to be recognized as unreliable by the entities they are interacting with and a faulty node that is providing inexact measurements can perform a self-diagnosis only by querying its neighbors about the quality of its own measurements.

II. The Trust Estimation problem formulation

In our model, we consider a network of N nodes, represented by a directed graph $G = (V, E)$ where an entity can interact with a certain subset of other nodes according to the edge set E.

We represent the real trustworthiness status of each node i with a bit variable $T_i \in \{-1, 1\}$, with the convention

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

so that the trust status of the whole network can be described by a *real trust vector* $T \in \{-1, 1\}^N$.

We assume that the complete *real trust vector* T is unknown to the nodes, and that they are able to collect some evidence about it on the base of the history of their previous interactions with their neighbors, that we assume to be statistically correlated with T. In particular we assume that T 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, assumed to be significant only if i and j are neighbors since based on the history of their previous interactions.

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. Remarkably, the estimate computed in this way is as good as the best one that could be obtained in a centralized way assuming that the entire matrix C is known.

A. The Gaussian case

For the sake of tractability we will mainly consider a special case of equation (1), where the opinion that node i has on j is modeled in the following way:

$$c_{ij} = \begin{cases} T_iT_j + w_{ij} & \text{if } (i,j) \in E \\ 0 & \text{if } (i,j) \notin E \end{cases} \tag{2}$$

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

Within this framework, the role of the trust estimation algorithm is to find the trust configuration $\hat{T} \in \{-1, 1\}^N$ 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 that $C = \overline{C}$.

The likelihood $LH(S; \overline{C})$ of any configuration S given an opinion matrix $\overline{C}$ is by definition:

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

where $p(T|\overline{C})$ is the probability of T conditioned that $C = \overline{C}$, so that the maximum likelihood (ML) estimate is

$$\arg\max_{T} LH(T; \overline{C})$$

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 the maximum likelihood estimate can be computed as

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

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

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

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

Therefore maximizing $p(C|T)$ is equivalent to maximize $U(T; C) := \sum_{(i,j)\in E} T_iT_jc_{ij}$, 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 of T 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

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

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, 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

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

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

We can therefore compute a ML estimate of the real trust vector by choosing

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

and 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 [9] in the presence of an external magnetic field of strength η that breaks the symmetry of the system. Again the physical interpretation confirms that when the a priori distribution of T is symmetrical, that is $p = 0.5$, the magnetic field disappears. 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 a class of systems known as Spin Glasses [9], that exhibit randomly distributed ferromagnetic and anti ferromagnetic interactions between spins, depending on the sign of the coupling coefficients c_{ij}.

So far we have shown how the original trust estimation problem 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. This problem has been proved to be NP-Complete for generic graph topologies in [3], and therefore a global search for a configuration that is provably a global minimum is computationally intractable. However a natural approach to solve the problem is a local search strategy based on Simulated Annealing, a method introduced to solve optimization problems by searching for the ground states of a proper energy function, that is exactly the same problem we need to solve.

An apparently similar approach has been previously proposed in [5], but using a model in which the energy was unrelated to statistical properties of the estimate. Moreover they did not solve the optimization problem completely, since they only used a Metropolis-like algorithm to generate a suitable Markov Chain to sample solutions from, without providing any guarantee on the quality of the results.

III. The Trust Estimation algorithm

For the system described in the previous section a simulated annealing scheme can be implemented 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 algorithm we introduce stochasticity into the rule so that we obtain the desired Markov Chain structure with the proper steady state probability distribution.

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 [5], the node i uses the following voting rule to compute $S_i(k+1)$

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

where $m_i(k)$ is defined to be

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

$\mathcal{N}_i$ 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

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

If 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. In fact 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 π, and it is reached from any initial probability distribution.

In the following proposition we will find an expression for π by showing that the resulting Markov Chain is *reversible*, that is the following *balance equation* is satisfied

$$\pi_S \ p_{S,R} = \pi_R \ p_{R,S} \qquad (7)$$

for any pair of states R, S.

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

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

where

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

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 denote this number by 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 of 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 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. ■

As in the standard Simulated annealing case ([8]), using the voting rule 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 $k \to \infty$.

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.

IV. Analysis

In this section we address the problem of understanding the average performance of the algorithm described, both from a theoretical point of view and by the means of Monte Carlo simulations.

From a qualitative point of view, we firstly note 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, so we will need to fix a topology in order to show meaningful results.

A. Case study: complete graphs

Even if it not representative of the topologies of most real world networks, we will focus our attention on the case of a complete communication graph, mainly because most analytical results from Spin Glass theory are derived for this topology.

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 incorrect estimates given by $\hat{T}$. Then

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 ([4]) we know 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. A totally symmetric argument can be developed for the case $\eta > 0$.

Recall that the symbol $w(T)$ denotes 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 $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 = \{1, \dots, N\}$ from which follows that their cardinality satisfy $|\mathcal{A}_{|}+|\mathcal{A}_2|+|\mathcal{A}_3| \geq |\{1,\dots,N\}|$, or equivalently $(N - h(\hat{T})) + (N - w(T)) + (N - w(\hat{T})) \geq N$, 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 therefore argue 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

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

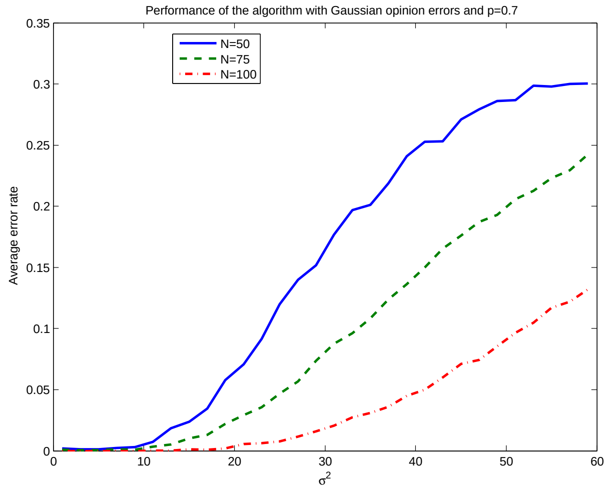

Fig. 1. Performance of the algorithm with a complete communication graph of N nodes for several values of N. The a priori probability p that a node is trustworthy is 0.7.

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

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

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 are not correctly identified, in expectation. If S^* is the ML 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 results.

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 $(1-p)$.

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 $(i,j) \in E$ then

$$c_{ij} = \begin{cases} T_i T_j & \text{with probability } 1 - p_e \\ -T_i T_j & \text{with probability } p_e \end{cases} \quad (9)$$

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

$$\sigma^2 = \mathbb{E}[(c_{ij} - T_i T_j)^2] = 4p_e \quad (10)$$

and it shows a good 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 (10). 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 glass systems.

These results can again be interpreted in the framework of Spin Glass theory. In particular in [4] 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 fact ensures a great degree of robustness, as it is confirmed by our simulative analysis.

Conclusions

The local interactions on which the algorithm is based are characterized by several levels of randomness, both unavoidable because modeling the uncertainty in the opinions that the nodes have on each other and artificially introduced by the algorithm in the voting rule. Despite that, a predictable and ordered global behavior is obtained, as suggested by Spin glass theory models developed by physicists.

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 individual nodes, but it is necessary to obtain a desired ordered behavior of the network as a whole.

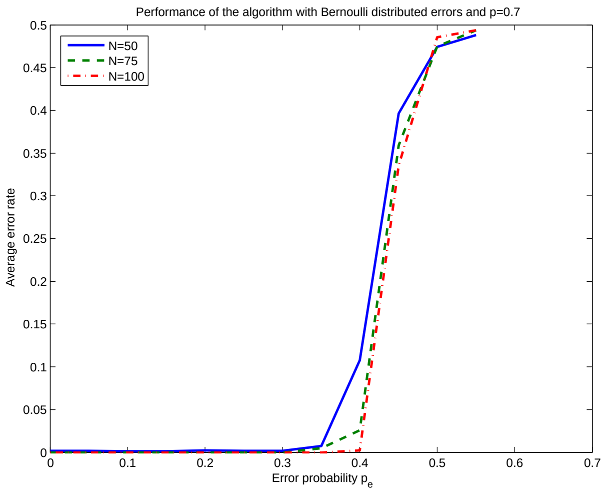

Fig. 2. Performance of the algorithm with a complete communication graph of N nodes for several values of N and opinions generated according to model (9). The a priori probability p that a node is trustworthy is 0.7.

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 design perspective from an engineering point of view.

Acknowledgment

This work was partially supported by NSF Expeditions in Computing grant 0832782.

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

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

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

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

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

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

[8] 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.

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

[10] 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.