An Agent Based Model for Trust and Information Sharing in Networked Systems

Kevin Chan
Army Research Laboratory
Adelphi, MD
Email: kevin.s.chan@us.army.mil

Sibel Adalı
Rensselaer Polytechnic Institute
Troy, NY
Email: sibel@cs.rpi.edu

***Abstract*—Vast amounts of information are generated, shared, and processed in tactical networks. In such systems, human cooperation is a crucial component for effective processing of information. However, human behavior is often mediated by social and organizational relationships, i.e. trust between team members and system level characteristics, such as network delays. The impact these different processes have on each other is not well understood. In this paper, we develop an agent based model for information sharing that incorporates trust into decision making. The nodes in our model are decision makers and are primarily responsible for the disambiguation of received information to obtain correct situation awareness as quickly as possible. Additionally, each node must share information with other nodes to enable the network to attain shared situation awareness. In our model, team members make trust evaluations for fellow team members that they cooperate with throughout a task. Most existing trust models concentrate on whether trust exists or not, and do not consider what task for which trust is being used. In contrast, we consider a new model of trust that incorporates two components: trust for competence and trust for throughput. In time constrained environments, both types of trust are crucial to mission success. Furthermore, this model allows us to study the impact of communication delays on overall trust. We also show how these trust values can be converted to labels that control agent's decision making behavior. To test out this proposed model, we use a command and control experiment platform called ELICIT (Experimental Laboratory for Investigating Collaboration, Information-sharing and Trust). We give initial experimental results that show how trust of nodes change in an ELICIT information sharing task for various settings of initial team trust, the biases of the nodes and possible communication channel disturbances.**

I. Introduction

Vast amounts of information are generated, shared, and processed in tactical networks. In such systems, human cooperation is a crucial component for effective processing of information. However, often human behavior is mediated by social and organizational relationships between team members and system level characteristics, such as network delays. For example, a team member trying to achieve a task may choose to maximize her energy by collaborating with the best people for a given problem, those who are providing the best information. But if those team members are not responsive, this is not the best choice. Instead, one might try to balance responsiveness with the ability. In many scenarios, being able to get high quality information is crucial [1]. However, if the volume of the available information is too low, this might lead to a loss in situation awareness as well [2]. For example, understanding what others in the network know about a specific problem reduces uncertainty in the team and allows one to assess when they might have a strategic advantage [3]. It is well known that social and organizational factors impact how information is shared among individuals [4]. For example, a team leader may be very competent but may not be willing to interact directly with a subordinate. So, the bandwidth from the leader to the subordinate may be limited. Among peers, one might believe information received from a competent peer over another peer. However, peers may choose to share information first with peers that they trust socially. All of these factors result in different behaviors in the network and have a tremendous impact on how the team functions as a whole and whether they are effective in achieving situation awareness. Being able to design such a system requires an understanding of the interaction of the humans or agents with the system [5], [6]. To be able to study this problem effectively, the appropriate model of the relationships between individuals in the team for the appropriate task must be developed.

In this paper, we introduce a model of trust that captures two very important aspects of team collaboration which we will call competence and willingness. Trust has been studied in many different contexts in the literature [7]–[10] and many different constructs have been proposed to explain why a trustee is trustworthy. One commonly used taxonomy [9] considers four main components: benevolence, competence, integrity and predictability. However, this taxonomy is context independent and does not really explain the dependence of the trustor on the trustee for solving a problem in a time-constrained environment. For example, trustor depends on competent and timely responses from the trustee. The timeliness component can be explained as the trustee's competence in sending timely information or the predictability of her responsiveness. In fact, the main problem we would like to address is trust in the context of the specific problem domain. A better suited taxonomy for our work comes from the trust in automation literature [5], which uses three main categories: performance, process and purpose. The performance mirrors the competence aspect both from [9] and in our work. Process considers to which degree the trustee is capable of achieving the trustor's goals. In this case, getting information quickly is a process component. The purpose deals with the motives, integrity of the trustee. In our case, we are assuming that the team shares the same values and goal. Hence, we are not modeling this aspect explicitly. Based on this reasoning, we introduce a trust model with two main components:

1) Competence: the ability of a team member to send pertinent or useful information. This is the performance based aspect of trust.
2) Willingness: the amount of effort a team member is willing to spend on the given node. This is the process based aspect of trust.

This model also brings together the notion of competence frequently studied in psychology and social science literature based on quality judgments [4] with the game theoretic approach to trust studied more in economics and cognitive science based on the past behavior of participants [11] and their reciprocity. Our model also allows us to study the impact of external factors on trust. For example, communication delays and packet loss may alter a peer's behavior and how much they are trusted within the scope of the specific team exercise.

Note that both types of trust can be impacted by previous experiences. If the trustor knows or has interacted with the trustee, he might have some prior information about how much the trustee should be trusted for competence and willingness. Even in cases where the trustee is not known, a prior trust may be assigned to her because she belongs in a specific social group (works in a specific organization) or holds a special role (team leader). These prior trust values have a direct impact on how the trust for this trustee evolves over time [11], [12]. We develop a Bayesian model to incorporate new evidence for trust into prior values. We show how we can incorporate individual differences in how trust information is processed into our model as well.

To illustrate our proposed model, we use a command and control experiment platform called ELICIT (Experimental Laboratory for Investigating Collaboration, Information-sharing and Trust). The ELICIT platform has configurable scenarios that enable groups to focus on the task of identifying details of a fictional insurgent threat. Packets of partial information are distributed so that no one participant receives all the information necessary to perform the task; thus, information sharing is required in order for any participant to determine a solution. Human-agent models have been created to enable ELICIT experiments to be run in place of human participants. We compute and illustrate how the trust of agents in an ELICIT scenario changes based on our model using preexisting ELICIT traces. We also add ELICIT to a communication network emulation environment EMANE (Extendable Mobile Ad-hoc Network Emulator) which allows for the study of the impact of communication network quality on the trust values of the participants. We give preliminary experimental results that show how trust of nodes are impacted by various tunable parameters in our network.

Our paper is a first step towards studying the impact of trust on situation awareness. We show how trust can be computed and incorporated into decision making in an information sharing scenario. In future work, we expect to modify agent behavior based on trust and situation awareness evaluation. We will use these models to examine the various network effects in different scenarios.

II. Trust Model

In this paper, we are considering an information sharing task. This scenario consists of a network of people exchanging messages over a communication channel. The people are treated as nodes in the network. Nodes can perform one or more of the following functions in information sharing: information provider, i.e. source and router of information. Additionally, there may be information posted to a set of websites, which will hold any posted information for the duration of the task. Identity of information source and poster will be known. In the modeled scenarios, information is not processed and cannot be fused together. One justification for this information policy is that (useful) information may be lost when information is combined. The nodes in the network share the goal of finding the correct information by gaining appropriate situation awareness as quickly as possible. We define situation awareness (SA) in this context broadly: the amount of information a node has, as well as the node's understanding about the reliability of the information and availability of other information are all part of the node's SA. In this context, it is equally important to have information at hand and to know what everyone else in the network knows.

To this end, the nodes try to disambiguate information received and share relevant information with other nodes. While performing these actions, they make decisions about which nodes to interact with to improve their success and "trust" is one of the crucial factors in this decision. Nodes share information with other nodes that are known to be reliable and have access to a great deal of useful information. This allows them to avoid flooding the communication network and also concentrate their energy on more promising activities. In Figure 1 we see our general model of trust. Agents observe network behavior to assess trust in two dimensions, competence and willingness. These two aspects of trust are converted to trust categories relevant to the mission (see Section III for details of an example). The agent then uses these categories, the overall situation awareness and other individual factors to decide on the next set of actions. Agent's behavior together with other agents change the information flow in the allows for a new set of observations. Literature on trust points to

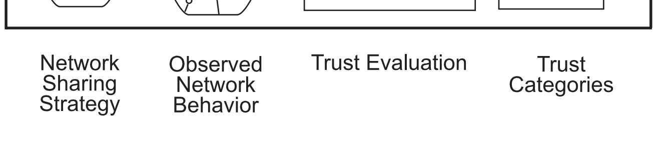

Fig. 1. Trust evolution process

many different definitions of trust, in essence modeling various types of expectations a trustor may have on how a trustee may behave in a specific context. In our scenario, only a few of these definitions are valid. For example, the notion of integrity, as to whether the trustee will tell the truth does not apply to our scenario. We assume all nodes are honest and do not purposefully give incorrect information to each other. However, nodes choose which nodes to interact with based on two important factors:

1) *Competence* is the capability of node being able to accomplish a task. One of the two components of competence is ability of the node to distinguish useful information from noise before sending it to neighboring nodes. One can consider this as a form of human capital. On the other hand, the node should be capable of getting useful information from others, i.e. through connectivity to other nodes who are equally competent themselves. This component can be considered the social capital of the node, i.e. the network effect.
2) *Willingness* is the amount of information that flows from one node to a neighboring node. In other words, how much of its bandwidth, i.e. attention it is willing to dedicate to a neighbor. We are assuming nodes will tell each other what they think is most important without alteration. However, nodes may not share information with each other due to hierarchy of the organization or that they are concentrating their limited energy on communicating with other nodes. Game theoretic approaches have shown that nodes typically are willing to share with others who they expect will reciprocate their actions. This belief may be based on past relations or current actions. However, the current actions are still impacted by the network that they are operating on. Nodes that have good bandwidth, i.e. reliable and fast links to other willing nodes are likely to be more willing.

In both cases, the nodes can only experience the interactions with their immediate neighbors. Therefore, they have to make inferences about competence and willingness of their neighbors based on local evidence, even though this evidence is impacted by the network structure and the communication channels. For example, if a competent node is hard to reach, then it is not worth trying to communicate with that node to accomplish the current task. The task specific trust modeled in these two axes, competence and willingness, in turn impacts how the nodes will act in the task. They will make decisions based on who they perceive as competent and/or willing to share. Before we introduce our decision model, we first describe how trust is computed.

A. Computing Trust

For each axis, i.e. competence and willingness, we assume a similar data model. In this model, a node x trustor needs to evaluate the trust for another node y, i.e. trustee. The trust is based initially on x's knowledge of y prior to the information sharing task. The initial trust can be based on many components, from x's propensity to trust, y's past behavior in interactions with x, y's reputation and other heuristics or biases x can employ to assess y's trustworthiness. For example, homophily or position in the organization are possible factors that can be used. This notion of trust has been studied in the literature from various disciplines, especially in psychology.

As x and y interact in the information sharing task, x gathers additional information about y. This information is considered evidence about y's competence or willingness. Evidence can be positive or negative. We assume a real value between $[0, 1]$ can best model this where 0 is a fully negative evidence and 1 is a fully positive evidence. Any value between the two indicates an evidence that is not fully negative or positive. Given new evidence, node x updates her trust for y. We use a Bayesian model to describe how the new evidence will be viewed by x based on her prior trust for y. Below, we provide some suggestions on possible models that can be used to compute trust.

We use a value between $[0, 1]$ to model trust where a value of 0 means there is no trust and a value of 1 means complete trust. Note that "no trust" is not equivalent to "distrust". No trust means that the person does not meet the criteria for trust, but this may change if sufficient positive evidence is found. However, a person who is distrusted may never be trusted even in the presence of overwhelming positive evidence. It is possible that there is a threshold for each person where any value below the threshold is considered "not trust".

We use probability distributions to model trust where the mean value can be used to estimate trust. In addition, we can also model uncertainty of trust. For example, if little is known about another person, even in the case of average amount of initial trust, there may still be a great deal of uncertainty about it. In other words, two distributions with the same mean may have different standard deviations. This means that if there is a large range of values trust can take, then the trust value is uncertain. We note that both are important in decision making, a high or low trust value is particularly useful only if there is some certainty associated with these values. This is similar to the notion of situation awareness for trust. Users can behave differently based on how well they can judge trust.

1) Priors: The prior value of trust is given by a probability distribution function with a given mean. We use probability 0.5 to represent neutral trust, whereas any value higher than 0.5 represents high trust and a value less than 0.5 represents low trust. A frequently used prior for either competence or willingness is given by the beta distribution [13], [14] which is given by two shape parameters α and β with mean $\frac{\alpha}{\alpha+\beta}$. One can consider α as the number of prior positive experiences and β as the number of prior negative experiences (α, β are both non-negative integers). If $\alpha > \beta$, trust is positive, i.e. mean is higher than 0.5. The beta distribution is given by

$$p(t) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)} \tag{1}$$

where $B(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$ and t is trust.

If there is no prior information, we can use the beta distribution where $\alpha = \beta$. Note that beta distribution has been used extensively in the literature to model trust especially in e-commerce settings [13]. We can use $\alpha/\beta = 1/K, 1, K$ for low, neutral and high trust for varying degrees K that models varying strengths of initial beliefs.

Another alternative model for prior trust is based on normal distribution. This model allows us to represent evidence that is not binary but is an evaluation of the true trust values. The initial trust can be set to low, neutral, and high trust by with $\mu_0 = 0, 0.5, 1.0$, respectively. The standard deviation in this case models the uncertainty involving the initial trust values, the lower it is, the more certain the trustor is about her trust evaluations.

$$p(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu_0)^2}{2\sigma^2}} \tag{2}$$

Both models are applicable for either competence or willingness. Whenever the evidence can be evaluated as binary, the first model can be used. However, in many cases, the evaluation of the evidence can be fuzzy as well. In these cases, the second model can be a better match.

2) Evidence: Given new evidence, the trust is updated to compute the new trust. We use Bayes' rule for this:

$$p(\text{trust}|\text{evidence}) = p(\text{evidence}|\text{trust}) * p(\text{trust}) \tag{3}$$

where $p(\text{trust})$ is the prior trust given as above. The factor $p(\text{evidence}|\text{trust})$ models how new evidence is viewed based on new evidence. We consider a number of models for evidence as well.

First, consider a model where evidence is collected after a set of interactions. At each step, evidence is evaluated using a binary method: either positive or negative. Based on this evaluation, let r, s be the number of positive and negative evidences that x has for y. Then, we can use the following formula:

$$p(\text{positive evidence}|t) = \binom{r+s}{r} t^r (1-t)^s \tag{4}$$

based on the beta distribution. Using the beta to represent the initial distribution of trust, we use the binomial distribution to represent evidence. As these two distributions are conjugate prior pairs, this results in a beta distribution. In this case, t is the (prior) trust. According to this model, the person is more likely to believe positive evidence (i.e. higher r values) if they have higher prior trust (high value of t).

In the alternate model, suppose we are using the model where evidence is an estimate of the true value of trust,a value between $[0, 1]$ and r is the overall value of the evidence so far. The model below shows how r is incorporated a trust estimate based on the current value of t and a tunable parameter σ_U:

$$p_\omega(\omega = \text{evidence}|t = \text{trust}) = \frac{1}{\sqrt{2\pi\sigma_U^2}} e^{-\frac{(\omega-t)^2}{2\sigma_U^2}} \tag{5}$$

In this model, evidence that is closer to trust is weighted higher than other evidence. The standard deviation σ_U can be used to model the individual differences in the evaluation of evidence based on the current trust. The higher the standard deviation, the wider range of evidence will be considered as probable. In other words, the lower the standard deviation, the more biased the individual towards initial trust values. One distinction between p_ω and p_b is that, p_b does not allow for any tunable parameters for individual differences. Past work has approached this problem by using an additional belief domain for evaluating evidence [14]. Such a model can be adopted in our case as well.

Based on a given model for prior and for evidence, we can now compute trust distribution between any pair of individuals as a function of its prior and new experiences. Given a probability distribution of trust, we can find the expected value of trust and the uncertainty as the standard deviation of the trust distribution.

Other models can be introduced based on the problem domain. For example, a model that considers timing effects of signals might be used to represent the impact of current experiences on the cognitive load. Evaluation of such models is part of our future work.

B. Information Agent Model

In this section, we describe how we can use the above model to compute competence and willingness in our information sharing scenario. As it is easier to evaluate evidence for competence in a binary fashion, we use the beta distribution to model the prior and evidence for competence.

Competence is based on skill. An ideal method would consider how valuable the information being received from another node. However, there is no known method to judge the value of a fact sent by another node without deep knowledge of the information domain. So, the only signal we can reliably use about skill is the number of new facts received. However, if better methods are available, we can substitute them instead of our model. Let f_k be a factoid received by x from y at time l. The value of f_k determines whether f_k is considered valuable, i.e. contributing to competence or not, denoted by $v(f_k, x, y, l)$. If f_k is not received by x before time l, then $v(f_k, x, y, l) = 1$, else $v(f_k, x, y, l) = 0$. If $v(f_k, x, y, l) = 0$, then f_k is considered a negative evidence of Y's competence by x. Otherwise, it is a positive evidence. The important thing for the agent x is to consider how much positive evidence she received from y. Note that this might be due y's ability or her social network. If the underlying agent model does not have the capability to filter data, then competence is purely based on the social network. The model p_b can be easily used in this case.

Willingness is based on how much information is expected from a person. The problem is that "amount of information" is a relative measure, cannot be defined quantitatively outside of the scope of the problem context. Furthermore, it is not a global measure either. When x is evaluating y's willingness, she is comparing y to others in the same situation, i.e. other peers x is interacting with. These relative values can also change over time as more is known in the network and fewer new facts are being sent. So, these qualitative values need to be updated as a function of time as well.

For the willingness model, the agent has to decide how much evidence there is for agent y's bandwidth to x. To do so, the agent x will look at how many messages are provided by all its neighbors (other agents sending information to x) over a period of time (or the last c time units). Instead of using direct number of messages which can be a noisy indicator, we sort all the nodes neighboring x in terms of how many message they sent in the last time interval. For example, if nodes y_1, y_2, y_3, y_4 sent 2, 2, 5, 0 messages, we obtain: $y_3, (y_1, y_2), y_4$ as y_1, y_2 are tied. We then assign score of 0 to the node that sent the least messages, $\frac{1}{n-1}$ to the node that that sent the second least messages, etc. The node that sent the most messages will be assigned score of 1.

C. Trust Model Estimation

Using the Bayesian model to estimate the posterior distribution of trust based on a prior trust and the given evidence, we derive expressions to find expected value and expected uncertainty with regard to both willingness $(E(t_\omega), \sigma^2_\omega)$ and competence $(E(t_c), \sigma^2_c)$.

1) Beta-Binomal: In our proposed trust model for competence, we measure the positive (r) and negative (s) evidence, which is the number of new or redundant messages received from each node. Given a prior distribution of Beta(α, β), the posterior distribution is Beta$(\alpha + r, \beta + s)$. Therefore, the posterior mean (expected trust) is

$$E(t_c) = \frac{\alpha + r}{\beta + s + \alpha + r} \tag{6}$$

and variance (uncertainty) is

$$\sigma^2(t_c) = \frac{(\alpha + r)(\beta + s)}{(\alpha + r + \beta + s)^2(\alpha + r + \beta + s + 1)} \tag{7}$$

2) Gaussian-Gaussian: In the trust model for willingness, each node assigns a t_w value between 0 and 1 based on the total number of messages received from neighboring nodes within a defined time period as discussed earlier.If a normal distribution is used to represent the prior and evidence, then the posterior distribution will also be normally distributed. Given prior $\mathcal{N}(\mu_0, \sigma^2_0)$, user defined standard deviation σ_U and evidence ω, the posterior distribution is $\mathcal{N}\left(\left(\frac{\mu_0}{\sigma^2_0} + \frac{\mu}{\sigma^2_U}\right) / \left(\frac{\sigma^2_0 + \sigma^2_U}{\sigma^2_0 \sigma^2_U}\right), \left(\frac{\sigma^2_0 \sigma^2_U}{\sigma^2_0 + \sigma^2_U}\right)\right)$. The posterior mean (expected trust) is

$$E(t_\omega) = \left(\frac{\mu_0}{\sigma^2_0} + \frac{\mu}{\sigma^2_U}\right) / \left(\frac{\sigma^2_0 + \sigma^2_U}{\sigma^2_0 \sigma^2_U}\right) \tag{8}$$

and variance (uncertainty) is

$$\sigma(t_\omega) = \frac{\sigma^2_0 \sigma^2}{\sigma^2_0 + \sigma^2_U} \tag{9}$$

III. Experiment Platform: ELICIT

In order to study the potential effect of the proposed model of trust in an information sharing scenario, we have implemented the proposed trust models in a simulation experiment platform called the Experimental Laboratory for Investigating Collaboration, Information-sharing, and Trust (ELICIT). ELICIT was developed as part of an initiative to develop and test organizational principles that enable transformation from traditional hierarchy-based command and control practices toward more agile practices that transfer more power and decision rights to the edge of the organization [15]. ELICIT is a Java-based software platform that can be used to run multi-user experiments focused on information, cognitive, and social domain phenomena. People participate in research sessions mediated by ELICIT by working together in teams that can be configured to reflect different organizational approaches (e.g., hierarchy, edge, hybrid and others) [16]. Additionally ELICIT can be performed with the participants being replaced by configurable human agents. The group of participants is tasked to gain shared situation awareness of the details of a fictitious insurgent threat. Specific packets of partial information, called factoids, are randomly seeded to the participants at the start of each experiment. Participants must then share and receive this information with others to obtain and distribute information to gain knowledge of the threat. The participants can send their knowledge of the threat to the ELICIT administrative system throughout the experiment. To measure their performance, we evaluate a measure called correctness to evaluate the level of situation awareness. Given varying amounts of information flow or processing ability, available individual or shared situation awareness will vary. The intention of this study is to understand better how trust can impact SA.

In previous work, we used the ELICIT agent parameters and represented communication network effects to determine the impact on task performance [17]. Previous to this work, the communications models within ELICIT were idealized with respect to the communication network. While decision-making performance was studied, the evolution of trust as a function of these communication network effects were not studied.

Our model based on competence and willingness trust with the ability to measure both the expected value and uncertainty provides us with a great deal of flexibility in reasoning about trust and modeling agent behavior. In the context of an ELICIT task, we can compute this information for all pairs participating in an information finding task. ELICIT supports five labels for agents: Trusted unknown, trusted non-discriminating, trusted discriminating, distrusted and no opinion. We can use our model to map to these labels in various ways. For example, suppose we are given thresholds l_T, h_T and l_U, h_U that describe which values of trust (T) and uncertainty (U) are considered high or low as part of our model. In other words, values higher than the upper threshold are considered high, lower than the lower threshold are considered low. In all other cases, the values do not have a category. These thresholds can be fixed for all agents or may change to represent different risk taking behavior on the part of the agents. Given trust values of x for competence and willingness of y and the appropriate thresholds, x thinks y is:

- **Trusted unknown** if there is low uncertainty and high expected value for willingness and high uncertainty for competence: models a case where y is expected to give

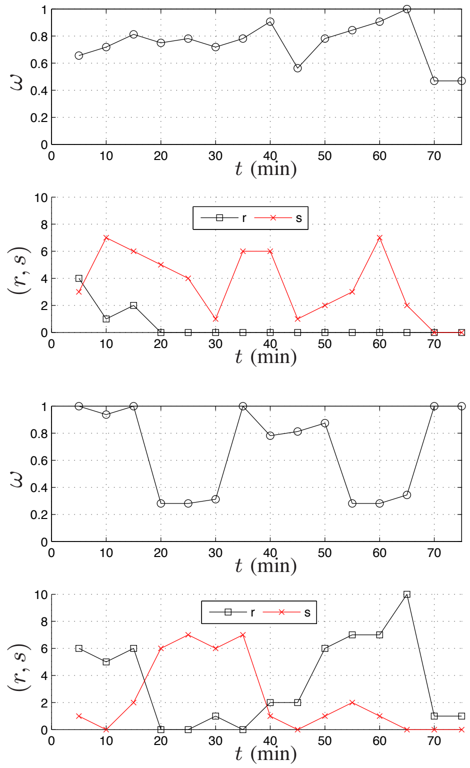

Fig. 2. Different evidence of activity to node x, from node 10 in Evidence 1 (top) and from Node 5 in Evidence 2 (bottom).

priority to x given its rank or relationship to x, but their competence in the problem domain is unknown.

- **Trusted non-discriminating** if there is low uncertainty and high expected value for willingness, low expected value or high uncertainty for competence: models a case where y's willingness is similar to the previous case, but y is not consistently competent in the problem domain.
- **Trusted discriminating** if there low uncertainty and high expected value for competence and high expected value of willingness: models a case where the agent is both competent and willing.
- **Distrusted/Untrusted** if there is low uncertainty and low expected value for competence and willingness: models a case where the agent is known to be not willing and not competent, no reason to trust him.
- **No opinion** in all other cases: for example an agent who is competent but not willing is not trusted to accomplish a task because not much actionable information can be obtained from her.

These labels are just an example of how our model can be used to model different types of trust related behavior. In different decision making domains, different labels and mappings may be more appropriate.

IV. Experimental Design

We use the results of ELICIT to evaluate trust between agents during experiments for a set of network quality of service configurations by varying message propagation delays. We evaluate every pairwise (t_{ij}: node i trust in node j) trust relationship within the organization. With regard to competence, positive evidence r_{ij} is modeled by the number of new messages received, and negative evidence s_{ij} is the number of factoids (including redundant ones) received from each node. In terms of willingness, we evaluate trust by tallying the total received number of factoids received w_{ij} and sorting assigning w_{ij} values as described in Section II-B. Note that the actual behavior of the nodes do not change with time as a function of trust as we are using an existing trace. Our purpose in this paper is to understand the evaluation of trust based on the factoid transmission behavior throughout the experiment to determine the validity of this model.

The experiment traces used in this paper consist of a set of 17 nodes randomly deployed into a unit area, and the communication radius is set to ensure connectivity. The same network topology is used for each experiment. Each experiment runs for 60 minutes and trust is evaluated every 5 minutes. A set of 68 factoids is randomly distributed into the network in three rounds at $0, 5, 10,$ minutes into the experiment. Factoids are seeded into the network once (no duplicate seeds). The factoids are shared between pairs of nodes that share links. Factoids are processed as they are received. Agents are provided with a message buffer, so that factoids are not dropped as a result of the agent being busy. Additionally, message propagation delay is varied to model the network quality of service. Communication delays of 8, 15, 30, 60, 120, 300 seconds.

In addition to the varying network delay, we consider the effect of trust initialization. We will consider priors for the beta distribution of $(\alpha, \beta) = \{(10, 100), (1, 10), (1, 1), (10, 1), (100, 10)\}$ to represent Low, Medium and High initial trust, respectively. The priors with the same α/β ratio indicate high and low uncertainty. For the Gaussian distributed prior, we will use $\mu_0 = \{0, 0.5, 1.0\}$ to represent Low, Medium and High initial trust, respectively. We also choose $\sigma_U = \{0.1, 1\}$ to represent Low and High initial uncertainty, respectively.

V. Results

To illustrate our model, we take two different types of agent behavior in ELICIT. These behaviors are given in Figure 2a (which is an actual behavior from ELICIT) and 2b (which is generated by us to show a distinct behavior). In evidence 1 from Figure 2, the most useful information arrives in the beginning of the simulation but then agent y sends a great deal of redundant information not knowing what agent x has received. Willingness evidence is medium (around 0.6) in most of the experiment, but more or less stable. We compare this with evidence 2 from Figure 2 where the node y exhibits erratic behavior towards node x. Both willingness and competence fluctuate throughout the experiment. Note that willingness is a relative measure and is computed by observing all the agents node x interacts with. We now plot competence and willingness as a function of these evidences and several different priors.

First, we study the impact of these different behaviors on competence as a function of time for different priors in Figure 3. For evidence 1, as over time less and less positive evidence is received, the trust will decrease. However, if there is more prior information, then the expected value of trust with respect to competence will vary less slowly than cases with little prior information. We note that the strength of the initial impression can be modeled by different priors. In case of a fast heuristic like facial expression [18], the prior may have lower strength, whereas a competence measure based on a prior social relationship (such as trust for a superior) has much higher strength and is expected to evolve much more slowly. We see that the case $(100, 10)$ has prior evidence that is strongly biased towards high trust. However, evidence 1 is opposite to this prior trust. Therefore, competence varies much less rapidly than the case where $(10, 1)$. In contrast, for Evidence 2, the evidence is more variable and the final expected competence is around 0.5 in most cases of priors.

In both cases of evidence, we expect that the uncertainty gets lower over time as more information becomes available. In Evidence 2, the uncertainty remains slightly higher for each prior, and the uncertainty is slightly higher for cases where there is less prior information and the evidence is opposite of the prior trust. In Evidence 1 and the prior of $(10, 100)$, the evidence is strongly negative. Even in the presence of highly positive information, competence changes slowly. In such cases, the uncertainty remains always low. Despite all the positive evidence, the trustee still strongly believes in low trust. However, in case of $(1, 10)$ indicating initial low trust, the evidence contrary to this prior initially increases uncertainty, which then slowly decreases. Note that if we only considered whether evidence is positive or negative, we fail to capture the relationship between these two nodes properly. Even in the case of Evidence 1, node y is supplying node x with a lot of information. This in itself may be useful in concluding that there is no new information circulating in the system and in fact y can be trusted to send any information it receives. To see this, we plot the willingness trust as a function of time based on both Evidence 1 and 2 for several prior distributions in Figure 4. We compute willingness as discussed before. We also tune the user defined parameter σ_U. Note that higher values of σ_U suggests that the user is more likely to accept information that does not conform with the prior trust. We look at the observed standard deviation of willingness σ_{WE} in the experiment. Higher values of σ_{WE} means that the information regarding willingness is unreliable, lower values signal the opposite. Hence, we assume the user adjusts σ_U by learning from the actual behavior by $\sigma_U = \sigma_U + \alpha(1/\sigma_{WE})$ by a learning factor α.

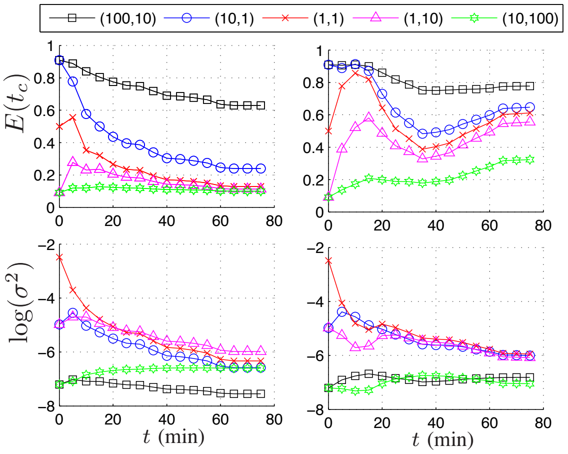

Fig. 3. Evolution of expected values (Ex, top) and uncertainty (log (s), bottom) of competence for various prior distributions (α, β) for Evidence 1 (L) and Evidence 2 (R)

The same trend exists where if the uncertainty is high, the expected willingness trust will consider the evidence with greater weight compared to the situation with low uncertainty. We see greater variability in the expected willingness values for Evidence 2, which in turn may impact the way the node behaves towards a node supplying this type of evidence. Next, we vary the communication delay in the communications of factoids between agents and determine what impact this delay has on trust during these experiments. We consider a single experiment where nodes are assigned individual message latencies varying from 8 seconds to 5 minutes. We consider the willingness trust from the perspective of one node towards a node with one of the assigned message delays values (8s, 30s, 1min, 2min, 5min). We expect that willingness will decrease as the message transmission delay is increases. This is shown in Figure 5. In Figure 5a, the expected willingness trust diminishes as message delay is increased as expected. Further, the uncertainty of willingness is the opposite, as nodes with higher delay have higher uncertainty since less raw evidence has been gathered for that particular link, which is shown in Figure 5b. With the evaluation of trust with

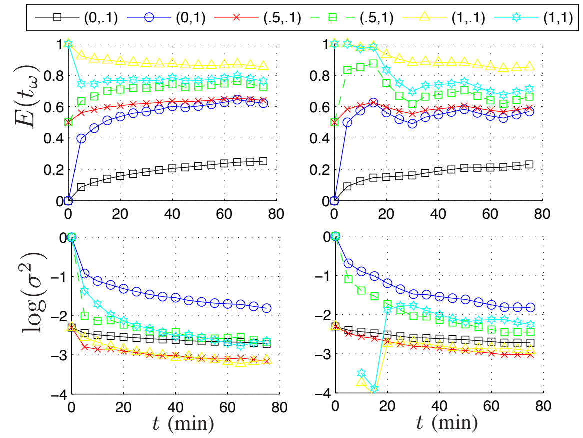

Fig. 4. Evolution of expected values (Ex, top) and uncertainty (log (s), bottom) of willingness for various prior distributions (μ_0, σ_U^2) for Evidence 1 (L) and Evidence 2 (R).

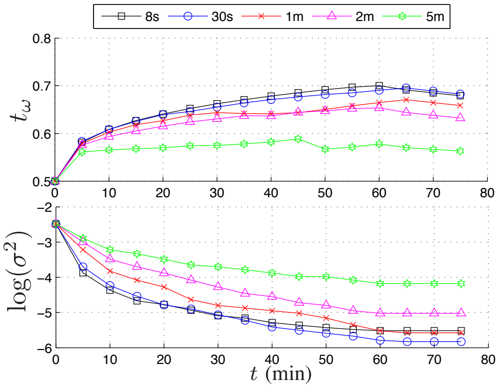

Fig. 5. a) Willingness trust for several nodes with varying transmission delay. b) Willingness uncertainty for several nodes with varying transmission delay.

regard to both willingness and competence, each node or agent can assign its neighboring nodes one of the trust categories described in Section III. Given this knowledge of its neighbors, an agent can modify its behavior in interacting with other nodes. An agent can choose to prioritize incoming messages from nodes with higher evaluated trust. Additionally, an agent can choose to reciprocate high competence or willingness by sending messages to those nodes whom have higher trust. Conversely, it can ignore any nodes with a low trust evaluation.

VI. Conclusions

We have shown a new model for trust that incorporates two measures: competence and willingness to capture different measures of how the trustee can be trusted in a specific team based task. We argued that one of the measures alone is not sufficient to capture the relationship between the agents. This new trust model allows us to explore various trust related behaviors in time critical missions and study the impact of trust in situation awareness.

There are a number of extensions of our work that we intend to study in future work. In the current setting, we note that most nodes end up in the Trusted non-discriminating case due to the way the agents are programmed. The new factoids are seeded randomly into the network within the first 10 minutes. Each node receives redundant copies of messages from its neighbors. Therefore, the competence trust for each of the nodes does degrade over time. The agents are configured to forward any message that affects any of the mental model of situation awareness. Therefore, redundant transmissions may occur in that other nodes may have already received the useful factoid through another node. However, using trust, agents are able to calibrate who they send messages to. In future work, we would like to study the impact of such reciprocating behavior in the system performance and which method is more robust the communication network delays. Note that the traces we used are for decisions made over a long period of time and do not allow us to study the impact of cognitive overload. In the future, we will also explore the use of some models developed in cognitive science literature [12] for decision scenarios involving quick decision making.

Another interesting direction is to analyze the network when the trust is distributed differently. This is especially important for willingness which is a relative measure. Hence, we can study the impact of having small but well connected groups of trusted individuals vs. random distributions of trust.

Acknowledgment

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] C. A. Bolstad and M. R. Endsley, "Tools for supporting team sa and collaboration in army operations," *Collaborative Technology Alliances Conferences: Advanced Decision Architecture Conference*, 2003.

[2] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, 1995.

[3] S. Saavedraa, K. Hagerty, and B. Uzzi, "Synchronicity, instant messaging, and performance among financial traders," *Proceeding of the National Academy of Sciences (PNAS)*, 2011.

[4] M. Granovetter, "The impact of social structure on economic outcomes," *The Journal of economic perspectives*, 2005.

[5] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, 2004.

[6] X. Fan, S. Oh, M. McNeese, J. Yen, H. Cuevas, L. Strater, and M. Endsley, "The influence of agent reliability on trust in human-agent collaboration," in *ECCE'08*.

[7] D. J. McAllister, "Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations," *Academy of Management Journal*, 1995.

[8] K. McCabe, D. Houser, L. Ryan, V. Smith, and T. Trouard, "A functional imaging study of cooperation in two-person reciprocal exchange," *Proceedings of the National Academy of Sciences*, 2001.

[9] D. McKnight and N. Chervany, "The meanings of trust," *Trust in Cyber-Societies-LNAI*, 2001.

[10] J. Cho, A. Swami, and I. R. Chen, "A survey on trust management for mobile ad hoc networks," *IEEE Comm. Surveys and Tutorials*, 2010.

[11] L. J. Chang, B. B. Doll, M. van't Wout, M. J. Frank, and A. G. Sanfey, "Seeing is believing: Trustworthiness as a dynamic belief," *Cognitive Psychology*, 2010.

[12] J. Rudoy and K. Paller, "Who can you trust? behavioral and neural differences between perceptual and memory-based influences," *Frontiers in Human Neuroscience*, 2009.

[13] A. Josang and R. Ismail, "The beta reputation system," *Proceedings of the 15th Bled Electronic Commerce Conference*, 2002.

[14] Y. Wang and M. Singh, "Evidence-based trust: A mathematical model geared for multiagent systems," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 2010.

[15] D. S. Alberts and R. Hayes, "Power to the edge," *DoD Command and Control Research Program*, 2003.

[16] M. Ruddy, "Instantiation of a sensemaking agent for use with elicit experimentation," in *Proceedings of the 14th International Command and Control Research and Technology Symposium, Washington, DC*.

[17] K. Chan and N. Ivanic, "Connections between communications and social networks using elicit," in *Proceedings of the 15th International Command and Control Research and Technology Symposium, Santa Monica, CA*.

[18] J. Willis and A. Todorov, "First impressions: Making up your mind after a 100-ms exposure to a face," *Psychological Science*, 2006.