A Socio-cognitive Approach to Modeling Policies in Open Environments

Tatyana Ryutov
Information Sciences Institute
University of Southern California
tryutov@isi.edu

Abstract

The richness of today's electronic communications mirrors physical world: activities such as shopping, business and scientific collaboration are conducted online. Current interactions have become a form of social exchange where participants must deal with complexity, uncertainty and risk.

We propose a policy specification approach that combines the social sciences and trust theory to facilitate ad-hoc interactions of self-interested parties in open environments. Our socio-cognitive approach allows us to reason about uncertainty and risk involved in a transaction, and automatically calculate the minimum **trust threshold** *needed to mitigate the vulnerabilities. The trust threshold comprises the core of security policies that govern the interactions. The threshold calculation is based on balancing objective and subjective trust components, which together guarantee that a transaction will result in an acceptable outcome. We propose to apply the Prospect Theory [20] to specify policies that determine a set of acceptable outcomes. We present the trust threshold negotiation primitives.*

1. Introduction

Early computer mediated interactions have tended to be one-way. Typically, employees had a privilege to access resources owned by an organization. The risk resided on the side of the resource provider and, therefore, security policies took into account only interests of the resource owner. The traditional security approaches were designed for closed, fixed environments where:

- a resource provider is completely trusted;
- user population is fairly small and fixed;
- only authorized users can access resources;
- authorization is based on authenticated user identity;
- untrusted users gain trust once authenticated.

In open environments, these assumptions often do not hold. A resource provider can betray trust by, for example, selling low quality products, or by disclosing personally identifiable information to others against the wishes of its customer. User population evolved from single organization to an open world. Maintaining identifiers and accounts for all potential users is not practical. Authenticating the identity (in traditional sense) of a stranger may not provide sufficient information for access control purposes.

Current interactions involve mutual exchange of resources that each party controls and values; as opposed to a one-sided nature of early communications. For example, users provide credit card numbers in exchange for goods or services. Online collaborations require exchanging resources and knowledge. Today's online interactions are effectively a form of social exchange where the distribution of risk between the parties is often symmetric. Handling risky mutual exchanges and establishing trust in open ad hoc environments are the new challenges of access control and authentication. Risk is a perception of uncertainty and adverse consequences of engaging in an activity [12]. A social exchange is broadly defined as an interaction in which one party is obligated to satisfy particular requirements, usually at some cost, in order to receive benefits from the other party. A risky transaction involves a social dilemma [21]: each party makes a choice to cheat, cooperate or not participate in the exchange. Therefore, these exchanges necessarily entail uncertainty and risk: the other party may take the benefits without satisfying the requirements.

To address the increased complexity, uncertainty and risks of online interactions, next generation security policies must explicitly model a way people interact in social environments when faced with these issues. The main contributions of this paper are:

1 application of assumptions and concepts from social sciences to modeling trust and risk in policies;

2 automatic calculation of the minimum trust threshold needed to mitigate the vulnerabilities in risky transactions;

3 facilitation of trust negotiation by providing a proposal construction, evaluation and modification mechanisms.

2. Trust and Social Exchange

Trust is a complex notion with many meanings. Introduction of a computationally usable notion of trust requires a simplified and focused definition. We adopt the following definition of trust for the purpose of our research*: trust is a decision to accept vulnerability (participate in an exchange) faced with positive or negative outcomes of an exchange which depend on the actions of the opponent.*

The social exchange theory posits that trust is more likely to develop between partners when exchange occurs without explicit negotiations or binding agreements (reciprocal exchanges). Under these conditions, the risk and uncertainty of exchange provide the opportunity for partners to demonstrate their trustworthiness. In reciprocal exchanges, the only form of assurance comes from the expectation of future interactions. For example, a corporation would share new product sketches with a prospective partner in order to obtain valuable input if it expects extensive future involvement. However, if future cooperation with the partner was in question, the firm would withhold the design information, which will be considered sensitive.

The most problematic interactions are the transient exchanges between strangers, one-shot interactions that do not necessarily repeat in the future. Building trust in such situations is a difficult matter. We believe that this requires establishing sufficient *subjective* and/or *objective* trust levels as the basis for a rational decision to engage or not to engage in a transaction; and in order to agree on security related terms of the exchange.

Subjective (or perceived) trust encompasses trusting attitude due to perceived qualities or abilities (e.g., reputation, skills, and profiles) of the other. During the subjective trust negotiation stage both parties try to acquire information that assures it of the likelihood of appropriate partner behavior with respect to resources involved in the transaction.

Objective trust means that one has formed an intention to trust (participate in exchange) irrespective of beliefs about the qualities of the other party due to any or all of the following:

1) it is economically not profitable for the other party to cheat because of severe penalties in the case of non compliance;
2) if trust is misplaced, some compensation mechanism will mitigate the vulnerability (e.g., insurance).

During the objective trust establishment, parties engage in a joint decision process, such as explicit bargaining, in which they reach an agreement on the terms of the exchange.

Subjective and objective trust types are complementary to each other. Subjective trust can fail to cope with the overwhelming uncertainty, for example, when no reliable reputation is available. In this case, objective trust can be used to constrain the interaction through certain mechanisms and reduce the uncertainty. On the other hand, the costs of safeguarding against untrustworthy behavior could be very high. In such cases, building subjective trust is important. Developing an approach to balance these two types of trust is the main objective of our research.

3. Trust Threshold

In order to decide whether to participate in a risky transaction, a certain level of trust (trust threshold) must be reached [24]. Cooperation is possible when the level of trust for the other exceeds a *minimum trust threshold* for each party. The difficult question is how can one automatically calculate the required trust threshold based on characteristics of interaction? Requiring too much trust may place unnecessary restrictions on the transaction or discourage a legitimate transaction. Acquiring too little trust may result in unjustified vulnerabilities. To our best knowledge no adequate solution to this problem exists yet.

In literature, trust is often represented by a set of discrete values (e.g., high/medium/low), or numerical intervals (e.g., [0, 1]) and the required trust threshold is specified in security policies. For example, [33] verifies that the trust level is greater than a certain number, [35] checks that at least a minimum number of evidence statements is presented. However, all of these thresholds are statically defined by the policy and there is no run-time evaluation of risk.Trust is very contextualized and nuanced in dynamic environments,

therefore fixed threshold assignment is not adequate. We strongly believe that the level of trust needed in a particular situation must be analyzed by performing run-time evaluation of risk, and then negotiate trust to reach the desired trust level. The question of determining a trust threshold becomes the question of:
1) finding all acceptable outcomes;
2) determining the set of attributes and rules which guarantees that the exchange will result in an outcome from the set of acceptable outcomes.
Therefore, a measure of trust is essentially a probability of a particular outcome [17].

4. Formal Definitions

From our perspective, an exchange involves thee phases:

Initial. Potential participants perform initial evaluation of an exchange. A decision process proceeds as follows:

- Determine what items are to be contributed by each party;
- Determine a set of matters of concern (e.g., quality, timeliness, etc.) for each item to be contributed or received;
- Calculate possible outcomes of the exchange in terms of gains and losses;
- Apply access control policy to find a set of acceptable outcomes;
- Determine a set of subjective/objective trust metrics (trust threshold) which guarantees the acceptable outcomes.

Negotiation. During this stage, the participants negotiate trust thresholds using private negotiation strategies. This process can be iterative. For example, during the trust negotiation one may realize that there are additional risks introduced by the mitigating measures used to build objective trust, and the risk/cost/benefit balance does not hold anymore. So, it may be necessary to repeat the steps of the Initial phase during the negotiation.

Final. After the exchange completes, the participants evaluate the actual outcomes of the exchange. The interaction history is updated according to the subjective appreciation of how well the issues were handled by the opponent.

We next introduce formal definitions and describe the cognitive process employed during the three phases.

4.1 Exchange and Outcome Representations

At a high level an exchange between participants *a* and *b* can be viewed as a function *Φ:*

$$\Phi(R^a, R^b, Y) \to (O^a, O^b) \quad \textbf{(1)}$$

which maps two sets of items contributed by each participant and the context of the exchange *Y* to outcomes observed by each party.

We define formal concepts from the point of view of a participant *a*. Since the situation is symmetric, we omit a similar set of concepts defined from the point of view of *b* for brevity.

R^a is a set of items controlled by a participant *a,* which *b* expects to receive after the exchange completes. Similarly, R^b is a set of items controlled by *b* which *b* must provide to *a*. An item can be a product, service, knowledge/information or money. Trust is context specific. Experiences in one environment not always can be directly converted to judgments in other environment; therefore the notion of context *Y* is critical.

To simplify the formal presentation, without the loss of generality, we consider an exchange that involves just two items: $\Phi(r^a, r^b, Y) = (O^a, O^b)$. For each item $r \in \{r^a, r^b\}$ to be exchanged, a participant may associate a finite set of matters of concern: $j(j = \{1, \ldots n\})$, for example, non-delivery/partial delivery, timeliness, quality, confidentiality, etc.

The domain of values taken by an issue j associated with an item r is noted as $X_j^r = [\min_j^r, \max_j^r]$. We constrain each domain of values with a delimited range of reasonable values (e.g., for an issue that represents timeliness of item delivery the domain of values can be: min=0 days, max=356 days).
Let $x_j^r \in X_j^r$ be a vale for issue j.
Each party has a set of scoring functions $S = \{s_1() s_2(), .. s_n()$. Each function $s_i()$ from this set maps the observed value that a particular matter takes to a *satisfaction rating*. The rating denotes a subjective degree of satisfaction with the performance of an opponent on a particular issue. The satisfaction rating is perceived in terms of the variation between the expected (advertised or negotiated) value for the issue and the executed one.

We use fuzzy sets to represent the domain of values to rate the degree of satisfaction with each matter. We assume that each participant defines a small set of linguistic labels to qualify the performance of an opponent on each matter. For example, to represent a degree of satisfaction with a matter *j* that denotes quality or timeliness of an item, we can define the domain of values to be fuzzy sets, such as *L= {l_1=excellent, l_2=good, l_3=average, l4= bad, l_5=very bad}* with membership functions μ_{l_i}. The membership function μ_{l_i} maps the universe of discourse (the domain of values $X_j^r = [\min_j^r, \max_j^r]$ taken by the issue *j* associated with an item *r*) to a degree of membership between 0 and 1 in the set l_i [13]. The satisfaction rating is represented by a vector consisting of the degrees of belongingness to each fuzzy set:

$$s_j(x_j^r) \rightarrow < \mu_{l_1}(x_j^r), \mu_{l_2}(x_j^r), \ldots, \mu_{l_n}(x_j^r) >,$$
$$s_j \in S, l_{1,m} \in L \quad \textbf{(2)}$$

A particular outcome O^a for participant *a* is a set of satisfaction ratings with the *b*'s performance on each matter. The satisfaction ratings are defined for each item to be exchanged. An outcome O^a is noted as:

$$O^a = \bigcup_{r \in \{r^a, r^b\}} \{< \mu_{l_1}(x_1^r), \ldots, \mu_{l_n}(x_1^r) >, \ldots,$$
$$< \mu_{l_1}(x_n^r), .., \mu_{l_n}(x_n^r) >\} \text{ where}$$
$$l_m \in L, x_j^{r_i} \in X_j^r \; \textbf{(3)}$$

Note that the sets of outcomes O^a and O^b are not necessarily equal, because participants may have different private interests and can assign a different set of matters depending on whether an item is to be sent or received. Furthermore, the judgment of how well an opponent performed on each issue is subjective.

The set of all possible outcomes Θ^a is infinite, even though the number of items to be exchanged and the sets of matters associated with each item in real life are always finite. This is because a membership function μ may take any value in the interval [0,1], thus permitting the gradual assessment of the membership of elements in relation to a set.

4.2 Outcome Evaluation

A particular outcome may have positive (gain) or negative (loss) consequences for each participant. We next define an outcome evaluation function.

Assume that a resource r^b has a subjective value $V(r^b)$ to the participant *a*:

$$V(r^b) = k \cdot V(r^a) \; \textbf{(4)}$$

Where $V(r^a)$ is a renouncement value of the resource contributed by *a* representing *a*'s investment in the exchange;

k denotes desirability of the resource r^b and indicates the willingness to contribute resource r^a in exchange for r^b. The value is measured with an abstract unit – a real number that can be mapped to other units (e.g., monetary value).

The function $C(O^a)$ denotes the consequences of the exchange in terms of gains and losses as perceived by the participant *a*.

$$C(O^a) = k \cdot V(r^a) \cdot \Omega - V(r^a) - \Psi \quad \textbf{(5)}, \quad \text{where}$$

$$\Omega = \min\{1, \sum_{j=1,n} \bar{W}_j^{r^b} \sum_{l_k \in L} W_{l_k}^{r^b}(l_k) \cdot \mu_{l_k}(x_j^{r^b})\}$$

$$\Psi = \sum_{j=1,n} \bar{W}_j^{r^a} \sum_{l_k \in L} W_{l_k}^{r^a}(l_k) \cdot \mu_{l_k}(x_j^{r^a}) \text{ where:}$$

Ω represents a weighted cumulative satisfaction rating with the values that the matters (associated with an item contributed by *b*) take.

$W_{l_k}^{r^a, r^b} : L \rightarrow [-1,1]$ denotes the relative importance of the fuzzy sets. In other words, it indicates how important it is for a satisfaction value with an issue *j* to belong to a particular fuzzy set. For example, if the delivery time is not very important for a user, the user can assign high weights (close to 1) to sets "excellent", "good", "average" and lower weights to the sets "bad" and "very bad". Note that the weights can take negative values. This is to express situations when dissatisfaction with an issue has a dramatic effect on the result of an exchange.

$\bar{W}_j^{r^b} : j \rightarrow (0,1]$ establishes the relative importance of a particular matter *j* associated with an item r^b. For example, if the quality of an item is more important than the delivery time, the higher weigh is assign to the matter that indicates the quality of an item. Note that the value is greater than 0 because we require that only issues which matter are considered.

$\bar{W}_j^{r^a} : j \rightarrow (0, \eta]$ establishes the relative importance of a particular matter *j* associated with an item r^a. Note that the weight can take a positive number no grater than *η*, therefore the cost of non satisfaction of issues associated with the contributed items may vary greatly. For example, a participant may associate a very high cost with a disclosure of sensitive information provided by the participant during the exchange. Function $C(O^a)$ returns a negative value (loss) if:

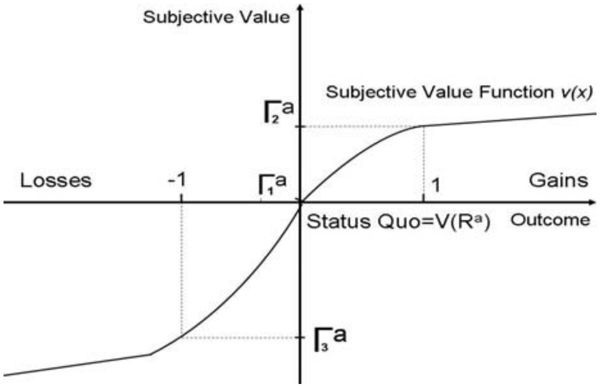
1) the value of item r^a contributed by a is greater than the value of item r^b contributed by b either because b provides less valuable item from the start (k<1); or because b handles some matters associated with the items unsatisfactorily, therefore the value of the item is reduced;
2) b handles some important matters associated with the items r^a contributed by a unsatisfactorily, therefore the value of the exchange is diminished;
3) all of the above.

Otherwise, $C(O^a)$ returns a non negative value. A positive value corresponds to gains, 0 indicates *status quo*. The status quo is assumed if either a participant decides not to participate in the exchange (and keeps r^a), or if the value of the outcome of the exchange is 0 (e.g., k=1, Ω=1, $\Psi = 0$ in Formula 5).

4.3 Exchange Policy

In the initial and negotiation phases of an exchange, a predicts and evaluates outcomes for an exchange under consideration. This is done using the Formula (5) that determines the value of an outcome in terms of gains and losses. Note that it is not enough to consider whether an outcome has a positive or negative expected gain. How much a participant can afford to loose is also important.

To determine the actual subjective value of an outcome, we use the value function $v(x), x = C(O^a)$ from the Prospect Theory [20]. The function is depicted in figure below.



Prospect theory provides a descriptive model of decision making under risk. The consequences are viewed in terms of changes from a reference point (usually the status quo, e.g., current wealth). The values of the outcomes for both positive and negative consequences are assigned to gains and losses rather than to final assets. The resulting value function is steeper for losses than for gains. The curve is concave for gains and convex for losses, implying that decision makers are risk averse when choosing between gains and risk seeking when choosing between losses. The function is constructed as follows [20]:

$$v(x) = \begin{cases} f(x) & if \quad x > 0 \\ 0 & if \quad x = 0 \\ \lambda * g(x) & if \quad x < 0 \end{cases}$$

where *f(x)* and *g(x)* are defined as follows:

$$f(x) = \begin{cases} x^{\alpha} & if \quad \alpha > 0 \\ \ln(x) & if \quad \alpha = 0 \\ 1-(1+x)^{\alpha} & if \quad \alpha < 0 \end{cases} \quad g(x) = \begin{cases} -(-x)^{\beta} & if \quad \beta > 0 \\ -\ln(-x) & if \quad \beta = 0 \\ (1-x)^{\beta}-1 & if \quad \beta < 0 \end{cases}$$

Parameters:

α: power for gains; 0.88
β: power for losses; 0.88
λ: loss aversion; 2.25

Risk attitudes are relative to individual traits of each partner, so loss aversion parameter λ will be different for different agents.

To determine a set of acceptable outcomes an access control policy A^a of the participant a is applied to determine the ***minimum acceptable value*** Γ^a of the exchange: $A^a(\Phi, k) \to \Gamma^a$.

Defining such a policy is our research issue. It is natural to assume that $\Gamma^a = 0$. However, the policy should not be purely economic in nature. It should take social exchange considerations into account as well. In particular, expectation of reciprocal exchange may allow for costs to outweigh the benefits (at this point of time) in the hope of receiving benefits from future interactions with the same entity. The policy should take into account the desirability of the resources (expressed by k) to be received in the exchange. If a resource is highly desirable, a participant may accept a negative value for the exchange. On the other hand, if the resource is not wanted much (e.g., because it is offered by other parties), the minimum value of an exchange can be $\Gamma^a = n, n > 0$.

For example, in the figure minimum acceptable value Γ_1^a requires that an outcome value be no less than 0, Γ_2^a requires gains no less than $V(r^a)$, and Γ_3^a tolerates loss no grater than $-V(r^a)$.

4.4 Trust Threshold Representation

Subjective trust is represented as a set of validated attributes (in practice - verified certificates) which describe various trust related aspects of a participant:

$$T^S = \{a_1, \ldots, a_n\}.$$

Objective trust is represented as a set of Boolean formulas $B = \{B_1, \ldots, B_p\}$built from elements called *conditions* and an associated set of *sanctions* $T^O = \{<B_1(c_1, \ldots, c_k), \{s_1, \ldots, s_m\}>, \ldots, <B_p(c_1, \ldots, c_l), \{s_1, \ldots, s_n\}> \}$.

Conditions describe issue-value comparisons for a particular resource *r*:

$$c_i = \{j_1^r \otimes x_1^r, \ldots, j_k^r \otimes x_k^r\}, \otimes \in \{=, >, \geq, <, \leq\}.$$

If specified conditions are met (the corresponding Boolean formula evaluates to *true*), then the associated set of sanctions must be carried out by the enforcement mechanisms.

A trust threshold $T_i = T_i^O \bigcup T_i^S$ predicts an exchange to result in an outcome with the value greater or equal to the minimum acceptable value:

$$v(C(O^a)) \geq \Gamma^a \textbf{ (6)}$$

The set of all acceptable thresholds is denoted as: $T = T_1 \bigcup \ldots \bigcup T_n$.

4.5 Determining Trust Threshold

To calculate the minimum trust threshold, we need a reasoning/belief model which maps a set of attributes (subjective trust), the set of conditional sanctions (objective trust), and other contextual information to a particular outcome. In real life people learn how much to trust from experiences and generate a set of mental trust rules which are then applied to each new situation. We will imitate this process for the purpose of finding a trust threshold by employing a neuro-fuzzy approach. A neuro-fuzzy system is a fuzzy system that uses a learning algorithm derived from neural network theory to determine its parameters (fuzzy sets and fuzzy rules) by processing data samples (training data). After the learning process completes, the resulting system can be interpreted in a form of fuzzy rules like:

IF **x1** is A1 and **x2** is A2
and **x3** is A3 and **x4** is A4
THEN the pattern (**x1,x2,x3,x4**) belongs to **class CL$_1$**,
where A1 - A4 are linguistic terms (e.g. small, medium, large) represented by fuzzy sets.

The constructed fuzzy rule base represents the relationships between a context of an exchange, negotiated objective and subjective trust, and the observed outcome. We codify an input to the system as a set of variables of three types which describe an exchange:

1. *Context* represents a context of exchange, e.g., on-line auction, scientific collaboration, history of prior interactions, etc.
2. *Attributes* represent the attributes of an opponent, e.g., reputation, location, skills.
3. *Mechanisms* indicate enforcement mechanisms used in the exchange, e.g., insurance, return policy, secure payment with PayPal.

We next codify the output classes CL_i to represent a set of outcomes. As discussed in section 4.1, the set of all possible outcomes Θ^a is infinite. In order to use the neuro- fuzzy system for the trust threshold calculation, we need to convert Θ^a to a small finite set of outcomes.

$$< \mu_{l_1}(x_j^r), \ldots, \mu_{l_n}(x_j^r) > \rightarrow < \sigma_1, \ldots, \sigma_n > \textbf{ (7)}$$

To achieve this, we convert continuous degree of membership to crisp one by replacing a membership number $\mu_{l_i}(x_j^r)$ with 1 if $\mu_{l_i}(x_j^r) > \mu_{l_j}(x_j^r)$ for $\forall j, j = \overline{1,n}, j \neq i$ and with 0 otherwise. If the highest membership value is not unique, we resolve the conflict by assigning 1 to a component of the vector with the highest index. This conversion reduces the size of all possible outcomes to n^m, given *n* fuzzy sets and *m* issues. We use these converted vectors to represent the outcomes related to each issue.

After the rules are generated we examine them to extract the trust threshold for a particular exchange as follows:

1. We select a set of fuzzy rules *F* where rule *antecedent* contains variables of the type "context" which match the context of current exchange.
2. From the set *F* we construct a subset *F'* by selecting rules where the rule *consequent* represents an acceptable outcome: $v(C(O^a)) \geq \Gamma^a$. Note that some of the mechanisms that enforce objective trust may introduce additional costs (e.g., insurance). This overhead is included into the $V(r^a)$ component of the outcome evaluation formula (5).
3. Next, for each fuzzy rule f_i from the set *F'* we construct a trust threshold $T_i = T_i^O \bigcup T_i^S$ by extracting a set of values of the type "attributes" (subjective trust T_i^S) and of the type "mechanisms" (objective trust T_i^O).
4. Finally we construct set *T* of all acceptable thresholds by taking a conjunction of the sets

constructed during the previous step: $T = T_1 \cup ... \cup T_n$

Another approach to define necessary objective trust level is using norms [8]. For example, all retailers must agree to a 30-day return policy on all items they sell; buyers must pay for the goods before they can be delivered. To support this, we need to (i) define a language to represent norms and (ii) define a mechanism for finding corresponding norms to mitigate non-plausible outcomes. For example, if predicted outcome indicates low product quality, the applied norm should guarantee the ability to return the product.

4.6 Negotiation

The purpose of the negotiation is for each party to reach its private trust threshold. Since there may be a choice of acceptable thresholds ($T = T_1 \cup ... \cup T_n$), different trust negotiation strategies are possible. The realization of this step is based on the agreement negotiation mechanism which employs the concepts of obligations and expectations. During the negotiations, parties exchange offers (agreement proposals) and counter-offers until the agreement is reached (private trust thresholds are reached by both parties) or failure is reported.

Obligations - an entity would be willing to participate in an exchange under an agreement with another participant. A member commits itself to provide an item under certain terms (expressed as the issue-value assignments) to another member. An obligation represents demanded resources to be contributed and the values that the issues must take.

Expectations – a wish list comprising of a set of obligations that the other entity must undertake as part of the agreement.

An agreement between members *a* and *b* is reached when expectations of one member are met by the obligations of the other member, and vice versa. A participant *a* enters a negotiation with a participant *b* with an agreement proposal $P_{a,b}$:

$$P_{a,b} = < E_{a,b}\, , O_{a,b}\, , \{\cancel{a}^{a}{}_{k}\}, \{a^{b}{}_{l}\} > \quad \textbf{(8)}$$

The agreement proposal consists of the following:

1. Expectations $E_{a,b}$ expressed as a set of resources $R^b = \{r_1^b, ..., r_d^b\}$ to be contributed by *b*, a set of issue-value assignments for each resource $I^{r_i^b} = \{j_1^{r_i^b} = x_1^{r_i^b}, ..., j_k^{r_i^b} = x_k^{r_i^b}\}$, $I^{R^b} = \{I^{r_1^b}, .. I^{r_d^b}\}$ and a set of conditioned sanctions in the case of non-performance expressed as a set of Boolean formulas over a set of conditions, and associated sanctions
$$K_{a,b} = \{< B_1(c_1, ..., c_s), \{s_1, ..., s_g\} >, ...,$$
$$< B_f(c_1, ..., c_w), \{s_1, ..., s_h\} >\}$$
$$E_{a,b} = \{R^b, I^{R^b}, K_{a,b}\}$$

2. A possibly empty set of obligations $O_{a,b}$ expressed a set of resources $R^a = \{r_1^a, ..., r_d^a\}$ to be contributed by *a*, a set of issue-value assignments for each resource $I^{r_i^a} = \{j_1^{r_i^a} = x_1^{r_i^a}, ..., j_k^{r_i^a} = x_k^{r_i^a}\}$, $I^{R^a} = \{I^{r_1^a}, ... I^{r_d^a}\}$ and conditioned sanctions expressed as a set of Boolean formulas over a set of conditions, and a set of associated sanctions.
$$K_{b,a} = \{< B_1(c_1, ..., c_s), \{s_1, ..., s_g\} >, ...,$$
$$< B_f(c_1, ..., c_w), \{s_1, ..., s_h\} >\}$$
$$O_{a,b} = \{R^a, I^{R^a}, K_{b,a}\}$$

3. A possibly empty set of verified attributes $\{\cancel{a}^{a}{}_{k}\}$;
4. A possibly empty set of attributes $\{a^{b}{}_{l}\}$ requested by *a* from *b*.

4.6.1 Construction of Initial Proposal

In the initial step of the decision making process an access control policy A^a of the participant *a* is applied to determine the minimum acceptable value of the exchange: $A^a(\Phi, k) \rightarrow \Gamma^a$.

To construct an *initial* agreement proposal, the party next determines the set of issues for each resource and creates the issue-value assignments to ensure that the predicted outcome satisfies the exchange policy: $\nu(C(O^a)) \geq \Gamma^a$.

The next step is to determine the trust threshold for the exchange. The threshold is calculated from the fuzzy rule base as described in Section 4.5.

The participant next chooses the "best" trust threshold $T_i = T_i^O \cup T_i^S$ from the set *T*. It next constructs expectations to match the objective trust component $K_{a,b} = T_i^O$ of the selected trust threshold. The subjective trust component T_i^S is used to request additional attributes (some attributes could be public and available to the participant before negotiation). The "best" trust threshold is chosen based on private participant's interests and goals. For example, an entity can attempt to reach a trust threshold that does not include expensive enforcement mechanisms which incur additional cost. This will most likely require building a “strong” subjective trust.

4.6.2 Proposal Evaluation

To evaluate a proposal $P_{b,a} = < E_{b,a}\ , O_{b,a}\ , \{\cancel{a}^{b}{}_{k}\}, \{a^{a}{}_{l}\}>$ received from an opponent, participant *a* checks whether the obligations fulfill the expectations. That is: $O_{a,b} = E_{b,a}$ and $E_{a,b} = O_{b,a}$.

If this is true, *a* checks whether the set of verified attributes $\mathcal{A}^b$ obtained from trusted public sources before the negotiation and provided by *b* during the negotiation, and the set of obligations *b* is willing to commit, match one of the acceptable thresholds. If this is true, the agreement is reached. If obligations/expectations do not match because either or all of the following:

- The set of resources to be contributed by *b* is a subset of the resources requested by *a*:
  $R^b \supset R^{b}{}', R^b \in O_{b,a}, R^{b}{}' \in E_{a,b}$
- The set of issue/value assignments declared in *a*’s expectations do not match the value-issue assignments defined in the *b*’s obligations
  $I^{R^b} \neq I^{R^b}{}', I^{R^b} \in O_{b,a}, I^{R^b}{}' \in E_{a,b}$

We need to re-evaluate the proposed assignments using formulas (5) and (6). If the equation (6) does not hold, the proposal is not acceptable and *a* may choose either to create a counter-proposal or decide to stop the negotiation.

4.6.3 Counter Offer Construction

If none of the acceptable thresholds is met, the participant creates a counter-offer with proposed changes to the expectations and obligations. A counter-proposal may extend the initial proposal and/or amend parts of the initial proposal. The changes can include requesting additional attributes and/or modifications to the set of issues and sanctions to match a chosen trust threshold.

4.6.4 After the Exchange

After the exchange completes, each participant employs the outcome construction and evaluation functions (3), (5), and (6) to assess the result of the exchange. The interaction history is updated according to the subjective satisfaction with performance of the opponent on each issue. Note that the subjective judgment is expressed using the continuous membership in the fuzzy sets. However, in order to store the evidence, the continuous scores must be converted to the crisp degrees as discussed in Section 4.5.The new evidence store is used to re-train the neuro-fuzzy system and update the rule base.

5. Implementation Approach

We plan to implement the system and test it in the on-line auction environments. We plan to use Nefclass software system as a core module for calculating a trust threshold. Nefclass [25] implements an algorithm that extracts interpretable fuzzy rules from data. The Nefclass system can be created with or without insertion of prior knowledge. This allows including pre-defined set of policies created by a domain expert. For the training and testing datasets we will use real life data collected from e-bay auction histories. The data will be converted to the Nefclass input format as described in section 4.3.The part of the system that implements the reasoning before and during the negotiation can be integrated with the existing trust negotiation systems, e.g., TrustBuilder [31].

6. Related Work

Early trust management systems such as PolicyMaker [3], KeyNote [4], REFEREE [6], Vigil [19] and Trust-Builder [31] concentrate on credential based distributed policy management. A trust decision is based on a given credential and its issuer. A common flaw with all these approaches is that they employ a static form of trust, supporting only limited trust dynamics in the form of credential revocation. These solutions fail to address fundamental questions such as how to express trust and how trust can be formed and updated.

New approaches attempt to address these shortcomings by modeling explicitly the trustworthiness of entities and supporting its formation and evolution. Although trust is an illusive concept, a number of definitions [27], [19], [22], [32], [2] and trust classifications [14], [18], [9] have been proposed. One of the first works to introduce a computational model for trust was [24]. The model attempts to incorporate all aspects of social trust. This makes the model large and complex and difficult to implement. [27] present an adaptation of Marsh's work to P2P environments.

More recent trust management approaches [1], [27], [5], [32], [22], [33], [10], [11] make decisions based on computational trust and trust/risk analysis rather than on credentials alone. SECURE [5] evaluates risk of every possible outcome of a particular action using a family of cost-PDFs parameterized by the outcomes' costs and benefits. This approach simplifies real life by modeling uncertainty based only on probability. In open environments, the probability distribution of different outcomes of the transaction is not known. It depends on the trustworthiness of the opponent.

[9] introduces a framework that depends on trust metrics, cost and utility parameters to produce a trust policy for a given interaction. The framework is defined at a very high level and lacks a formal definition of the involved concepts.[23] develops a model based on trust-related variables such as the cost of the transaction and its history, and defines risk-trust decision matrices. [17] expands the Manchala's model by refining the relationship between trust and risk: a three-dimensional decision surface for balancing trust and risk is introduced.

A fuzzy logic approach to security was first suggested by [15]. The proposed new computer security paradigm is no longer concerned with a perfect security but rather intents to employ risk management appropriate for particular tasks. [26] introduced a fuzzy version of the Bell-LaPadula model. [7] proposed a fuzzy technique to evaluate a security level for a given policy against a set of reference policy levels. [34] employs fuzzy logic to support federated trust management. In this work, trust is classified as objective and subjective. Subjective trust is similar to our definition of the term. It is based on one's beliefs and, therefore, is uncertain. The notion of objective trust differs from our definition since it represents a type of trust that can be evaluated with certainty, e.g., data provided by an entity can be objectively verified using scientific instruments. This work is complementary to our approach. The derivation rules can be applied to calculate the subjective trust component of our system. [29], [30], [28] use a fuzzy logic based analytical model for multiple-issue, two-party, negotiations in order to alleviate the complexity of negotiation.

7. Conclusions

We presented a new risk/trust balancing approach (based on the concepts from social sciences) to model policies in open competitive environments. We believe that the system will be useful in different environments which include various forms of peer-to-peer interactions between human and artificial agents. In all of the environments listed below the interactions involve mutual exchange of resources; therefore handling uncertainty and addressing social and trust issues is important.

Peer2peer systems. These systems are based on free participation of partners engaged in the production of common resources, often without monetary compensation as a key motivation. However, these voluntary efforts of entities are motivated by the returns they are expected to get from others (reciprocal exchanges).

Socio-cognitive grids. A socio-cognitive grid is an emerging paradigm – an extension of grid computing: utilization of resources from a number of networked computers *and people* to solve a particular problem.

Scientific and commercial collaborations. These settings require on-demand sharing of resources, services and knowledge.

Ad hoc on line trading. This environment includes electronic auctions, swap meets, small scale online business, specifically existing in the absence of trusted regulating institutions.

Semantic web. The proposed system could be useful in semantic web where user agents interact with possibly unknown artificial agents. For example, an agent acting on behalf of a patient could use the semantic web to identify a medical specialist and make an appointment. In this example, the agents must exchange sensitive user information, possibly perform pre-payment (or agree on a fee in the case of non-show – objective trust), verify trustworthiness of the health provider, ensure the necessary medical facilities are in place (subjective trust), etc.

9. References

[1] K. Aberer and Z. Despotovic. Managing trust in a peer 2-peer information system. In Proceedings of the Conference on Information and Knowledge Management, 2001.

[2] F. Almenarez, A. Mar´ın, C. Campo, C. Garc´ıa. A Pervasive Trust Management Model for dynamic Open Environments. First Workshop on Pervasive Security and Trust at MobiQuitous 2004.

[3] M. Blaze, J. Feigenbaum, and A. D. Keromytis. Keynote: Trust management for public-key infrastructures. In Secure Internet Programming: Issues in Distributed and Mobile Object Systems, LNCS. Springer-Verlag, 1998.

[4] M. Blaze, J. Feigenbaum, and J. Lacy. Decentralized trust management. Conference on Security and Privacy, 1996.

[5] Cahill, V., Shand, B., Gray, E., Bryce, C., Dimmock, N.: Using trust for secure collaboration in uncertain environments. IEEE Pervasive Computing 2, 2003.

[6] Chu, Y.H., Feigenbaum, J., LaMacchia, B., Resnick, P., Strauss, M.: REFEREE: Trust management for Web applications. Computer Networks and ISDN Systems,1997.

[7] Casola V., Preziosi R., Rak M., Troiano L. Security Level Evaluation: Policy and Fuzzy Technique, in Proceedings of International Conference on Information Technology: Coding and Computing , ISBN 0-7695-2108-8, 2004.

[8] F. Dignum. Autonomous agents with norms. Artificial Intelligence and Law, 7(1):69–79, 1999.

[9] T. Dimitrakos. A Service-Oriented Trust Management Framework. In R. Falcone, S. Barber, L. Korba, and M. Singh, editors, Trust, Reputation, and Security: Theories and Practice, LNAI 2631, pages 53–72. Springer, 2002.

[10] Dimmock, N., Belokosztolszki, A., Eyers, D., Bacon, J., Moody, K.: Using trust and risk in role-based access control policies. In: Proceedings of Symposium on Access Control Models and Technologies, ACM, 2004.

[11] Dimmock, N. and Bacon, J. and Moody, K. and Ingram, D. Risk models for Trust-Based Access Control (TBAC). In: 3rd International Conference on Trust Management, 2005.

[12] Dowling, G. R., & Staelin, R. A model of perceived risk and intended risk-handling activity. Journal of Consumer Research, 21 , 119-134, 1994.

[13] Faratin, P. and Sierra, C. and Jennings, N. R. Negotiation Decision Functions for Autonomous Agents, Journal of Robotics and Autonomous Systems, 1998.

[14] Grandison, T. and Sloman, M., A Survey of Trust in Internet Applications, IEEE Communications Surveys. Fourth Quarter 2000.

[15] H. Hosmer, "Security Is Fuzzy: Appling The Fuzzy Logic Paradigm to the Multiplicity Paradigm". pp. 175-184, Communication of ACM, 1993.

[16] N. R. Jennings, P. Faratin, A. R. Lomuscio, *S.* Parsons, C. Sierra, and M. Wooldridge. Automated negotiation: Prospects, methods and challenges. International Journal of Group Decision and Negotiation, 2001.

[17] Josang, A. and Lo Presti, S. (2004) Analysing the Relationship Between Risk and Trust. In Proceedings of Second International Conference on Trust Management (iTrust 2004) LNCS 2995, pp. 135-145.

[18] A. Josang, R. Ismail, and C. Boyd. A Survey of Trust and Reputation Systems for Online Service Provision. Decision Support Systems, 2005.

[19] L. Kagal, J. L Undercoffer, F. Perich, A. Joshi, and . Finin. A security architecture based on trust management for pervasive computing systems. In Grace Hopper Celebration of Women in Computing, October 2002.

[20] Kahneman, D., & Tversky, A. Prospect theory: An analysis of decision under risk. Econometrica, 47, 1979.

[21] Kollock, P. The production of trust in online markets. Advances in Group Processes (Vol. 16), 1992.

[22] N. Li and John C. Mitchell. RT: A role based trust management framework. In 3rd DARPA Information Survivability Conference and Exposition, 2003.

[23] D.W. Manchala. Trust Metrics, Models and Protocols for Electronic Commerce Transactions. In Proc. of the 18 International Conference on Distributed Computing Systems, pages 312–321. IEEE Computer Society, 1998.

[24] Marsh, S. Formalising Trust as a Computation Concept. PhD thesis, University of Stirling, 1994.

[25] D. Nauck, F. Klawonn and R. Kruse. Foundations of Neuro-Fuzzy Systems. Wiley, Chichester, 1997 (ISBN: 0-471-97151-0).

[26] Ovchinnikov, S. Fuzzy Sets and Secure Computer Systems. Proceedings of the workshop on New security paradigms, pp. 54-62, 1994.

[27] A. Rahman and S. Hailes. Supporting trust in virtual communities. In Hawaii International Conference on System Sciences 33, pages 1769.1777, 2000.

[28] K. M. Sim and S.Y. Wang. Flexible Negotiation Agent with Relaxed Decision Rules. IEEE Transactions on Systems, Man and Cybernetics, Vol. 34, No. 3., 2004.

[29] X. Wang, X. Shen, N.D. Georganas, "A Fuzzy Logic Based Intelligent Negotiation Agent in E-Commerce", Proc. IEEE Canadian Conference on Electrical and Computer Engineering, 2006.

[30] Wasfy A.M., and Y. Hosni,; Two -Party Negotiation Modeling: An Integrated Fuzzy Logic Approach", Journal of Group Decision and Negotiation, No. 7: 491-518; Kluwer Academic Publishers, Norwell, MA, 1998.

[31] Winsborough, W.H., Seamons, K.E., Jones, V.E.: Automated trust negotiation. In: DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00

[32] Li Xiong and Ling Liu. A reputation-based trust model for peer-to-peer ecommerce communities. In ACM Conference on Electronic Commerce, 2003.

[33] B. Yu and M. P. Singh. An evidential model of distributed reputation management. International Joint Conf. on Autonomous Agents and MultiAgent Systems, 2002.

[34] Wu, Zhengping and Weaver, A. C., Application of Fuzzy Logic in Federated Trust Management for Pervasive Computing, 2006 Workshop on Security, Privacy, and Trust for Pervasive Applications, 2006.

[35] Y. Zhong and B. Bhargava. Authorization based on evidence and trust, Data Warehousing and Knowledge Discovery, volume 2454, Lecture Notes in Computer Science, pages 94.103. Springer, 2002.