An Argumentation-Based Approach to Handling Trust in Distributed Decision Making

Simon Parsons and Elizabeth Sklar
Dept Computer & Information Science,
Brooklyn College,
Brooklyn, NY 11210
parsons@sci.brooklyn.cuny.edu
sklar@sci.brooklyn.cuny.edu

Munindar Singh
Department of Computer Science
North Carolina State University
Raleigh, NC 27695
mpsingh@ncsu.edu

Karl Levitt and Jeff Rowe
Department of Computer Science,
University of California, Davis,
CA 95616.
levitt@cs.ucdavis.edu
rowe@cs.ucdavis.edu

Abstract
Our work aims to support decision making in situations where the source of the information on which decisions are based is of varying trustworthiness. Our approach uses formal argumentation to capture the relationships between such information sources and conclusions drawn from them. This allows the decision maker to explore how information from particular sources impacts the decisions they have to make. We describe the formal system that underlies our work, and a prototype implementation of that system, applied to a problem from military decision making.

Introduction
Trust is a mechanism that mitigates the effects of uncertainty in our knowledge of those around us, providing a way of predicting what those around us will do. Sztompka (1999), for example, defines trust as:

... a bet about the future contingent actions of others.

and similar definitions are suggested by other authors (Gambetta 1990; McKnight & Chervany 1996; Mui, Moteashemi, & Halberstadt 2002).

Our concern with trust is in the context of modern networked warfare. In such a context, there are many issues of trust that bear on decision makers. They must make decisions using information from sources that are of variable trustworthiness, where this trustworthiness can be a matter of relying on sensors that provide data of variable quality and on intelligence reports garnered from more or less reliable informants. The information may also arrive over communication networks of variable trustworthiness, since nodes or links in the networks may be compromised, especially in situations where the security of the communications infrastructure cannot easily be verified. In coalition operations, additional factors relating to the reliability of coalition partners to deliver on promised actions, or the reliability of the equipment they are using, may arise.

One might, as (Sztompka 1999) does implicitly1, decide that trust can be quantified as a probability. Or one might, as Castelfranchi and Falcone (2000) argue, decide that trust is more complex than probability alone, and instead has a rational basis in reasons for beliefs about the future actions of others. In our work we follow (Castelfranchi & Falcone 2000) in adopting a reason-based model of trust (albeit one that can be combined with numerical estimates of trust) since we believe that people are able to take critical decisions under time pressure soundly and with high confidence only if they understand the bases for their decisions. We concentrate, in particular, on the trust that decision-makers have in information sources.

The approach that we take is based on formal argumentation (Rahwan & Simari 2009). Formal argumentation models construct arguments (reasons) for and against adopting beliefs and actions. These arguments explicitly record the agents that need to be trusted in the adoption. As we have suggested before (Parsons, McBurney, & Sklar 2010), the fact that argumentation records the steps used in reaching conclusions makes it appropriate for reasoning where the provenance of information is important, as is widely acknowledged in handling trust (Geerts, Kementsietsidis, & Milano 2006; Golbeck 2006). Linking the provenance information to conclusions means that a decision maker may reason about the sources of evidence, any independence assumptions about the evidence, and about the reasoning process itself—all of which lead to a more nuanced notion of trust than approaches that rely on numeric weights without a clear interpretation of their meanings.

In previous work, we have defined an argumentation system that can reason using the trust relations between a group of individuals, and the information held by those individuals (Tang et al. 2012), and we have developed a prototype implementation of this system (Tang et al. 2011). Here we describe how the formal system and its implementation can be used for the kind of decision making we are interested in.

Related work
As computer systems have become increasingly distributed, and control of those systems has become more decentralized, computational approaches to trust have become steadily more important (Grandison & Sloman 2000). Some of this work has directly been driven by changes in technol-
ogy, for example considering the trustworthiness of nodes in peer-to-peer networks (Abrams, McGrew, & Plotkin 2004; Feldman et al. 2004; Kamvar, Schlosser, & Garcia-Molina 2004), or dealing with wireless networks (Govindan, Mohapatra, & Abdelzaher 2010; Karlof & Wagner 2003; Sun et al. 2005). Other research has been driven by changes in the way that technology is used, especially the use of the Internet. One early challenge is related to the establishment of trust in e-commerce (Mui, Moteassemi, & Halberstadt 2002; Resnick & Zeckhauser 2002; Yu & Singh 2002), and the use of reputation systems to enable this trust (Khopkar, Li, & Resnick 2005; Khosravifar et al. 2010). Another issue is the problem of deciding which of several competing sources of conflicting information one should trust (Adler & de Alfaro 2007; Dong, Berti-Equille, & Srivastava 2009).

Additional issues have arisen with the development of the social web. For example, the question of how social media can be manipulated (Lang, Spear, & Wu 2010; Lerman & Galstyan 2008), and how one should revise one’s notions of trust based on the past actions of individuals (Hang, Wang, & Singh 2008). In this area is some of the work that is most relevant for that we describe here, work that investigates how trust should be propagated through a network of individuals (Guha et al. 2004; Josang, Hayward, & Pope 2006; Katz & Golbeck 2006; Wang & Singh 2006). In this latter work, the input is a network of individuals with links annotated with the degree to which one trusts the other, and the output is the trust inferred between any two nodes in the network. In Figure 1, the input consists of the nodes and the solid edges, and the output consists of the dashed edges.

There is also work that looks at the use of argumentation to handle trust, for example Harwood’s work on networks of trust and distrust (Harwood, Clark, & Jacob 2010), Strangers’ coupling of argumentation with fuzzy trust measures (Stranders, de Weerdt, & Witteveen 2008), Matt’s (Matt, Morge, & Toni 2010), Villata’s use of metalevel argumentation to describe trust (Villata et al. 2011), and Oren’s (Oren, Norman, & Preece 2007) coupling of argumentation and subjective logic (used in (Josang, Hayward, & Pope 2006) to handle trust measures). However, none of this covers the same ground as our work.

**Figure 1: Propagating trust through a social network**

---

**Argumentation**

Our formal argumentation system (Tang et al. 2012) starts with the idea that we deal with a set of individuals $A_g$ where each $A_g$ has access to a knowledge base, $\Delta_i$, containing formulae in some language $L$. The $A_g$ are related by a social network that includes estimates of how much agents trust their acquaintances. These values can be propagated to relate agents that are not directly connected in the social network. An argument is then:

**Definition 1 (Argument)** An argument $A$ from a knowledge base $\Delta_i \subseteq L$ is a pair $(G, p)$ where $p$ is a formula of $L$ and $G \subseteq \Delta_i$ such that:

1. $G$ is consistent;
2. $G \vdash p$; and
3. $G$ is minimal, so there is no proper subset of $G$ that satisfies the previous conditions.

$G$ is called the grounds of $A$, written $G = \text{Grounds}(A)$ and $p$ is the conclusion of $A$, written $p = \text{Conclusion}(A)$. Any $g \in G$ is called a premise of $A$. The key aspect of argumentation is the association of the grounds with the conclusion, in particular the fact that we can trace conclusions to the source of the grounds.

The particular language $L$ that we use is $L^{DHC}$ the language of defeasible Horn clauses, that is a language in which formulae are either atomic propositions $p_i$ or formulae of the form $p_1 \wedge \ldots \wedge p_n \Rightarrow c$, where $\Rightarrow$ is a defeasible rule rather than material implication. Inference in this system is by a defeasible form of generalized modus ponens (DGMP):

$$p_1, \ldots, p_n \quad p_1 \wedge \ldots \wedge p_n \Rightarrow c \quad c$$

and if $p$ follows from a set of formulae $G$ using this inference rule alone, we denote this by $G \vdash^{DHC} p$. In decision making situations, argumentation helps in two ways.

First, it is typical that from the data a given individual $A_g$ has about a situation, we can construct a set of arguments that conflict with each other. We might have an argument $(G, p)$ in favor of some decision option, and another argument $(G’, \neg p)$ against it (in this case we say the arguments rebut each other). We might also have a third argument $(G''', \neg g)$ where $g \in G$ is one of the grounds of the first argument (in this case we say that $(G''', \neg g)$ undermines $(G, p)$). Finally, we might have a fourth argument $(G''''\neg i)$ where $i$ is one of the conclusions to one of the steps in $(G, p)$. (This is another form of rebut, rebuttal of a sub-argument.) Argumentation provides a principled way—or rather a number of alternative ways—for $A_g$ to establish which of a conflicting set of arguments it is most reasonable to accept (Baroni, Caminada, & Giacomin 2011).

Second, the grounds of an argument $G$, can be related back to the sources of that information. If that information comes from some individual $A_g$ that $A_g$ knows, then $A_g$ can weight it according to how much they trust $A_g$, (an extension of Liau’s (2003) principle that you believe information from individuals that you trust), and the same principle can be applied to other sources of information.

---

2Military intelligence traditionally separates information into that which comes from human sources, that which comes from sig-
This weight can be used to resolve conflicts between arguments, and it is possible to provide the decision maker with links between information that feeds into a decision and the source of that information, allowing them to explore the effect of trusting particular sources.

To see more concretely how this can be useful, let’s look at a simple decision-making example.

An example

The example we will use is the following, loosely based on Operation Anaconda (Naylor 2005) and depicted in Figure 2. In this example, a decision is being made about whether to carry out an operation in which a combat team will move into a mountainous region to try to apprehend a high value target (HVT) believed to be in a village in the mountains.

We have the following information. If there are enemy fighters in the area, then an HVT is likely to be in the area. If there is an HVT in the area, and the mission will be safe, then the mission should go ahead. If the number of enemy fighters in the area is too large, the mission will not be safe. UAVs that have flown over the area have provided images that appear to show the presence of a significant number of campfires, indicating the presence of enemy fighters. The quality of the images from the UAV is not very good, so they are not very trusted. A reconnaissance team that infiltrated the area saw a large number of vehicles in the village that the HVT is thought to be inhabiting. Since enemy fighters invariably use vehicles to move around this is evidence for the presence of many enemy fighters. The informants near the combat team base claim that they have been to the area in question and that a large number of fighters are present. In addition we have the default assumption that missions will be safe, because in the absence of information to the contrary we believe that the combat team will be safe.

Thus there is evidence from UAV imaging that sufficient enemy are in the right location to suggest the presence of an HVT. There is also some evidence from informants that there are too many enemy fighters in the area for the mission to be safe. Since informants are paid, their incentive is often to make up what they think will be interesting information and so they are not greatly trusted. However, this conclusion is supported by the findings of the reconnaissance team who are highly trusted.

We might represent this information as follows 3:

\[
\begin{align*}
\text{InArea(campfires)} \\
\text{InArea(vehicles)} \\
\text{Many(enemy)} \\
\text{Safe(mission)} \\
\text{InArea(campfires)} \rightarrow \text{InArea(enemy)} \\
\text{InArea(vehicles)} \rightarrow \text{Many(enemy)} \\
\text{InArea(enemy)} \rightarrow \text{HVT} \\
\text{InArea(enemy)} \\
\land \text{Many(enemy)} \rightarrow \neg\text{Safe(mission)} \\
\text{HVT} \land \text{Safe(mission)} \rightarrow \text{Proceed(mission)}
\end{align*}
\]

From this information we can construct arguments such as:

\[
\left\{ \begin{align*}
\text{InArea(campfires)}, \\
\text{InArea(campfires)} \rightarrow \text{InArea(enemy)}, \\
\text{InArea(enemy)} \land \text{Safe(mission)} \rightarrow \text{HVT}, \\
\text{HVT} \Rightarrow \text{Proceed(mission)}
\end{align*} \right\}
\]

which is an argument for the mission proceeding, based on the fact that there are campfires in the area, these suggest enemy fighters, that enemy fighters suggest the presence of an HVT, and that the presence of an HVT (along with the default assumption that the mission will be safe) suggests that the mission should go ahead.

We can build other arguments from the available information, and, since these will conflict, then compute a subset

\[\text{Proceed(mission)}\]

3While stressing that this is purely illustrative — a real model of this example would be considerably more detailed.
that are acceptable. (Approaches to this computation are discussed in (Baroni, Caminada, & Giacomin 2011).) We can build other arguments from the full information that is available. For example, from the informants’ information we can conclude that there are many enemies in the area and hence the mission will not be safe:

\[
\begin{align*}
\{\text{InArea(vehicles),} \\
\text{InArea(enemy),} \\
\text{InArea(vehicles) } \Rightarrow \text{Many(enemy)} \\
\text{InArea(enemy) } \land \text{Many(enemy) } \Rightarrow \neg \text{Safe(mission)} \\
\neg \text{Safe(mission)}
\end{align*}
\]

This conflicts with the previous argument by undermining the assumption about the mission being safe. Since in our scenario the informants are not highly trusted, the first argument is not defeated and so is then acceptable. The relation between trust in the source of an argument and defeat between arguments is explored in (Parsons et al. 2011). Given all the information from the scenario, we can also construct an argument against the safety of the mission based on information from the recon team. Since the recon team is highly trusted, this argument would defeat the argument for the mission to proceed, rendering it not acceptable.

**ArgTrust**

We have a prototype implementation of the system sketched above which we call ArgTrust. For full details see (Tang et al. 2011). The system currently takes as input an XML file in a format which we sketch here. First, we have a specification of how much sources of information are trusted, for example:

```
<beliefbase>
  <belief>
    <agent> recon </agent>
    <trust>
      <truster> me </truster>
      <trustee> recon </trustee>
      <level> 0.95 </level>
    </trust>
  </belief>
...
</beliefbase>
```

which specifies the individuals involved (including “me”, the decision maker) and the trust relationships between them, including the level of trust (specified as a number between 0 (no trust) and 1 (completely trustworthy)). The current implementation uses these values to compute the trust that one agent places on another using a choice of TidalTrust (Golbeck 2005) or the mechanism described in (Wang & Singh 2006).

The XML file also contains the specification of each individual’s knowledge, for example:

```
<beliefbase>
  <belief>
    <agent> recon </agent>
    <fact> enemy_in_area </fact>
    <level> 0.9 </level>
  </belief>
...
</beliefbase>
```

Here the numbers reflect the belief each individual has in its information about the world.

Given the XML input file, the system can answer queries about whether a given conclusion can be derived by a given agent. The system is invoked from the Unix command line, and generates output in the form of an annotated dot description. This can be converted to any graphical format.

Since displaying all the available information rapidly overwhelms the user, we are working on approaches to providing a zoomable interface. The current prototype version of the software can generate a simple click-and-drill-down HTML interface. The top-level view of the graph for our example is shown in Figure 3(a). Drilling down, as in Figure 3(b), reveals that the one piece of evidence behind the conflict over the safety of the mission is because of a combination of evidence from the UAV and the informants. Further consideration of the situation can then focus on the reliability of these pieces of data, and the data behind the second argument for the mission not to be safe, against the data supporting the argument for the mission to succeed. Were it the case, for example, that the default information about the safety of the mission was considered more reliable than the information from the recon team, then the result of the assessment would be reversed and the conclusion that the mission should proceed would become acceptable. Such “what if” reasoning is supported by our implementation, which has the ability to modify information about a scenario through the command line.

**Future work**

The work we describe here can be extended in four main ways. First, we are continuing to work on the prototype software. This work focuses in particular on the user interface, extending it to allow additional forms of navigation of the argument structure. Second, we are continuing to work on the underlying argumentation model, extending the representation. Here we are working on the use of argument schemes (Parsons et al. 2012) reasoning about why individuals should trust each other, an extension of the current model that just takes the trust in individuals as input. Third, we are evaluating our approach in a series of user studies. Finally, following (Grosse, Chesñevar, & Maguitman 2012), we are examining the feasibility of extracting

---

4http://www.graphviz.org/

5Future releases of the prototype will allow this to be controlled through the GUI.
arguments from natural language. This would make it possible, for example, to extract information from intelligence reports.

**Summary**

This paper has described how argumentation—a form of reasoning that records the reasons behind conclusions, and uses the interactions between the reasons to establish the validity of the conclusions—can be used to support decision-making in situations where information comes from sources of variable trustworthiness. We have demonstrated how a formal system of argumentation can be applied to an example of military decision-making, and have described the current state of a prototype implementation of this formal system.

**Acknowledgement**

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.

The authors are grateful to Timothy Hanratty and Susan Toth, from the Army Research Laboratory, for their comments on drafts of this paper.

**References**


Guha, R.; Kumar, R.; Raghavan, P.; and Tomkins, A. 2004.


