be the ABox assertion added to $\mathcal{A}^*$ while extending $\mathcal{A}$. We know that $b(w_n) \geq b(w)$ and $d(w) \geq d(w_n)$. Belief in $\exists R_n^{\mathcal{I}}(\mathtt{a})$ does not constrain *belief* in $R_n^{\mathcal{I}}(\mathtt{a}, \mathtt{b})$. However, in a consistent knowledge base, disbelief in $R_n^{\mathcal{I}}(\mathtt{a}, \mathtt{b})$ cannot be lower than disbelief in $\exists R_n^{\mathcal{I}}(\mathtt{a})$ or $\exists R_n^{-\mathcal{I}}(\mathtt{b})$. Hence, disbelief in $\exists R_n^{\mathcal{I}}(\mathtt{a})$ is constrained by the following TBox axioms in $\mathcal{T}^*$:

- $B_1 \sqsubseteq \neg\exists R_n, B_2 \sqsubseteq \neg\exists R_n, \ldots, B_{a-1} \sqsubseteq \neg\exists R_n, B_a \sqsubseteq \neg\exists R_n$
- $\exists R_n \sqsubseteq B_{a+1}, \exists R_n \sqsubseteq B_{a+2}, \ldots, \exists R_n \sqsubseteq B_{b-1}, \exists R_n \sqsubseteq B_b$
- $\exists R_n \sqsubseteq \neg B_{b+1}, \exists R_n \sqsubseteq \neg B_{b+2}, \ldots, \exists R_n \sqsubseteq \neg B_{c-1}, \exists R_n \sqsubseteq \neg B_c$

Disbelief in $\exists R_n^{-\mathcal{I}}(\mathtt{b})$ is constrained by the axioms in $\mathcal{T}^*$:

- $B_{c+1} \sqsubseteq \neg\exists R_n^-, B_{c+2} \sqsubseteq \neg\exists R_n^-, \ldots, B_{d-1} \sqsubseteq \neg\exists R_n^-, B_d \sqsubseteq \neg\exists R_n^-$
- $\exists R_n^- \sqsubseteq B_{d+1}, \exists R_n^- \sqsubseteq B_{d+2}, \ldots, \exists R_n^- \sqsubseteq B_{e-1}, \exists R_n^- \sqsubseteq B_e$
- $\exists R_n^- \sqsubseteq \neg B_{e+1}, \exists R_n^- \sqsubseteq \neg B_{e+2}, \ldots, \exists R_n^- \sqsubseteq \neg B_{f-1}, \exists R_n^- \sqsubseteq \neg B_f$

Let w_i refer to the opinion related to the ABox assertion for $B_i(\mathtt{a})$ if $1 \leq i \leq c$ and for $B_i(\mathtt{b})$ if $c+1 \leq i \leq f$. The most general opinion w'' that satisfies the constraints for $R_n^{\mathcal{I}}(\mathtt{a}, \mathtt{b})$ is computed as follows:

$$\begin{aligned} w'' = (b(w), \max(S_d)) \quad \texttt{where} \\ S_d = \{d(w), b(w_1), \ldots, b(w_a), b(w_{c+1}), \ldots, b(w_d), \\ d(w_{a+1}), d(w_{a+2}), \ldots, d(w_b), \\ d(w_{d+1}), d(w_{d+2}), \ldots, d(w_e), \\ b(w_{b+1}), b(w_{b+2}), \ldots, b(w_c), \\ b(w_{e+1}), b(w_{e+2}), \ldots, b(w_f)\} \end{aligned} \tag{2}$$

If $b(w'') + d(w'') > 1$, there is no opinion satisfying the constraints defined by the semantics for $R_n(\mathtt{a}, \mathtt{b})$; otherwise, we take w'' as interpretation of $R_n(\mathtt{a}, \mathtt{b})$.

Let us now compute the interpretation of *roadBombedBy*(R, G*1*) in our example scenario. It is constrained by the interpretations of two other assertions ∃*roadBombedBy*(R) and ∃*roadBombedBy*⁻(G*1*). The disbelief in the interpretation of ∃*roadBombedBy*(R) is constrained by TBox axioms t_6, t_7, t_8, t_9, and t_{12} of Table 4. We have opinion assertions only for the assertions *Blocked*(R), *Safe*(R), and *BombedRoad*(R) in the extended ABox derived from Table 2. Therefore, based on Equation 2, we compute the interpretation as $(0, \max(\{0, 0.63, 0.3, 0.09\})) = (0, 0.63)$. Table 4 shows the interpretations computed for the scenario using Equations 1 and 2.

The computational complexity of these calculations is $O(n)$ in the size of $\mathcal{T}^*$ and $\mathcal{A}^*$. Now, we introduce Theorem 1 , which defines the conditions necessary and sufficient for inconsistency.

THEOREM 1. *An **extended** $\mathcal{S}$DL-Lite KB $\mathcal{K}^* = (\mathcal{T}^*, \mathcal{A}^*)$ with a coherent $\mathcal{T}^*$ is inconsistent with respect to the semantics in Table 3 if and only if one of the following conditions hold:*

1. *Given $B_m(\mathtt{a})$:w_m, $B_n(\mathtt{a})$:$w_n \in \mathcal{A}^*$, and $B_m \sqsubseteq B_n \in \mathcal{T}^*$, we have $b(w_m) + d(w_n) > 1$*
2. *Given $B_m(\mathtt{a})$:w_m, $B_n(\mathtt{a})$:$w_n \in \mathcal{A}^*$, and $B_m \sqsubseteq \neg B_n \in \mathcal{T}^*$, we have $b(w_m) + b(w_n) > 1$*

Proof: *The inconsistency arises if and only if at least one class or role does not have a valid interpretation satisfying the semantics in Table 3. Let us first analyse the inconsistencies due to the interpretations of classes. The most general interpretation for $B_n(\mathtt{a})$ is computed as in Equation 1 and referred to as is w_n''. Let w_n be the opinion for $B_n(\mathtt{a})$ in $\mathcal{A}^*$. If $b(w_n'') + d(w_n'') > 1$, there is no opinion satisfying the constraints defined by the semantics for $B_n(\mathtt{a})$ and $\mathcal{K}^*$ is inconsistent. To have $b(w_n'') + d(w_n'') > 1$, one of the following conditions must hold based on Equation 1:*

- $b(w_n'') = b(w_n)$
  - *$d(w_n'') \in \{b(w_{l+1}), b(w_{l+2}), \ldots, b(w_i)\}$: This implies that there exists a TBox axiom $B_m \sqsubseteq \neg B_n$ with ABox assertions $B_m(\mathtt{a})$:w_m and $B_n(\mathtt{a})$:w_n such that $b(w_n) + b(w_m) > 1$.*
  - *$d(w_n'') \in \{d(w_{i+1}), d(w_{i+2}), \ldots, d(w_j)\}$: This implies that there exists a TBox axiom $B_n \sqsubseteq B_m$ with ABox assertions $B_m(\mathtt{a})$:w_m and $B_n(\mathtt{a})$:w_n such that $b(w_n) + d(w_m) > 1$.*
  - *$d(w_n'') \in \{b(w_{j+1}), b(w_{j+2}), \ldots, b(w_k\}$: This implies that there exists a TBox axiom $B_n \sqsubseteq \neg B_m$ with ABox assertions $B_m(\mathtt{a})$:w_m and $B_n(\mathtt{a})$:w_n such that $b(w_n) + b(w_m) > 1$.*
- $b(w_n'') \in \{b(w_0), b(w_1), \ldots, b(w_l)\}$
  - *$d(w_n'') = d(w_n)$: This implies that there exists a TBox axiom $B_m \sqsubseteq B_n$ with ABox assertions $B_m(\mathtt{a})$:w_m and $B_n(\mathtt{a})$:w_n such that $b(w_m) + d(w_n) > 1$.*
  - *$d(w_n'') \in \{b(w_{l+1}), b(w_{l+2}), \ldots, b(w_i)\}$: This implies TBox axioms $B_x \sqsubseteq B_n$ and $B_y \sqsubseteq \neg B_n$ with ABox assertions $B_x(\mathtt{a})$:w_x and $B_y(\mathtt{a})$:w_y. These TBox axioms imply that the extended $\mathcal{T}^*$ contains $B_x \sqsubseteq \neg B_y$ and $b(w_x) + b(w_y) > 1$.*
  - *$d(w_n'') \in \{d(w_{i+1}), d(w_{i+2}), \ldots, d(w_j)\}$: This implies TBox axioms $B_x \sqsubseteq B_n$ and $B_n \sqsubseteq B_y$ with ABox assertions $B_x(\mathtt{a})$:w_x and $B_y(\mathtt{a})$:w_y. These TBox axioms imply that $\mathcal{T}^*$ contains $B_x \sqsubseteq B_y$ and $b(w_x) + d(w_y) > 1$.*
  - *$d(w_n'') \in \{b(w_{j+1}), b(w_{j+2}), \ldots, b(w_k\}$: This implies TBox axioms $B_x \sqsubseteq B_n$ and $B_n \sqsubseteq \neg B_y$ with ABox assertions $B_x(\mathtt{a})$:w_x and $B_y(\mathtt{a})$:w_y. These TBox axioms imply that $\mathcal{T}^*$ contains $B_x \sqsubseteq \neg B_y$ and $b(w_x) + b(w_y) > 1$.*

Now, we look into the interpretations for role axioms to test consistency. The most general interpretation for $R_n(\mathtt{a}, \mathtt{b})$ is computed as in Equation 2 and referred to as w''. Let w be the opinion for $R_n(\mathtt{a}, \mathtt{b})$ in $\mathcal{A}^*$. If $b(w'') + d(w'') > 1$, there is no opinion satisfying the constraints defined by the semantics for $R_n(\mathtt{a}, \mathtt{b})$ and $\mathcal{K}^*$ is inconsistent. To have $b(w'') + d(w'') > 1$, one of the following conditions must hold based on Equation 2:

- *$B_m \sqsubseteq \neg\exists R_n$ or $B_m \sqsubseteq \neg\exists R_n^-$ and $b(w_n) + b(w_m) > 1$ since $b(w) + b(w_m) > 1$*
- *$\exists R_n \sqsubseteq B_m$ or $\exists R_n^- \sqsubseteq B_m$ and $b(w_n) + d(w_m) > 1$ since $b(w) + d(w_m) > 1$*
- *$\exists R_n \sqsubseteq \neg B_m$ or $\exists R_n^- \sqsubseteq \neg B_m$ and $b(w_n) + b(w_m) > 1$ since $b(w) + b(w_m) > 1$*

As shown, in the case of inconsistency, one of the conditions defined in Theorem 1 must hold. Furthermore, if none of these conditions holds in $\mathcal{K}^*$, we guarantee that $\mathcal{K}^*$ is consistent. ■

In an inconsistent extended $\mathcal{S}$DL-Lite knowledge base $\mathcal{K}^* = (\mathcal{T}^*, \mathcal{A}^*)$, the inconsistencies exist only because of *conflicting opinions*. Two opinions w_m and w_n, which are about $B_m(\mathtt{a})$ and $B_n(\mathtt{a})$ respectively, are in conflict if they satisfy one of the conditions in Theorem 1. We label the portion of w_m which conflicts with w_n as c_{mn}, and refer to it as the *conflicting portion*. If the conflict is due to the axiom $B_m \sqsubseteq B_n \in \mathcal{T}^*$, then the conflict arises because $b(w_m) + d(w_n) > 1$; hence $c_{mn} = b(w_m)$ and $c_{nm} = d(w_n)$. On the other hand, if the conflict is due to the axiom $B_m \sqsubseteq \neg B_n \in \mathcal{T}^*$, we have conflict because $b(w_m) + b(w_n) > 1$; hence $c_{mn} = b(w_m)$ and $c_{nm} = b(w_n)$. If all conflicts in $\mathcal{K}^*$ are resolved, then the knowledge base becomes consistent.

In the rest of the paper, we assume that the opinion about a specific ABox assertions is provided by a single source. When there is more than one source for an assertion, only one of them is chosen (e.g. based on their trustworthiness). This will be relaxed in future.

Table 5: Extended ABox for case II

a_1 : $\mathit{roadBombedBy}(\mathtt{R}, \mathtt{G}_1)$:$(0.67, 0.083, 0.247)$
a_2 : $\mathit{roadBombedBy}^{-}(\mathtt{G}_1, \mathtt{R})$:$(0.67, 0.083, 0.247)$
a_3 : $\exists \mathit{roadBombedBy}(\mathtt{R})$:$(0.67, 0.0, 0.33)$
a_4 : $\exists \mathit{roadBombedBy}^{-}(\mathtt{G}_1)$:$(0.67, 0.0, 0.33)$
a_5 : $\mathit{Blocked}(\mathtt{R})$:$(0.71, 0.09, 0.2)$
a_6 : $\mathit{Safe}(\mathtt{R})$:$(0.63, 0.066, 0.304)$
a_7 : $\mathit{BombedRoad}(\mathtt{R})$:$(0.2, 0.3, 0.5)$

Having described $\mathcal{S}$DL-Lite we now examine a novel application of the system, describing how evidence from multiple sources can be reasoned about based on the trust placed in these sources.

4. TRUST-BASED EVIDENCE ANALYSIS

Here we get to the crux of the problem being addressed in this paper: how can we draw reliable conclusions regarding the state of the world, given evidence acquired from disparate sources (agents), about whom we have variable trust? We refer to this process as trust-based evidence analysis. Our aim is not to offer a new mechanism for assessing the trustworthiness of information sources; in fact, we exploit a widely-studied model [10] for this purpose based on Beta distributions as described in Section 2.2. The novelty of this work lies in the use of such models to guide evidence analysis.

4.1 Handling Inconsistencies

$\mathcal{S}$DL-Lite presented in the previous section provides a tractable means to capture and interpret evidence acquired from other agents. The fact that we have evidence from multiple agents, however, means that there are likely to be inconsistencies in the evidence received. Thus, given evidence (i.e., opinions) from various sources, our knowledge-base may not be consistent. This is despite the use of *discounting* through DST. Discounting provides us with a "best-guess" of the reliability of agents based on an aggregation of our prior experiences with, and other knowledge of them as evidence sources. As with any computational model of trust, the trust assessments that drive discounting are vulnerable to: lack of evidence about other agents and the effects of whitewashing [2]; a conflation of the probability of malicious behaviour and lack competence/expertise in the evidence-provider; strategic liars; and collusion among evidence-providers. In our running example, for instance, local police and civilian sources have relatively low trustworthiness, not because of any perceived malicious intent but due to a belief that they lack experience in providing precise information. With more evidence, trustworthiness of information sources may be modelled more accurately, but our challenge is to support the analysis of evidence given the status quo.

To illustrate this challenge, consider an adaptation of our example (case II) in which additional evidence is received from a third source, agent C, about R: C reports that R was bombed by $\mathtt{G}_1$ with opinion $(0.8, 0.1, 0.1)$. With this additional report, our ABox contains $\mathit{roadBombedBy}(\mathtt{R}, \mathtt{G}_1)$:$(0.67, 0.083, 0.247)$ after discounting the opinion with C's trustworthiness 0.83 listed in Table 1 (a_1 in Table 5 where the resulting extended ABox is presented). The extended ABox will now has a conflict between a_1 and a_6, because $0.67 + 0.63 > 1$ and the extended TBox contains $\exists \mathit{roadBombedBy} \sqsubseteq \neg \mathit{Safe}$. Let $w_1 = (0.63, 0.066, 0.304)$ and $w_2 = (0.67, 0.0, 0.33)$. The *conflicting portions* of w_1 and w_2 are $c_{12} = 0.63$ and $c_{21} = 0.67$. Let us refer to the trustworthiness of the sources of w_1 and w_2 as t_1 and t_2 respectively. In our example, from Table 1, $t_1 = 0.83$ and $t_2 = 0.786$. In order for us to transform our inconsistent knowledge-base into a *consistent* knowledge-base, from which we can draw valid conclusions given our semantics, we need to determine additional discounting factors x_1 and x_2 for *opinions* w_1 and w_2 such that $0 \leq c_{12}.x_1 + c_{21}.x_2 \leq 1$.

In this paper, we specify this problem as that of finding *additional* discounting factors for the belief-mass distributions of pieces of evidence to make our knowledge-base consistent. In general, our conflict resolution problem is a tuple $\langle \mathcal{C}, \mathcal{X} \rangle$ where $\mathcal{C}$ is the set of conflicting portions that appear in the extended knowledge base $\mathcal{K}^*$, and $\mathcal{X}$ is a set of additional discounting factors corresponding to $\mathcal{C}$. We require that, in $\langle \mathcal{C}, \mathcal{X} \rangle$, $\forall c_{ij} \in \mathcal{C}$, $\exists c_{ji} \in \mathcal{C}$ and $\exists x_i, x_j \in \mathcal{X}$. Then, a solution to this problem is an assignment of values to each $x_i \in \mathcal{X}$ such that

$$\forall c_{ij}, c_{ji} \in \mathcal{C}, \forall x_i, x_j \in \mathcal{X} \quad 0 \leq c_{ij}.x_i + c_{ji}.x_j \leq 1$$

There are many heuristic approaches to solving this problem, among them being to consider only consistent knowledge to draw conclusions from the evidence received; i.e. $\forall x_i \in \mathcal{X}, x_i = 0$. This, however, could lead to a significant loss of evidence. Here, we explore a nuber of increasingly refined approaches that guarantee the generation of a consisitent knowledge-base: *trust-based deleting*, *trust-based discounting* and *evidence-based discounting*.

4.2 Trust-based deleting

If two opinions w_1 and w_2 are in conflict, the opinion from the less trustworthy source is deleted, and if both sources are equally trustworthy both opinions are deleted. Thus, if the trust we have in the source of opinion w_1 is greater than that of the source of w_2 ($t_1 > t_2$) then $x_2 = 0$ and $x_1 = 1$, and in the event that $t_1 = t_2$ we assign $x_1 = x_2 = 0$. In our example, the local police sources P are slightly less trustworthy than the local civilian sources C. Hence, the opinion about $\mathit{Safe}(\mathtt{R})$ is changed to $(0, 0, 1)$ and the conflict is resolved. This approach, however, neglects the amount of evidence used to calculate trust, and it does not consider the difference between trust values ($t_{\mathtt{C}} = 0.83 \approx t_{\mathtt{P}} = 0.786$).

4.3 Trust-based discounting

If two opinions w_1 and w_2 are in conflict, they are discounted in proportion to the trustworthiness of their sources. That is, the additional discounting factor for w_1 and w_2 is computed using $t_1/(c_{12}t_1 + c_{21}t_2)$ and $t_2/(c_{12}t_1 + c_{21}t_2)$, respectively, where t_1 and t_2 are the trustworthiness of the sources of the opinions. In our example, an additional discount factor of $\mathit{roadBombedBy}(\mathtt{R}, \mathtt{G}_1)$ is 0.79 and that of $\mathit{Safe}(\mathtt{R})$ is 0.75, since the trustworthiness of C and P are 0.83 and 0.786, respectively. Therefore, to resolve the conflict, the original opinion of C about $\mathit{roadBombedBy}(\mathtt{R}, \mathtt{G}_1)$ is discounted by $0.83 \times 0.79 = 0.65$ and that of P about $\mathit{Safe}(\mathtt{R})$ is discounted by $0.786 \times 0.75 = 0.59$. However, this approach neglects the amount of evidence used to calculate trust in sources.

4.4 Evidence-based discounting

Within the evidence analysis domain, the information that we have to work with relates to past experiences with a specific agent (i.e., information source) ϱ_k where information received has proven reliable or unreliable according to some criteria (as would be captured in any trust assessment model). In other words, the amount of positive evidence we have for agent ϱ_k, namely r_k, and the amount of negative evidence for that agent, namely s_k. From this evidence, we calculate trustworthiness of ϱ_k, denoted as t_k described in Section 2.2. When we receive opinion w_i^k from ϱ_k, we discount it by t_k and add the resulting opinion w_i to our knowledge base. However, as explained before, additional discounting by factor x_i is required when w_i is in conflict with another opinion in the knowledge base. Discounting w_i by x_i implies discounting the original opinion w_i^k by $t_k.x_i$. This corresponds to revising the trustworthiness of w_i^k as $t_k.x_i$ by speculating about the trustworthiness of ϱ_k regarding this single opinion. That is, even though the trustworthiness of ϱ_k is t_k based on the existing evidence (r_k, s_k), it becomes $t_k.x_i$ for this specific opinion w_i^k; so, $t_k x_i$ effectively becomes the trust in w_i^k.

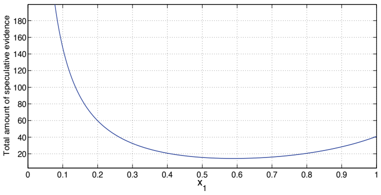

Figure 1: Speculative evidence required for case II ($\kappa = 1$).

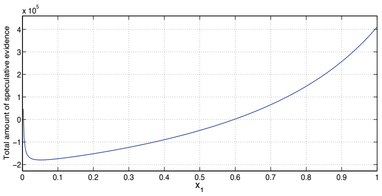

Figure 2: Speculative evidence required for case III ($\kappa = 1$).

Here, we create a metric to measure how much we speculate about the trustworthiness of ϱ_k regarding w_i^k.

First, to decrease trust from t_k to $t_k.x_i$, we need additional negative evidence, which is called *speculative evidence* and denoted as ρ_i. Our intuition is that it is less likely for a trustworthy agent to present additional negative *speculative evidence* than it is for an untrustworthy agent, and thus the receipt of such evidence should be tempered by $(\bar{t}_k)^\kappa$. Here, $\bar{t}_k$ represents the *distrust* we have in agent ϱ_k; i.e. the likelihood that we will receive additional negative evidence given our experiences with the source. The calibration constant $\kappa \geq 0$ enables us to vary the influence that prior experience has on our prediction that an individual will present negative evidence in the future. If $\kappa = 0$, for example, we assume that all sources are equally likely to provide negative evidence. Now, using the Beta distribution formula for trust, we obtain:

$$\begin{aligned} t_k.x_i = \frac{r_k+1}{r_k+s_k+2} \cdot x_i &= \frac{r_k+1}{s_k+r_k+2+\rho_i.(\bar{t}_k)^\kappa} \\ &= \frac{r_k+1}{s_k+r_k+2+\rho_i.(\frac{s_k+1}{r_k+s_k+2})^\kappa} \end{aligned}$$

Rearranging this for ρ_i yields:

$$\rho_i = \frac{\nu_i(1-x_i)}{x_i} \quad \text{where} \quad \nu_i = \frac{(r_k+s_k+2)^{\kappa+1}}{(s_k+1)^\kappa} \tag{3}$$

To illustrate this, let us return to case II above in which agent C reports that road R is bombed by G_1. Using Equation 3, we can compute the total amount of *speculative evidence* necessary to discount w_1 and w_2 by x_1 and x_2, respectively. If we assume that $c_{12}.x_1 + c_{21}.x_2 = 1$, we have $x_2 = (1 - c_{12}.x_1)/c_{21}$. Then, the total amount of *speculative evidence* (i.e. $\rho_1 + \rho_2$) can be formulated as a function of single variable x_1 by Equation 4, which is plotted in Figure 1. This function has a minimum at $x_1 = 0.5892$ in the interval $[0, 1]$ and the corresponding x_2 is 0.9607. That is, for a consistent knowledge base, trust in C's opinion about $roadBombedBy(\mathtt{R}, \mathtt{G}_1)$ should be reduced to 0.489 from 0.83, but the trust in the opinion of P about $Safe(\mathtt{R})$ is reduced only slightly to 0.755 from 0.786. This reflects the relative level of positive and negative evidence we have from prior experience from both parties, and results in a consistent knowledge-based from which we can draw conclusions.

$$f(x_1) = \frac{\nu_1(1-x_1)}{x_1} + \frac{\nu_2(1-\frac{(1-c_{12}.x_1)}{c_{21}})}{\frac{(1-c_{12}.x_1)}{c_{21}}} \tag{4}$$

Table 6: Extended ABox for case III

$roadBombedBy(\mathtt{R}, \mathtt{G}_1)$:$(0.67, 0.083, 0.247)$
$roadBombedBy^-(\mathtt{G}_1, \mathtt{R})$:$(0.67, 0.083, 0.247)$
$\exists roadBombedBy(\mathtt{R})$:$(0.67, 0.0, 0.33)$
$\exists roadBombedBy^-(\mathtt{G}_1)$:$(0.67, 0.0, 0.33)$
$Blocked(\mathtt{R})$:$(0.71, 0.09, 0.2)$
$Safe(\mathtt{R})$:$(0.63, 0.066, 0.304)$
$BombedRoad(\mathtt{R})$:$(0.2, 0.6, 0.2)$
$SabotagedRoad(\mathtt{R})$:$(0.801, 0, 0.199)$

Table 7: After extra discounting for case III ($\kappa = 1$)

Extended ABox
$roadBombedBy(\mathtt{R}, \mathtt{G}_1)$:$(0.0342, 0.0043)$
$roadBombedBy^-(\mathtt{G}_1, \mathtt{R})$:$(0.0342, 0.0043)$
$\exists roadBombedBy(\mathtt{R})$:$(0.0342, 0)$
$\exists roadBombedBy^-(\mathtt{G}_1)$:$(0.0342, 0)$
$Blocked(\mathtt{R})$:$(0.71, 0.09)$
$Safe(\mathtt{R})$:$(0.63, 0.066)$
$BombedRoad(\mathtt{R})$:$(0.2, 0.6)$
$SabotagedRoad(\mathtt{R})$:$(0.0304, 0)$

Computed Interpretations
$Blocked(\mathtt{R})$:$(0.71, 0.09)$
$Safe(\mathtt{R})$:$(0.63, 0.066)$
$BombedRoad(\mathtt{R})$:$(0.2, 0.63)$
$SabotagedRoad(\mathtt{R})$:$(0.0342, 0.63)$
$\exists roadBombedBy(\mathtt{R})$:$(0.0342, 0.63)$
$\exists roadBombedBy^-(\mathtt{G}_1)$:$(0.0342, 0)$
$\exists roadBombedBy^-(\mathtt{G}_2)$:$(0, 0, 1)$
$roadBombedBy(\mathtt{R}, \mathtt{G}_1)$:$(0.0342, 0.63)$
$roadBombedBy(\mathtt{R}, \mathtt{G}_2)$:$(0, 0.63)$
$roadBombedBy^-(\mathtt{G}_1, \mathtt{R})$:$(0.0342, 0.63)$
$roadBombedBy^-(\mathtt{G}_2, \mathtt{R})$:$(0, 0.63)$

Until now, we considered only one conflict between two opinions. When we have multiple conflicts, they may interact in such a way that resolving one may also affect the resolution of another. To illustrate this, consider two new intelligence reports (case III):

- A reports a bomb explosion on R with opinion $(0.2, 0.6, 0.2)$.
- M_2 informs M_1 that R is sabotaged with opinion $(0.9, 0, 0.1)$.

The resulting ABox is shown in Table 6 and implies three relevant conflicts: $0.67 + 0.63 > 1$, $0.67 + 0.6 > 1$, and $0.63 + 0.801 > 1$. Let us refer to $(0.2, 0.6, 0.2)$ and $(0.801, 0, 0.199)$ as w_3 and w_4, respectively. We refer to the conflicting portions as $c_{31} = 0.6$ and $c_{42} = 0.801$. We also use x_3 and x_4 to refer to the additional discounting necessary for w_3 and w_4, respectively, to resolve the conflicts. The overall amount of *speculative evidence* necessary to resolve all of these relevant conflicts is computed as in Equation 5.

$$\begin{aligned} f(x_1) &= \frac{\nu_1(1-x_1)}{x_1} + \frac{\nu_2(1-x_2)}{x_2} + \frac{\nu_3(1-x_3)}{x_3} + \frac{\nu_4(1-x_4)}{x_4} \\ \text{such that} \quad & 0 \leq c_{12}.x_1 + c_{21}.x_2 \leq 1 \text{ and} \\ & 0 \leq c_{13}.x_1 + c_{31}.x_3 \leq 1 \text{ and} \\ & 0 \leq c_{24}.x_2 + c_{42}.x_4 \leq 1 \end{aligned} \tag{5}$$

Since these conflicts are relevant, we can write x_2, x_3 and x_4 in terms of x_1 if we set $c_{12}x_1 + c_{21}x_2 = 1$, $c_{13}x_1 + c_{31}x_3 = 1$, and $c_{24}x_2 + c_{42}x_4 = 1$. The resulting function is shown in Figure 2 and has a minimum at $x_1 = 0.0514$ in the interval $[0, 1]$. The other discounting factors are computed as $x_2 = 1$, $x_3 = 1$, and $x_4 = 0.043$ in the same interval. That is, trust in the opinion of C about $roadBombedBy(\mathtt{R}, \mathtt{G}_1)$ is reduced to 0.0427 and trust in the opinion of M about $SabotagedRoad(\mathtt{R})$ is reduced to 0.0338. The ABox and the computed interpretations after extra discounting is shown in Table 7.

We generalise this approach for any number of conflicts with arbitrary relations. Assume we have a set of conflicting opinions $\{\langle w_i, w_j\rangle, \ldots, \langle w_m, w_n\rangle\}$ and, derived from trust evidence about agents, coefficients $\{\nu_i, \nu_j, \ldots, \nu_m, \nu_n\}$. To determine the optimum discounting factors $\{x_i, x_j, \ldots, x_m, x_n\}$ for these opinions, we construct the following optimisation problem with a multivariate non-linear objective function and linear constraints.

$$\begin{aligned} & \arg\min_{\overrightarrow{x}} f(\overrightarrow{x}) \quad \text{where} \\ f(\langle x_i, x_j, \ldots, x_m, x_n\rangle) &= \frac{\nu_i(1-x_i)}{x_i} + \frac{\nu_j(1-x_j)}{x_j} + \ldots \\ & \quad \frac{\nu_m(1-x_m)}{x_m} + \frac{\nu_n(1-x_n)}{x_n} \\ \text{such that} \quad & 0 \leq x_i \leq 1, 0 \leq x_j \leq 1, \ldots \\ \text{and} \quad & 0 \leq c_{ij}x_i + c_{ji}x_j \leq 1, \ldots \end{aligned} \tag{6}$$

Existing constrained non-linear programming methods can be used to solve this problem in order to estimate the best discounting factors. There are various techniques that may be used including *Interior-Point* and *Active-Set* algorithms. In this work, we use *Interior-Point* approximation. Details of these methods are out of the scope of this paper and can be found elsewhere [14].

In this section we have formalised the problem of computing additional discounting factors for *opinions* received about the world from other agent so that we may formulate a consistent $\mathcal{S}$DL-Lite knowledge-base from which we can draw reliable conclusions. We have presented a number of approaches to the resolutions of inconsistencies between opinions including an optimisation-based approach, evidence-based discounting. Next, we evaluate these approaches with respect to their robustness in the face of liars.

5. EVALUATION

We have evaluated our approach through a set of simulations. In each simulation, we define the domain by randomly generating an $\mathcal{S}$DL-Lite TBox that contains 100 concepts and roles, as well as axioms over those, e.g., $B_1 \sqsubseteq B_2$ and $B_2 \sqsubseteq \neg\exists R_3$. For each role or concept, there is one information source that provides opinions about its instances, e.g., $B_1(\mathsf{a}){:}(0.8, 0, 0.2)$ and $R_3(\mathsf{a}, \mathsf{b}){:}(0.5, 0.1, 0.4)$. There are 10 information sources in total, each is an expert on 10 concepts and roles, and provides its opinions about those.

In our simulations, we assume there is one information consumer that uses the information from sources to make decisions. Each simulation is composed of 10 iterations. At each iteration t, the consumer needs to gather information about an individual a. We generate ground truth about a, which is composed of one assertion about a for each concept and role with an associated opinion. Each information source knows the ground truth only about the concepts and roles of their expertise. However, they may not provide the ground truth to the consumer when it is requested. Behaviours of the information sources are determined by their behavioural type, which are summarised as follows.

- **Honest:** Most of the time, this type of sources provide the ground truth about the assertion of their expertise with small Gaussian noise $N(0, 0.01)$. With probability P_b, honest sources behave like malicious ones and provide bogus information.
- **Malicious:** This type of sources aim at misleading the information consumer by providing bogus opinions. More specifically, given $(b, d, \_)$ is the ground truth about an assertion, a malicious source provides the opinion $(abs(\epsilon_1), 0.9 + \epsilon_2, \_)$ if $b \approx d$; otherwise it provides the opinion $(d + \epsilon_1, b + \epsilon_2, \_)$, where $\epsilon_1, \epsilon_2 \in [-0.05, 0.05]$. There are two types of malicious sources, which are defined as follows:
    i. **Simple liars:** they always provide bogus opinions.
    ii. **Strategic liars:** they behave like honest sources to build trust and then provides bogus information exploiting the built trust. After providing misleading information to the consumer, they change their identity to avoid negative evidence against them.

After collecting opinions about different assertions from information sources, the information consumer uses its trust in these sources to discount these opinions and uses the proposed reasoning mechanisms for $\mathcal{S}$DL-Lite to compute interpretations. Ideally, these interpretations should be close to the ground truth if all sources are accurate and their trustworthiness is modelled correctly. If there are some malicious sources, there may be conflicts in the collected information. In the case of conflicts, the consumer resolve the conflicts using *Naive Deleting* (NDL), *Trust-based Deleting* (TDL), *Trust-based Discounting* (TDC), or *Evidence-based Discounting* (EDC) with $\kappa = 1$. In NDL, all conflicting opinions are deleted from the knowledge base to resolve the conflicts. The consumer computes the interpretations for concept and role assertions related to a, after resolving the conflicts if any. Then, we measure the performance as the *mean absolute error* in the computed interpretations. Let (b, d, u) be the ground truth and (b', d', u') be the computed interpretation for assertion $B(\mathsf{a})$, then the *absolute error* in the interpretation is computed as $err_{B(a)} = abs(\delta_b) + abs(\delta_d)$, where $\delta_b = b' - b$ and $\delta_d = d' - d$. For instance, if the ground truth about $B(\mathsf{a})$ is $(0.9, 0.05, 0.05)$, but the computed interpretation is $(0.05, 0.9, 0.05)$, then the error would be 1.7.

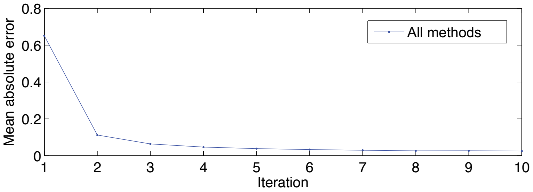

Figure 3: Simple liars ($R_{liar} = 0.5$ and $P_b = 0$)

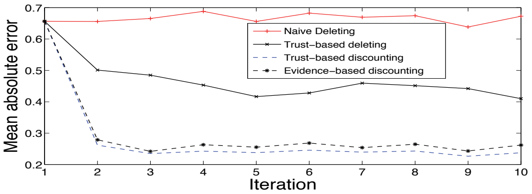

Figure 4: Simple liars ($R_{liar} = 0.5$ and $P_b = 0.1$)

At the end of each iteration, the consumer learns the ground truth and updates the trustworthiness of the information sources with new evidence (r^t, s^t) computed as in Equation 7, which is based on the intuition that the information is still useful if it has a small amount of noise or is slightly discounted.

$$(r^t, s^t) = \begin{cases} (0, 1), & \text{if } \delta_b > 0.1 \text{ or } \delta_d > 0.1 \\ (1, 0), & \text{if } -0.1 \le \delta_b \le 0.01 \text{ and } -0.1 \le \delta_d \le 0.01 \\ (0, 0), & \text{otherwise.} \end{cases} \qquad (7)$$

Each of our simulations are repeated 10 times and our results are significant based on *t-test* with a confidence interval of 0.95.

Without any evidence, the trustworthiness of sources is computed as 0.5. Thus, there are is no conflict in the beginning of our simulations. If all sources have deterministic behaviours, i.e., malicious sources are simple liars and $P_b = 0$, then trustworthiness of sources are easily modelled over time and the opinions from liars are significantly discounted. In such settings, conflicts are totally avoided and information consumers using either of the four proposed methods have the same level of success. Figure 3 shows an example of this setting where honest sources always provides the truth ($P_b = 0$) and malicious sources are simple liars. Here, the *ratio of liars* (R_{liar}) is 0.5, i.e., half of the sources are malicious.

When honest sources provide bogus information occasionally, the conflicts may arise in the knowledge base of the consumer, because the information from these sources are not significantly discounted. Figure 4 shows our results for $R_{liar} = 0.5$ and $P_b = 0.1$, where all malicious sources are simple liars. In this setting, NDL leads to significant errors in the computed interpretations. While TDL does much better than NDL, it is outperformed by discounting based approaches TDC and EDC. Both of these approaches have similarly good performance though TDC does slightly better.

Simple liars may not be enough to model malicious sources in real life. That is why we change the type of malicious sources

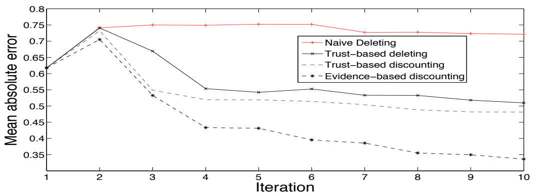

Figure 5: Strategic liars ($R_{liar} = 0.5$ and $P_b = 0.1$)

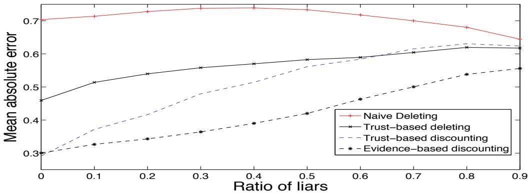

Figure 6: Strategic liars with varying R_{liar} ($P_b = 0.1$)

to strategic liars and repeat our simulations. Figure 5 shows our results for $R_{liar} = 0.5$ and $P_b = 0.1$. In this settings, trust evaluations become misleading, since strategic liars build trust, make their impact and then change their identity to avoid any negative evidence. As a result, as shown in the figure, TDC fails significantly more than EDC after a few iterations. We repeat the simulations with strategic liars for different R_{liar} values; our results are shown in Figure 6. Our results indicate that evidence-based discounting is much more robust in the presence of realistic malicious behaviour than trust-based discounting or deletion.

6. DISCUSSION

DL-Lite is a tractable subset of DLs with a large number of application areas [4]. Its scalability makes it very useful especially for the settings where large amount of data should be queried. However, in a network of heterogeneous sources, any information provided by the sources could be uncertain, incomplete, and even conflicting. DL-Lite cannot accommodate such information. Pan et al. [11] proposed a framework of tractable query answering algorithms for a family of fuzzy query languages over large fuzzy DL-Lite [16] ontologies. On the other hand, DST and its extensions such as Subjective Logic explicitly takes into account *uncertainty* and *belief ownership* [9].

Gobeck and Halaschek [8] present a belief revision algorithm for OWL-DL, which is based on trust degrees to remove conflicting statements from a knowledge base. However, as the authors point out, the proposed algorithm is not guaranteed to be optimal. In our work, we embed statement retraction implicitly into the opinion revision procedure with a global optimal criteria which is grounded on a Beta distribution formalisation of trust.

Fact-finding algorithms aim to identify the *truth* given conflicting claims. Pasternack and Roth [12] propose to translate these claims to a linear program, which is solved to obtain belief scores over claims. For example, with TruthFinder [17], the belief scores obtained can be interpreted as the result of simultaneously minimising the frustration coming from the sources against the claims. These approaches do not consider semantics while reasoning about belief and trustworthiness as we do here.

Costa Periera and Tettamanzi [6] deal with belief changes in an agent's mental state considering trust in information sources. Dong *et al*. [7] propose to resolve conflicts in information from multiple sources by a voting mechanism. Double counting in votes is avoided by taking into account information dependence among the sources. The dependence is derived from Bayesian analysis over data sets held by the sources with a statistical interpretation.

In this paper, we propose $\mathcal{S}$DL-Lite, which is expressive enough to represent and reason about uncertain information using trust and domain knowledge. It allows us to efficiently identify conflicting information with respect to domain constraints. Then, these conflicts are resolved through the methods we propose for trust revision. Through simulations, we show that our approach can successfully handle highly misleading information in challenging settings. The simulations also show that the approach is robust in the face of strategic liars. In this paper, mostly for clarity, we assume opinions about each assertion is provided by a single information source. In the future, we will extend our approach to handle multiple sources.

7. REFERENCES

[1] F. Baader, D. L. McGuiness, D. Nardi, and P. Patel-Schneider, editors. *Description Logic Handbook: Theory, implementation and applications*. Cambridge University Press, 2002.

[2] C. Burnett, T. J. Norman, and K. Sycara. Bootstrapping trust evaluations through stereotypes. In *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems*, pages 241–248, 2010.

[3] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Data complexity of query answering in description logics. In *Proc. of KR 2006*, pages 260–270, 2006.

[4] D. Calvanese, G. Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The dl-lite family. *J. Autom. Reason.*, 39(3):385–429, 2007.

[5] D. Calvanese, G. D. Giacomo, M. Lenzerini, R. Rosati, and G. Vetere. DL-Lite: Practical Reasoning for Rich DLs. In *Proc. of the DL2004 Workshop*, 2004.

[6] C. da Costa Pereira and A. G. B. Tettamanzi. Goal generation with relevant and trusted beliefs. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*, Estoril, Portugal, 2008.

[7] X. L. Dong, L. Berti-Equille, and D. Srivastava. Integrating conflicting data: The role of source dependence. In *Proc. of the 35th International Conference on Very Large Databases*, Lyon, France, August 2009.

[8] J. Golbeck and C. Halaschek-Wiener. Trust-based revision for expressive web syndication. *Journal of Logic and Computation*, 19(5):771–790, Oct. 2009.

[9] A. Jøsang. *Subjective Logic*. Book Draft, 2011.

[10] A. Jøsang and R. Ismail. The beta reputation system. In *Proc. of the 15th Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy*, pages 48–64, 2002.

[11] J. Z. Pan, G. Stamou, G. Stoilos, S. Taylor, and E. Thomas. Scalable Querying Services over Fuzzy Ontologies. In *the Proc. of the 17th International World Wide Web Conference (WWW2008)*, 2008.

[12] J. Pasternak and D. Roth. Knowing what to believe (when you already know something). In *Proc. of the 23rd International Conference on Computational Linguistics*, Beijing, China, 2010.

[13] A. Rogers, R. K. Dash, and N. R. Jennings. Computational mechanism design for information fusion within sensor networks. In *In Proceedings of The 9th International Conference on Information Fusion*, 2006.

[14] A. Ruszczynski. *Nonlinear optimization*, volume 13. Princeton university press, 2011.

[15] G. Shafer. *A mathematical theory of evidence*. Princeton university press, 1976.

[16] U. Straccia. Answering vague queries in fuzzy DL-Lite. In *Proc. of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 2238–2245, 2006.

[17] X. Yin, J. Han, and P. S. Yu. Truth discovery with multiple conflicting information providers on the web. In *Proceedings of the Conference on Knowledge and Data Discovery*, 2007.